UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA – UFSCar-USP

Wesley Bertoli da Silva

A new class of discrete models for the analysis of zero-modified count data

Thesis submitted to the Department of Statistics – DEs/UFSCar and to the Institute of Mathematics and Computer Sciences – ICMC-USP in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics.

Advisor: Prof. Dr. Francisco Louzada Neto

São Carlos
April 2020

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA – UFSCar-USP

Wesley Bertoli da Silva

Uma nova classe de modelos discretos para a análise de dados de contagem zero-modificados

Tese apresentada ao Departamento de Estatística – DEs/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Francisco Louzada Neto

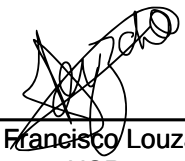
São Carlos
Abril de 2020

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

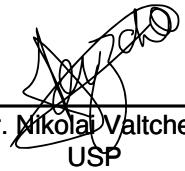
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Wesley Bertoli da Silva, realizada em 03/04/2020:



Prof. Dr. Francisco Louzada Neto
USP

Prof. Dr. Marinho Gomes de Andrade Filho
USP

Prof. Dr. Caio Lucidius Naberezny Azevedo
UNICAMP



Prof. Dr. Nikolai Valtchev Kolev
USP

Prof. Dr. Paulo Henrique Ferreira da Silva
UFBA

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Francisco Louzada Neto Marinho Gomes de Andrade Filho, Caio Lucidius Naberezny Azevedo, Nikolai Valtchev Kolev, Paulo Henrique Ferreira da Silva e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Francisco Louzada Neto

ACKNOWLEDGEMENTS

Gostaria de expressar meus sinceros agradecimentos

Aos meus pais, Edmilson e Sirlene, pela vida, pelos valores ensinados, pelo amor e apoio incondicionais, por compreenderem a ausência do primogênito ao longo dos últimos anos, e por serem os maiores exemplos de luta, trabalho e honestidade que eu levo para a vida,

À minha noiva Deisy (que assumiu esse *status* exatamente no período em que ambos estavam em suas jornadas de doutorado), também pelo amor e apoio incondicionais, pelo carinho e paciência, pela sensibilidade e empatia, pelo companherismo, por ser minha inspiração e motivo de orgulho, por ser fantástica como ser humano, por ser a melhor companheira de viagem e, principalmente, por ter sido a grande responsável por hoje eu ser uma pessoa melhor,

Aos meus familiares que me acompanham e me apoiam desde o início desta tragetória e que, apesar da ausência e da distância, acreditam na minha capacidade de superar barreiras, vencer desafios e alcançar objetivos,

Ao Professor Francisco Louzada pela orientação, pela liberdade concedida e por acreditar na minha capacidade para desenvolver este trabalho,

À Professora Katiane e ao Professor Marinho pela coorientação informal, pelas ideias perspicazes, e por terem sido diretamente responsáveis pela mudança no rumo do meu projeto de pesquisa,

Aos professores e funcionários do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs) por todo o conhecimento compartilhado e pelo apoio nas questões burocráticas,

Aos amigos da minha turma de doutorado que, durante os quatro anos de curso, compartilharam experiências, opiniões e trouxeram momentos de descontração ao longo dos intensos dias de estudo; aos amigos que residem na cidade de São Paulo, pela amizade, pelos exemplos de luta e pelos divertidos momentos de confraternização,

À Universidade de São Paulo (USP) e à Universidade Federal de São Carlos (UFSCar) que, por meio de suas estruturas organizacionais, forneceram condições adequadas para a realização deste trabalho,

À Universidade Tecnológica Federal do Paraná (UTFPR) que, por meio de seu programa de aperfeiçoamento pessoal e profissional, concedeu afastamento integral para que eu pudesse me dedicar integralmente aos estudos e à realização deste trabalho,

À Fundação Araucária que, por meio de sua parceria com CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), forneceu apoio financeiro durante parte do período de desenvolvimento deste projeto,

E, finalmente, a todos que de alguma forma contribuíram para a realização deste trabalho.

“Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less.”
(Marie Curie)

RESUMO

BERTOLI, W. **Uma nova classe de modelos discretos para a análise de dados de contagem zero-modificados**. 2020. 321 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Neste trabalho, uma nova classe de modelos discretos para a análise de contagens zero modificados foi introduzida. A classe proposta é composta pelas versões *hurdle* das distribuições de Poisson-Lindley, Poisson-Shanker e Poisson-Sujatha, que são misturas uniparamétricas de Poisson, capazes de acomodar diferentes níveis de sobredispersão. Diferentemente da formulação tradicional das distribuições zero modificadas, a principal suposição acerca de um modelo *hurdle* é que as observações positivas são inteiramente representadas por distribuições zero-truncadas. No sentido de estender a aplicabilidade dos modelos teóricos, também foi desenvolvida uma estrutura de regressão com efeitos fixos, na qual tanto a probabilidade de se observar o valor zero, quanto o número médio de observações positivas por indivíduo, puderam ser modelados na presença de covariáveis. Além disso, também foi desenvolvida uma estrutura ainda mais flexível, permitindo a inclusão simultânea de efeitos fixos e aleatórios nos preditores lineares do modelo *hurdle*. Na estrutura de efeitos mistos derivada, considerou-se o uso de efeitos aleatórios escalares para quantificar a heterogeneidade entre as observações de um mesmo indivíduo, que decorre de agrupamentos ou medidas repetidas. Neste trabalho, todos os procedimentos inferenciais foram conduzidos sob uma perspectiva totalmente Bayesiana. Diferentes distribuições a priori foram consideradas (por exemplo, *Jeffreys'* e *g-prior*), e a tarefa de gerar valores pseudo-aleatórios de uma distribuição a posteriori sem forma fechada foi realizada por um dos três algoritmos a seguir (dependendo da estrutura de cada modelo): Amostragem por Rejeição, Random-walk Metropolis, e Metropolis Adaptativo. Estudos intensivos de simulação de Monte Carlo foram realizados como forma de avaliar o desempenho das metodologias Bayesianas adotadas. A utilidade dos modelos zero modificados propostos foi ilustrada usando vários conjuntos de dados reais que apresentavam diferentes estruturas e fontes de variação. Além de estimar os parâmetros, foram realizadas análises de sensibilidade para identificar pontos influentes e, para avaliar os modelos ajustados, foram computados os p-valores Bayesianos, os resíduos quantílicos aleatorizados, entre outras medidas. Por fim, quando comparados com distribuições bem estabelecidas que são úteis para a análise de dados de contagem, a competitividade dos modelos propostos foi comprovada em todos os exemplos fornecidos.

Palavras-chave: Dados zero modificados, distribuições de mistura de Poisson, métodos Bayesianos, modelo *hurdle* com efeitos mistos, sobredispersão.

ABSTRACT

BERTOLI, W. **A new class of discrete models for the analysis of zero-modified count data**. 2020. 321 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

In this work, a new class of discrete models for the analysis of zero-modified count data has been introduced. The proposed class is composed of hurdle versions of the Poisson-Lindley, Poisson-Shanker, and Poisson-Sujatha baseline distributions, which are uniparametric Poisson mixtures that can accommodate different levels of overdispersion. Unlike the traditional formulation of zero-modified distributions, the primary assumption under hurdle models is that the positive observations are entirely represented by zero-truncated distributions. In the sense of extending the applicability of the theoretical models, it has also been developed a fixed-effects regression framework, in which the probability of zero-valued observations being generated as well as the average number of positive observations per individual could be modeled in the presence of covariates. Besides, an even more flexible structure allowing the inclusion of both fixed and random-effects in the linear predictors of the hurdle models has also been developed. In the derived mixed-effects structure, it has been considered the use of scalar random-effects to quantify the within-subjects heterogeneity arising from clustering or repeated measurements. In this work, all inferential procedures were conducted under a fully Bayesian perspective. Different *prior* distributions have been considered (e.g., *Jeffreys'* and *g-prior*), and the task of generating pseudo-random values from a *posterior* distribution without closed-form has been performed by one out of the three following algorithms (depending on the structure of each model): Rejection Sampling, Random-walk Metropolis, and Adaptive Metropolis. Intensive Monte Carlo simulation studies were performed in order to evaluate the performance of the adopted Bayesian methodologies. The usefulness of the proposed zero-modified models was illustrated by using several real datasets presenting different structures and sources of variation. Beyond parameter estimation, it has been performed sensitivity analyses to identify influent points, and, in order to evaluate the fitted models, it has been computed the Bayesian p-values, the randomized quantile residuals, among other measures. Finally, when compared with well-established distributions for the analysis of count data, the competitiveness of the proposed models has been proved in all provided examples.

Keywords: Bayesian methods, mixed-effects hurdle models, overdispersion, Poisson mixture distributions, zero-modified data.

LIST OF FIGURES

Figure 1 – Supremum of set $\mathcal{A}_\theta$ depending on each element of the class $\mathcal{F}_2$ 40
Figure 2 – Behavior of each element of the class $\mathcal{F}_3$ for different values of θ and p . . 42
Figure 3 – *Posterior* expected frequencies under the fitted models 60
Figure 4 – Behavior of the $\mathcal{ZMPS}_h$ distribution for different values of μ and p 66
Figure 5 – *Posterior* estimates of parameter p using zero-inflated samples $(n = 1{,}000)$ 79
Figure 6 – *Posterior* estimates of parameter p using zero-deflated samples $(n = 1{,}000)$ 81
Figure 7 – Frequency distribution of the response variable 82
Figure 8 – Sensitivity analysis for diagnostic of influential points 84
Figure 9 – *Posterior* estimates of parameters p_0, p, and μ 86
Figure 10 – Normal probability plot for the randomized quantile residuals 87
Figure 11 – *Posterior* expected frequencies under zero-modified models 88
Figure 12 – Behavior of the $\mathcal{ZMPL}$ distribution for different values of μ and p 96
Figure 13 – *Posterior* estimates of parameter p using zero-inflated samples $(n = 500)$ . . 109
Figure 14 – *Posterior* estimates of parameter p using zero-deflated samples $(n = 500)$. 109
Figure 15 – Sensitivity analysis for diagnostic of influential points 112
Figure 16 – Estimated *posterior* densities of vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ 114
Figure 17 – Frequency distribution and Half-Normal plot with simulated envelope for the RQRs 116
Figure 18 – *Posterior* expected frequencies under zero-modified models 116
Figure 19 – *Posterior* estimates of parameter p using zero-modified samples $(n = 100)$. 137
Figure 20 – Summary of the number of dicentrics and centric rings aberrations 139
Figure 21 – Trace plots and marginal *posterior* distributions of the $\mathcal{ZMPS}_u$ model parameters 140
Figure 22 – Sensitivity analysis for diagnostic of influential points 142
Figure 23 – Frequency distribution and Half-Normal plot with simulated envelope for the RQRs 143
Figure 24 – *Posterior* estimates of parameters p_0 and p 144
Figure 25 – *Posterior* expected frequencies and dose-response curve fitted by the zero-modified models 144
Figure 26 – Descriptive summary (per treatment) of grooming counting data 172
Figure 27 – *Posterior* descriptive summary for the $\mathcal{P}$, $\mathcal{NB}$, and $\mathcal{PL}$ fitted models 173
Figure 28 – Trace plots and marginal *posterior* distributions of parameters of the $\mathcal{ZMPL}_2$ model 175

Figure 29 – Sensitivity analysis for diagnostic of influential points 178
Figure 30 – Frequency distribution and Half-Normal plot with simulated envelope for the RQRs 179
Figure 31 – *Posterior* estimates (per treatment) of representative parameters λ_{ij} and ω_{ij} . 180
Figure 32 – *Posterior* estimates (per treatment) of parameters p_0 and p over time 180
Figure 33 – *Posterior* expected frequencies (per treatment) under the $\mathcal{ZMPL}_2$ fitted model 181
Figure 34 – BRE of estimated $\mathcal{ZMPL}$ distribution when $\mathcal{PL}$ is the true model 241
Figure 35 – BRE of estimated $\mathcal{ZMPS}_h$ distribution when $\mathcal{PS}_h$ is the true model 241
Figure 36 – BRE of estimated $\mathcal{ZMPS}_u$ distribution when $\mathcal{PS}_u$ is the true model 243
Figure 37 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 1 and 2) 258
Figure 38 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 3 and 4) 259
Figure 39 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 1 and 2) 266
Figure 40 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 3 and 4) 267
Figure 41 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 1 and 2) 274
Figure 42 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 3 and 4) 275
Figure 43 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 1 and 2) 282
Figure 44 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 3 and 4) 283
Figure 45 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 1 and 2) 290
Figure 46 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 3 and 4) 291
Figure 47 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 1 and 2) 298
Figure 48 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 3 and 4) 299

LIST OF ALGORITHMS

Algorithm 1 – Rejection Sampling 48
Algorithm 2 – Sequential-Search 51
Algorithm 3 – Random-walk Metropolis 72
Algorithm 4 – Sequential-Search 75
Algorithm 5 – Random-walk Metropolis 102
Algorithm 6 – Sequential-Search 103
Algorithm 7 – Random-walk Metropolis 251
Algorithm 8 – Sequential-Search 251
Algorithm 9 – Adaptive Metropolis 302
Algorithm 10 – Sequential-Search 302
Algorithm 11 – Inverse transform sampling 303

LIST OF TABLES

Table 1 – Characterization of each element of the class $\mathcal{F}_1$ 36
Table 2 – Characterization of each element of the class $\mathcal{F}_2$ 37
Table 3 – Expressions of τ_1, τ_2 and τ_3 for each element of the class $\mathcal{F}_2$ 38
Table 4 – Some theoretical measures regarding each element of the class $\mathcal{F}_2$ 39
Table 5 – Value of some theoretical measures regarding each element of the class $\mathcal{F}_3$. 43
Table 6 – Actual parameter values for simulation of zero-modified artificial datasets . . 50
Table 7 – Response variables and some descriptive statistics for each dataset 55
Table 8 – Frequency distribution of each dataset 56
Table 9 – *Posterior* estimates of parameter θ and 95% BCIs from the proposed models 56
Table 10 – Comparison criteria and adequacy of the fitted models 57
Table 11 – *Posterior* estimates of parameter ω and 95% BCIs from the proposed models 58
Table 12 – *Posterior* estimates of parameter p and 95% BCIs from the proposed models 59
Table 13 – *Posterior* estimates of extra parameters from the proposed models 59
Table 14 – Empirical properties of the Bayesian estimators using zero-inflated samples . 77
Table 15 – Empirical properties of the Bayesian estimators using zero-deflated samples . 78
Table 16 – Coverage probabilities (%) of the BCIs using zero-inflated samples 79
Table 17 – Coverage probabilities (%) of the BCIs using zero-deflated samples 80
Table 18 – *Posterior* parameter estimates and 95% BCIs from $\mathcal{ZMPS}_h$ fitted model (full dataset) 83
Table 19 – Comparison criteria for the fitted models 84
Table 20 – *Posterior* parameter estimates and 95% BCIs from $\mathcal{ZMPS}_h$ model (without influential points) 85
Table 21 – *Posterior* estimates of the expected number of zeros (n_0) 87
Table 22 – Summary of Bayesian and classical estimation using zero-inflated samples . 106
Table 23 – Summary of Bayesian and classical estimation using zero-deflated samples . 107
Table 24 – Coverage probabilities (%) of the BCIs/ACIs using zero-inflated samples . . 108
Table 25 – Coverage probabilities (%) of the BCIs/ACIs using zero-deflated samples . . 108
Table 26 – *Posterior* parameter estimates and 95% BCIs/HPDIs from $\mathcal{ZMPL}$ fitted model 110
Table 27 – Sensitivity analysis to evaluate the effect of different *prior* specifications . . . 111
Table 28 – Comparison criteria for the fitted models 113
Table 29 – *Posterior* parameter estimates and 95% BCIs/HPDIs from $\mathcal{ZMPL}$ fitted model (without influential point) 113
Table 30 – *Posterior* estimates of extra parameters and goodness-of-fit evaluation 115

Table 31 – Actual parameter values for simulation of zero-modified artificial datasets. . . 132
Table 32 – Empirical properties of the Bayesian estimators using zero-inflated samples . 134
Table 33 – Empirical properties of the Bayesian estimators using zero-deflated samples . 135
Table 34 – Coverage probabilities (%) of the HPDIs using zero-inflated samples 136
Table 35 – Coverage probabilities (%) of the HPDIs using zero-deflated samples 136
Table 36 – Descriptive summary of the number of dicentrics and centric rings aberrations 139
Table 37 – *Posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPS}_u$ fitted model . 141
Table 38 – Comparison criteria and adequacy measures for the fitted models 143
Table 39 – *Posterior* parameter estimates and goodness-of-fit evaluation 145
Table 40 – Actual parameter values for simulation of zero-modified artificial datasets . . 169
Table 41 – Descriptive summary of grooming counting data per rat 171
Table 42 – Comparison criteria and adequacy measures for the fitted models 174
Table 43 – *Posterior* estimates for the Pseudo-Bayes Factor ($\mathcal{ZMPL}$ and $\mathcal{ZMNB}$-based models) 176
Table 44 – *Posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPL}_2$ fitted model . 177
Table 45 – *Posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPL}_2$ fitted model (without influential point) 178
Table 46 – Function $\Psi(\theta)$ for different values of θ 202
Table 47 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 0.50$) 205
Table 48 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 5.00$) 206
Table 49 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-inflated case with $\theta = 0.50$) 207
Table 50 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-inflated case with $\theta = 5.00$) 208
Table 51 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 0.50$) 209
Table 52 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 5.00$) 209
Table 53 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 0.50$) 210
Table 54 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 5.00$) 211
Table 55 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-inflated case with $\theta = 0.50$) 212
Table 56 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-inflated case with $\theta = 5.00$) 213

Table 57 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 0.50$) . . . 214
Table 58 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 5.00$) . . . 214
Table 59 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 0.50$) . . . 215
Table 60 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 5.00$) . . . 216
Table 61 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-inflated case with $\theta = 0.50$) . . . 217
Table 62 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-inflated case with $\theta = 5.00$) . . . 218
Table 63 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 0.50$) . . . 219
Table 64 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 5.00$) . . . 219
Table 65 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 2.50$) . . . 220
Table 66 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 6.00$) . . . 221
Table 67 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-deflated case with $\theta = 2.50$) . . . 222
Table 68 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-deflated case with $\theta = 6.00$) . . . 223
Table 69 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 2.50$) . . . 224
Table 70 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 6.00$) . . . 224
Table 71 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 2.50$) . . . 225
Table 72 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 6.00$) . . . 226
Table 73 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-deflated case with $\theta = 2.50$) . . . 227
Table 74 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-deflated case with $\theta = 6.00$) . . . 228
Table 75 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 2.50$) . . . 229

Table 76 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 6.00$) 229
Table 77 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 2.50$) 230
Table 78 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 6.00$) 231
Table 79 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-deflated case with $\theta = 2.50$) 232
Table 80 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-deflated case with $\theta = 6.00$) 233
Table 81 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 2.50$) 234
Table 82 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 6.00$) 234
Table 83 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPL}$ is the true model (zero inflation) 235
Table 84 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPL}$ is the true model (zero deflation) 236
Table 85 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_h$ is the true model (zero inflation) 237
Table 86 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_h$ is the true model (zero deflation) 238
Table 87 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_u$ is the true model (zero inflation) 239
Table 88 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_u$ is the true model (zero deflation) 240
Table 89 – Real datasets used in the paper 242
Table 90 – MC estimators for some φ-divergence measures and its calibration 252
Table 91 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2) 253
Table 92 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4) 254
Table 93 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2) 255
Table 94 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4) 256
Table 95 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2) 257

Table 96 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4) . . . 257
Table 97 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2) . . . 261
Table 98 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4) . . . 262
Table 99 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2) . . . 263
Table 100 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4) . . . 264
Table 101 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2) . . . 265
Table 102 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4) . . . 265
Table 103 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2) . . . 269
Table 104 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4) . . . 270
Table 105 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2) . . . 271
Table 106 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4) . . . 272
Table 107 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2) . . . 273
Table 108 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4) . . . 273
Table 109 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2) . . . 277
Table 110 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4) . . . 278
Table 111 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2) . . . 279
Table 112 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4) . . . 280
Table 113 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2) . . . 281
Table 114 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4) . . . 281

Table 115 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2) . . . 285
Table 116 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4) . . . 286
Table 117 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2) . . . 287
Table 118 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4) . . . 288
Table 119 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2) . . . 289
Table 120 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4) . . . 289
Table 121 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2) . . . 293
Table 122 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4) . . . 294
Table 123 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2) . . . 295
Table 124 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4) . . . 296
Table 125 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2) . . . 297
Table 126 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4) . . . 297
Table 127 – AMC estimators for some well-known φ-divergence measures and its calibration 301
Table 128 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2) . . . 304
Table 129 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2) . . . 305
Table 130 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2) . . . 306
Table 131 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4) . . . 307
Table 132 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4) . . . 308
Table 133 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4) . . . 309
Table 134 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 5 and 6) . . . 310

Table 135 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 5 and 6) . . . 311
Table 136 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 5 and 6) . . . 312
Table 137 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2) . . . 313
Table 138 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2) . . . 314
Table 139 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2) . . . 315
Table 140 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4) . . . 316
Table 141 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4) . . . 317
Table 142 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4) . . . 318
Table 143 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 5 and 6) . . . 319
Table 144 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 5 and 6) . . . 320
Table 145 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 5 and 6) . . . 321

CONTENTS

1 INTRODUCTION 29
1.1 Background 29
1.2 Research goals 32
1.3 Organization and content of chapters 33

2 A NEW CLASS OF ZERO-MODIFIED POISSON MIXTURE MODELS 35
2.1 Introduction 35
2.2 Poisson mixture models 36
2.3 Zero-modified models 39
2.4 Hurdle models 44
2.5 Inference 45
2.5.1 Maximum likelihood *45*
2.5.2 Prior distributions *46*
2.5.3 Posterior distributions and estimation *47*
2.5.4 Posterior predictive distribution *49*
2.6 Simulation studies 50
2.6.1 Study 1 . . . . . *51*
2.6.2 Study 2 . . . . . *53*
2.6.3 Study 3 . . . . . *54*
2.7 Application to real data 55
2.8 Concluding remarks 60

3 THE ZERO-MODIFIED POISSON-SHANKER REGRESSION MODEL 63
3.1 Introduction 63
3.2 The PS_h distribution 64
3.3 The $ZMPS_h$ distribution 65
3.4 Hurdle version of the PS_h distribution 67
3.5 The $ZMPS_h$ regression model 68
3.6 Inference 69
3.6.1 Prior and Posterior distributions *70*
3.6.2 Influential points *72*
3.6.3 Residual analysis *74*

3.7 Simulation study . . . 74
3.7.1 Zero-inflated artificial data . . . *80*
3.7.2 Zero-deflated artificial data . . . *81*
3.8 Fetal deaths data analysis . . . 81
3.9 Concluding remarks . . . 88

4 THE ZERO-MODIFIED POISSON-LINDLEY REGRESSION MODEL 91
4.1 Introduction . . . 91
4.2 The PL distribution . . . 92
4.3 The ZMPL distribution . . . 93
4.4 The ZMPL regression model . . . 97
4.5 Inference . . . 98
4.5.1 Prior distributions . . . *99*
4.5.2 Posterior distributions and estimation . . . *100*
4.6 Simulation study . . . 102
4.6.1 Zero-inflated artificial data . . . *105*
4.6.2 Zero-deflated artificial data . . . *105*
4.7 Takeover bids data analysis . . . 109
4.8 Concluding remarks . . . 117

5 THE ZERO-MODIFIED POISSON-SUJATHA REGRESSION MODEL 119
5.1 Introduction . . . 119
5.2 The ZMPS$_u$ regression model . . . 121
5.3 Inference . . . 123
5.3.1 Prior distributions . . . *124*
5.3.2 Posterior distributions and estimation . . . *125*
5.3.3 Posterior predictive distribution . . . *127*
5.3.4 Influential points . . . *127*
5.3.5 Model comparison and adequacy . . . *129*
5.3.6 Residual analysis . . . *131*
5.4 Simulation study . . . 131
5.5 Chromosomal aberration data analysis . . . 138
5.6 Concluding remarks . . . 145

6 THE ZERO-MODIFIED POISSON-LINDLEY REGRESSION MODEL WITH MIXED-EFFECTS . . . 147
6.1 Introduction . . . 147
6.2 The ZMPL mixed-effects regression model . . . 149
6.2.1 Testing for zero modification . . . *152*
6.3 Approximate Bayesian inference and prediction . . . 153

6.3.1 Prior *distributions* . . . 156
6.3.1.1 *Fixed-effects parameters* . . . 156
6.3.1.2 *Random-effects parameters* . . . 158
6.3.2 Posterior *distribution and estimation* . . . 159
6.3.3 *Random-effects prediction* . . . 161
6.3.4 Posterior *predictive distribution* . . . 162
6.3.5 *Influential points* . . . 163
6.3.6 *Model comparison and adequacy* . . . 165
6.3.7 *Residual analysis* . . . 167
6.4 Simulation study . . . 168
6.5 Real data application . . . 171
6.6 Concluding remarks . . . 181

7 CONCLUSIONS AND PERSPECTIVES . . . 185
7.1 Perspectives . . . 187

BIBLIOGRAPHY . . . 189

APPENDIX A SUPPLEMENT FOR CHAPTER 2 . . . 201
A.1 Maximum likelihood estimation . . . 201
A.2 Model comparison and suitability . . . 203
A.3 Results from the simulation studies . . . 204
A.3.1 *Study 1* . . . 205
A.3.1.1 *Zero-inflated artificial data* . . . 205
A.3.1.2 *Zero-deflated artificial data* . . . 220
A.3.2 *Study 2* . . . 235
A.3.3 *Study 3* . . . 240
A.4 Datasets . . . 242

APPENDIX B SUPPLEMENT FOR CHAPTER 4 . . . 245
B.1 Influential points . . . 245
B.2 Model comparison . . . 247
B.3 Residual analysis . . . 248

APPENDIX C SUPPLEMENT FOR CHAPTER 5 . . . 251
C.1 Algorithms . . . 251
C.1.1 *Random-walk Metropolis* . . . 251
C.1.2 *Sequential-Search* . . . 251
C.2 Divergence measures . . . 252
C.3 Results from the simulation study . . . 252

C.3.1 Zero-inflated artificial data . . . 252
C.3.1.1 Using logit link function . . . 252
C.3.1.2 Using probit link function . . . 261
C.3.1.3 Using complementary log-log link function . . . 269
C.3.2 Zero-deflated artificial data . . . 277
C.3.2.1 Using logit link function . . . 277
C.3.2.2 Using probit link function . . . 285
C.3.2.3 Using complementary log-log link function . . . 293

APPENDIX D SUPPLEMENT FOR CHAPTER 6 . . . 301
D.1 Divergence measures . . . 301
D.2 Algorithms . . . 302
D.2.1 Adaptive Metropolis . . . 302
D.2.2 Sequential-Search . . . 302
D.2.3 Inverse transform sampling . . . 302
D.3 Results from the simulation study . . . 303
D.3.1 Zero-inflated artificial data . . . 303
D.3.2 Zero-deflated artificial data . . . 312

CHAPTER

1

INTRODUCTION

1.1 Background

Most of the classical statistical methods were developed under the strong assumption that observations from a random variable are independent and identically distributed, usually drawn from homogeneous populations. However, the primary challenge for modern statisticians is to handle data structures that violate such assumptions. Particularly, the modeling of count data deserves special attention (MOLENBERGHS; VERBEKE, 2005) as it may be necessary to use models that accommodate different sources of variation arising, for example, from an alteration in the observed frequency of a specific outcome.

The ordinary Poisson distribution is often adopted for the analysis of count data, mainly due to its simplicity and by having its computational implementation available for most of the usual statistical packages. However, it is well-known that such a model is not suitable to describe counts in which the variance-to-mean ratio is not (at least) close to 1. Apart from data transformation, the most popular approach to circumvent such an issue is the use of finite mixture models (MCLACHLAN; PEEL, 2004) that can accommodate different levels of overdispersion (KARLIS; XEKALAKI, 2005). A unified approach for obtaining mixtures of a Poisson distribution and a model belonging to the exponential family is provided by (BARRETO-SOUZA; SIMAS, 2016).

The Negative Binomial distribution (that may arise as a Poisson mixture model by using a Gamma distribution for the continuous part) is undoubtedly the most popular alternative to model extra-Poisson variability. Nevertheless, there is extensive literature regarding other discrete mixed distributions that can accommodate different levels of overdispersion, for example, the Poisson-Lindley (SANKARAN, 1970), the Poisson-Lognormal (BULMER, 1974), the Poisson-Inverse Gaussian (SHABAN, 1981), the Negative Binomial-Lindley (ZAMANI; ISMAIL, 2010), the Poisson-Janardan (SHANKER *et al.*, 2014), the two-parameter

Poisson-Lindley (SHANKER; MISHRA, 2014), the Poisson-Amarendra (SHANKER, 2016b), the Poisson-Shanker (SHANKER, 2016c), the Poisson-Sujatha (SHANKER, 2016d), the quasi Poisson-Lindley (SHANKER; MISHRA, 2016), the Weighted Negative Binomial-Lindley (BAKOUCH, 2018), the Poisson-Weighted Lindley (SHANKER; SHUKLA, 2018), the Binomial-Discrete Lindley (KUŞ *et al.*, 2019), among many others.

Unfortunately, there is a significant drawback regarding such mixture models, which is the fact that they do not fit well when data presents a modification in the frequency of zeros (typically underestimate the data dispersion and the frequency of zero-valued observations). The most common case in practice is the presence of an excessive number of zero-valued observations along with a nonhomogeneous/skewed distribution of the positive values. In this way, developing two-part models (zero-inflated/hurdle models) based on the Poisson distribution became necessary. Prominent works addressing this task are Cohen (1960), Umbach (1981), Mullahy (1986), Lambert (1992), Zorn (1996), Deb and Trivedi (2002), McDowell (2003), and Wagh and Kamalja (2018).

Several authors have considered these approaches for the analysis of real data, and here we point out a few. Gurmu and Trivedi (1996) have sought to deal with the excess of zeros on data from recreational trips. Bohara and Krieg (1996) have shown that the modeling of migratory frequency data can be improved by using zero-inflated Poisson models. Ridout, Demétrio and Hinde (1998) have exploited the apple shoot propagation data, and they have addressed the modeling task by using several zero-inflated regression models. In the social sciences, Bahn and Massenburg (2008) have considered the hurdle version of the Poisson model for the number of homicides in Chicago (State of Illinois, US). An application to private health insurance count data using ordinary and zero-inflated Poisson regression models was provided by Mouatassim and Ezzahid (2012). Further applications of these models were considered in quantitative studies about HIV-risk reduction (HEILBRON; GIBSON, 1990; HU; PAVLICOVA; NUNES, 2011), for the modeling of some occupational allergic diseases in France (NGATCHOU-WANDJI; PARIS, 2011), for the analysis of DNA sequencing data (BEUF *et al.*, 2012), and the modeling of several datasets on chromosomal aberrations induced by radiation (OLIVEIRA *et al.*, 2016). A Bayesian approach for the zero-inflated Poisson distribution was considered by Rodrigues (2003), and by Ghosh, Mukhopadhyay and Lu (2006) in a regression framework with fixed-effects.

Noticeably, the majority of developed works are focused on the modeling of zero inflation, but zero-deflated data are also frequently observed in practice. However, still, there are very few studies addressing this case solely (DIETZ; BÖHNING, 2000; ANGERS; BISWAS, 2003), even if this situation is often referred to in works handling zero inflation. In this context, a more comprehensive approach is provided by zero-modified models, which are flexible tools to handle count data with inflation/deflation at zero when there is no information about the nature of such a phenomenon. Some of the most relevant works about zero-modified and hurdle models are cited in the following. Dietz and Böhning (2000) have introduced the zero-modified

Poisson regression model, and Conceição, Andrade and Louzada (2013) have considered such a model as an alternative for the analysis of Brazilian leptospirosis notification data. Min and Agresti (2005) have pointed out the differences between the hurdle and zero-inflated versions of the Poisson distribution with an interesting comparison based on simulations. The possible loss due to the specification of a zero-modified Poisson model for the analysis of samples without zero modification was studied by Conceição, Andrade and Louzada (2014) using the Kullback-Leibler divergence. A dynamic hurdle model for zero-inflated count data was derived by Baetschmann and Winkelmann (2017). Besides, the hurdle version of the Power Series distribution was presented and well discussed by Conceição *et al.* (2017a), and Conceição *et al.* (2017b) have adopted a Bayesian approach for the zero-modified Poisson model to predict match outcomes of the Spanish La Liga (Season 2012-2013).

Beyond overdispersion and zero modification, it is increasingly common in many areas of quantitative research the obtaining, for example, of correlated experimental outcomes, such as clustered data, repeated measures, and longitudinal (panel) data. Clustered data comes from a subdivision of the target population into disjoint subsets (clusters), which can be composed either by a single or a collection of individuals (states, municipalities, farms, hospitals, schools, families, and so on). In some studies involving clusterization, the characteristic of interest is sequentially measured (repeated measurements) or repeatedly observed over time (longitudinal data). In these cases, the assumption of independence within clusters and the specification of a single distribution to describe different subpopulations may lead to erroneous inferences and misleading sample-based conclusions, mainly if the phenomenon under investigation is exposed to varying experimental conditions.

A handy model-based approach to address lack of independence within clusters (due to the intrinsic correlation between measurements) consists of including fixed-effects (under the availability of covariates) in a flexible regression framework that can also accommodate random-effects. Mixed-effects models are robust parametric tools, widely used in the last half-century, mainly because of its versatility to handle correlated data, its flexibility to characterize underlying populations, and its capability to fit heterogeneous data with variability due to multiple sources. Usually, clustered non-normal correlated data can be analyzed using a large class of distributions through generalized linear mixed models (KACHMAN; STROUP, 1994).

Mixed models have great practical appeal in many fields of quantitative research, including agriculture, biology, medicine, sociology, and so on. The most popular model in this context is the linear mixed model, which provides a unifying approach for the analysis of a variety of correlated data (VERBEKE; MOLENBERGHS, 2000). This model is built under a hierarchical structure, in which the response variable/random-effects are assumed to be Gaussian. However, there are plenty of experiments whose primary outcome is non-normal, and a brief search on the extensive literature handling this case brings us some examples of the analysis of repeated measures/longitudinal data, in which the response variable is binary (WILLIAMSON *et al.*, 1996;

LANDERMAN; MUSTILLO; LAND, 2011; GAO; PAN; HABER, 2012; YIN *et al.*, 2014); is ordinal (JEYASEELAN; ANTONYSAMY; JOHN, 1996; LI; SCHAFER, 2008; BÜRGIN; RITSCHARD, 2015; RANA; ROY; DAS, 2018); is a proportion (HUNGER; DÖRING; HOLLE, 2012; XU *et al.*, 2013; GALVIS; BANDYOPADHYAY; LACHOS, 2014; BONAT; RIBEIRO JR.; ZEVIANI, 2015).

Robust approaches for the analysis of correlated count data (specifically under zero inflation) were proposed by Hall (2000), Min and Agresti (2005), and Alfò and Maruotti (2010). Particularly, the mixed-effects zero-inflated Poisson model was considered by some authors, including Yau and Lee (2001), which have considered such a model to evaluate an occupational injury prevention program, and Wang, Yau and Lee (2002), which have analyzed diagnosis-related groups with the majority of same-day hospital stays. Besides, Neelon, O'Malley and Normand (2010) have modeled repeated measures obtained from a study of outpatient psychiatric service use, Buu *et al.* (2012) have analyzed longitudinal zero-inflated count data in the substance abuse field, and Gupta, Szczesniak and Macaluso (2015) have dealt with the excess of zeros in an experimental study with repeat measurements on the number of problems with female condom use reported by women at high risk of contracting sexually transmitted diseases. Zero-inflated longitudinal measures were also analyzed by Zhu, Luo and DeSantis (2017), which have analyzed data from the largest clinical trial of alcohol dependence performed in the US. While zero deflation has not been addressed in this context until now, a comprehensive approach for the zero-modified Poisson and zero-modified Negative Binomial models in a regression framework with mixed-effects is provided by Neelon, O'Malley and Smith (2016a), with real data examples available in Neelon, O'Malley and Smith (2016b).

Far beyond motivating the development of this work, the content of this section provides a complete overview with several arguments supporting the objectives of this project – the research goals are presented in the following. Overall, the search for new discrete models that can accommodate different structures (e.g., overdispersion, zero modification, individual heterogeneity due to repeated measurements) may lead us to obtain potent statistical tools for analysis of count data.

1.2 Research goals

This doctoral thesis is the final result of a broad research project that was guided by the primary goal of introducing in the literature a brand new class of discrete models as an alternative for the analysis of zero-modified count data. In particular, we have worked on proposing a class of zero-modified distributions (expressed as hurdle models) in which the baseline components were uniparametric Poisson mixtures, namely, the Poisson-Lindley, the Poisson-Shanker, and the Poisson-Sujatha distributions.

The second leading goal of this work was twofold. We have focused on extending the

obtained theoretical zero-modified models to allow the inclusion of covariates in a comprehensive regression framework with (i) fixed-effects only and then, with (ii) fixed and random-effects (mixed-effects structure). This thesis comprises case (i) for all models of the proposed class, while case (ii) was developed only for the zero-modified Poisson-Lindley distribution.

1.3 Organization and content of chapters

A brief description of the chapters that form the core of this thesis is provided in the following. This work was organized in order to present a cohesive structure. However, as the chapters are (in general) independent from each other, the interested reader can adopt any reading order.

Chapter 2 is based on the paper entitled "A Bayesian approach for some zero-modified Poisson mixture models," published by Statistical Modelling in 2019 (<https://doi.org/10.1177/1471082X19841984>). In this work, we have proposed a new class of zero-modified models based on the hurdle version of Poisson mixture distributions (Poisson-Lindley/Shanker/Sujatha). As the title suggests, the inferential procedures were conducted under a fully Bayesian perspective. Besides, three different kinds of simulation studies were performed as a way of evaluating the behavior of the adopted methodology in some specific situations. The usefulness and competitiveness of the proposed class of models were illustrated using three real datasets.

Chapter 3 is based on the paper entitled "On the zero-modified Poisson-Shanker regression model and its application to fetal deaths notification data," published by Computational Statistics in 2018 (<https://doi.org/10.1007/s00180-017-0788-1>). In this work, we have introduced a new fixed-effects regression model based on the hurdle version of Poisson-Shanker distribution. All inferential procedures were conducted under a fully Bayesian perspective, and an intensive simulation study was performed, so we were able to evaluate the empirical properties of the Bayesian estimators in some specific situations. The proposed regression model was considered for the analysis of a real dataset on fetal death reported in all cities of Bahia State (Brazil) in 2014.

Chapter 4 is based on the paper entitled "Bayesian approach for the zero-modified Poisson-Lindley regression model," published by Brazilian Journal of Probability and Statistics in 2019 (<https://doi.org/10.1214/19-BJPS447>). In this work, we have introduced a new fixed-effects regression model based on the hurdle version of Poisson-Lindley distribution. The main inferential procedures were conducted under a fully Bayesian perspective. The empirical properties of the Bayesian estimators were assessed through an intensive simulation study, and the performance of the maximum likelihood estimators was also evaluated. The usefulness and competitiveness of the proposed regression model were highlighted using the classical Takeover Bids dataset.

Chapter 5 is based on the paper entitled "A new regression model for overdispersed

and zero-modified count data," firstly submitted for publication in 2019. In this work, we have introduced a new fixed-effects regression model based on the hurdle version of Poisson-Sujatha distribution. All inferential procedures were conducted under a fully Bayesian perspective. An intensive simulation study was performed, so we were able to evaluate the empirical properties of the Bayesian estimators under the specification of different link functions for the parameter that overlaps the probability of observing positive values. The proposed regression model was considered for the analysis of a real dataset on the number of cytogenetic chromosomal aberrations.

Chapter 6 is based on the paper entitled "A new mixed-effects regression model for the analysis of zero-modified hierarchical count data," firstly submitted for publication in 2020. In this work, we have extended the fixed-effects model proposed in Chapter 4 in order to provide a more flexible regression framework, including random-effects. Approximate Bayesian procedures were considered for the task of making inferences from the proposed model. An intensive simulation study was performed, so we were able to evaluate the empirical properties of the Bayesian estimators in some specific situations. The usefulness and competitiveness of the proposed mixed-effects regression model were illustrated using a real dataset on the number of grooming movements practiced by six rats in a neurophysiologic experiment.

Finally, Chapter 7 is devoted to final comments, overall conclusions, and perspectives. This chapter is followed by the bibliography and appendix sections.

CHAPTER

2

A NEW CLASS OF ZERO-MODIFIED POISSON MIXTURE MODELS

2.1 Introduction

In this chapter, we propose a class of zero-modified Poisson mixture models as an alternative to model overdispersed count data exhibiting inflation or deflation of zeros. A relevant feature of this class is that the zero modification can be incorporated using a zero truncation process, and consequently, the proposed models can be expressed in the hurdle version. This procedure leads to the fact that the proposed models can be fitted without any previous information about the zero modification present in a given dataset. A fully Bayesian approach has been considered for estimation and inference concerns. Three different simulation studies have been conducted to illustrate the performance of the developed methodology. The usefulness of the proposed class of models has been assessed by using three real datasets provided by the literature. General model comparison with some well-known discrete distributions has been presented.

More specifically, we introduce a class containing zero-modified versions of the Poisson-Lindley, the Poisson-Shanker, and Poisson-Sujatha distributions, which are themselves overdispersed. The zero-modified models are naturally more flexible than the baseline distributions since they take into account inflation/deflation of zeros, besides modeling data with overdispersion that does not come only from the zero modification in the sample. Also, since the proposed models can handle different levels of overdispersion, they can be naturally considered more flexible and robust than the traditional zero-modified Poisson model as they may fit some points that the latter may not accommodate well.

This chapter is organized as follows. In Section 2.2, we present a class containing the Poisson-Lindley, the Poisson-Shanker, and the Poisson-Sujatha distributions and some of its mathematical properties. In Section 2.3, we present the class of zero-modified versions of these models, demonstrating its flexibility to deal with zero-inflated and zero-deflated data. In Section

2.4, we present the zero-modified distributions as hurdle models. In Section 2.5, the maximum likelihood function, the *prior*, and the *posterior* distributions are derived, and suitable Bayesian estimators are obtained for the unknown model parameters. In Section 2.6, three different simulation studies are presented. In Section 2.7, real applications of the proposed class of models are presented. Concluding remarks are addressed in Section 2.8.

2.2 Poisson mixture models

We are interested in the cases in which the underlying assumption on the Poisson distribution (equidispersion) is violated. For this reason, we give space to the issue of modeling heterogeneous independent Poisson samples, and the Negative Binomial model gives us the necessary motivation. It is well-known that this distribution may arise when it is believed that the rate parameter λ) of a Poisson distribution behaves according to a Gamma random variable on the parameter space $\mathbf{\Lambda} = \mathbb{R}_+$. The interpretation of such a specification is straightforward in the sense that if this law is correctly applied, then modeling independent and heterogeneous Poisson samples can be done by assuming that the parameter λ varies randomly from place to place through the Negative Binomial distribution.

A Gamma specification is quite reasonable for the rate parameter (λ) since its shape flexibility is very useful for the sake of the Negative Binomial model competitiveness. However, this choice gives rise to a biparametric discrete model that may present tricky problems of estimation in some cases (AL-KHASAWNEH, 2010). In this way, we may look to other specifications that provide flexibility when describing parameter λ and that lead, for example, to an uniparametric distribution as a final product. Nonetheless, it will always be preferable to choose a competitive model that has the least number of parameters and, when looking to alternatives for the Poisson distribution, it would be interesting to consider models that have such a characteristic.

Table 1 – Characterization of each element of the class $\mathcal{F}_1$.

Model	$c_1(\theta)$	$m_1(x,\theta)$
$\mathcal{L}$	$(\theta+1)\,\theta^{-2}$	$x+1$
$\mathcal{S}_h$	$\left(\theta^2+1\right)\theta^{-2}$	$x+\theta$
$\mathcal{S}_u$	$\left(\theta^2+\theta+2\right)\theta^{-3}$	x^2+x+1

Source: Elaborated by the author.

In this section, we seek to present a class of uniparametric models that can be seen as alternatives to the standard Poisson distribution on the task of modeling counts in the presence of heterogeneity and that can be competitive with the Negative Binomial distribution. The class we are interested contains the Lindley (LINDLEY, 1958), the Shanker (SHANKER,

2015) and the Sujatha (SHANKER, 2016a) uniparametric distributions as being the possible choices to describe the behavior of the rate parameter (λ) of the Poisson distribution. Despite the large gap of time between the release of these models, they share the feature of having the continuous part of the composition being defined by the mixture of Exponential(θ) and Gamma$(2,\theta)$ distributions, only changing the mixing probabilities for each model. For the Lindley distribution, the mixing probabilities are $\theta(1+\theta)^{-1}$ and $(1+\theta)^{-1}$. For the Shanker distribution, we have $\theta^2(\theta^2+1)^{-1}$ and $(\theta^2+1)^{-1}$. Finally, for the Sujatha distribution, the mixture is defined by the weights $\theta^2(\theta^2+\theta+2)^{-1}$, $\theta(\theta^2+\theta+2)^{-1}$ and $2(\theta^2+\theta+2)^{-1}$. These models are considered competitive for modeling real lifetime in biomedical and engineering contexts and a comprehensive discussion about their mathematical properties such as moments, hazard function, stochastic orderings and parameter estimation can be found in the papers in which these models were introduced.

These three models provide flexibility concerning its shape, which is controlled by a single parameter $\theta \in \mathbf{\Theta} = \mathbb{R}_+$. For a unified approach, let $\mathcal{F}_1 = \{g(\cdot;\theta) : \theta \in \mathbf{\Theta}\}$ be a class of probability density functions that, for a continuous positive random variable X, has elements of the form

$$g(x;\theta) = \frac{m_1(x,\theta)}{c_1(\theta)} e^{-\theta x}, \quad x \in \mathbb{R}_+,$$

where $c_1(\theta) > 0$ is the normalization constant and θ is the shape parameter. Table 1 characterizes the elements of $\mathcal{F}_1$ in terms of the functions $c_1(\theta)$ and $m_1(x,\theta)$. Note that we are restricting $\mathcal{F}_1$ to have only three elements. For simplicity, let us consider that $\mathcal{L}$ states for the Lindley distribution, $\mathcal{S}_h$ states for the Shanker distribution and $\mathcal{S}_u$ states for the Sujatha distribution. Further details on these models can be found in the aforementioned papers where they were first introduced.

Table 2 – Characterization of each element of the class $\mathcal{F}_2$.

Model	$c_2(\theta)$	$m_2(y,\theta)$
$\mathcal{PL}$	$(\theta+1)^3\theta^{-2}$	$(\theta+y+2)(\theta+1)^{-y}$
$\mathcal{PS}_h$	$(\theta+1)^2(\theta^2+1)\theta^{-2}$	$(\theta^2+\theta+y+1)(\theta+1)^{-y}$
$\mathcal{PS}_u$	$(\theta+1)^3(\theta^2+\theta+2)\theta^{-3}$	$[y^2+y(\theta+4)+(\theta^2+3\theta+4)](\theta+1)^{-y}$

Source: Elaborated by the author.

Now, let $\mathcal{F}_2 = \{f(\cdot;\theta) : \theta \in \mathbf{\Theta}\}$ be a class whose elements are Poisson mixture distributions defined from any $g(\cdot;\theta) \in \mathcal{F}_1$. The Poisson-$\mathcal{F}_1$ is a class of probabilistic models that arises when each element of $\mathcal{F}_1$ is chosen to describe the rate parameter (λ) of the Poisson distribution. In fact, the class of uniparametric models in which the mentioned composition leads to a closed-form model is finite and here we will work with a class containing three of them. In this case, a discrete random variable Y is distributed according to an element of $\mathcal{F}_2$ if

the stochastic representation $Y|\lambda \sim f(y;\lambda)$ and $\lambda \sim g(\lambda;\theta) \in \mathcal{F}_1$ holds, where $f(y;\lambda)$ is the conditional probability mass function (pmf) of a Poisson variable.

The unconditional distribution of the random variable Y will be denoted by $f(y;\theta)$. Let $\mathcal{Y}_s = \{s, s+1, \ldots\}$ be the set of the integers greater than or equal to s. The definition is completed by stating that a random variable Y, defined on $\mathcal{Y}_0$, will be distributed according to an element of $\mathcal{F}_2$ if its pmf can be derived as

$$\begin{aligned} f(y;\theta) &= \int_{\Lambda} f(y;\lambda)\, g(\lambda;\theta)\, \mathrm{d}\lambda \\ &= \int_{\Lambda} \frac{\lambda^y e^{-\lambda}}{y!} \frac{m_1(\lambda,\theta)}{c_1(\theta)} e^{-\theta\lambda}\, \mathrm{d}\lambda \\ &= \frac{1}{y!\, c_1(\theta)} \int_{\Lambda} \lambda^y e^{-\lambda(\theta+1)} m_1(\lambda,\theta)\, \mathrm{d}\lambda \\ &= \frac{m_2(y,\theta)}{c_2(\theta)}, \quad y \in \mathcal{Y}_0, \end{aligned}$$

for $\theta \in \boldsymbol{\Theta}$. The fourth equality comes after straightforward algebraic calculations (omitted here) using the gamma integral. Table 2 characterizes the elements of $\mathcal{F}_2$ in terms of the functions $c_2(\theta)$ and $m_2(y,\theta)$. Again, for simplicity, consider that $\mathcal{PL}$ states for the Poisson-Lindley distribution, $\mathcal{PS}_h$ states for the Poisson-Shanker distribution and $\mathcal{PS}_u$ states for the Poisson-Sujatha distribution. Also, the Poisson distribution will be denoted by $\mathcal{P}$ and, when necessary, the Negative Binomial distribution will be denoted by $\mathcal{NB}$ from now on.

The $\mathcal{PL}$ distribution was introduced by Sankaran (1970) and well discussed by Ghitany and Al-Mutairi (2009). More recently, the $\mathcal{PS}_h$ and the $\mathcal{PS}_u$ distributions were introduced and extensively studied by Shanker (2016c) and Shanker (2016d), respectively. From the results provided by these authors, one can generalize one of the most important features concerning the elements of $\mathcal{F}_2$. Let $Y^{(r)} = Y(Y-1)\cdots(Y-r+1)$ the r-order factorial of the random variable Y. If Y is distributed according to an element of $\mathcal{F}_2$, then its r-order factorial moment about the origin is given by

$$\mu'_r = \frac{r!\,(\theta+\tau_1+\tau_2)}{\theta^r\,(\theta+\tau_3)}, \tag{2.1}$$

for $\theta \in \boldsymbol{\Theta}$. The τ_i $(i = 1,2,3)$ is a function of parameter θ but only τ_1 and τ_2 may depend on the order of the factorial moment. The expressions for each τ_i depending on the elements of $\mathcal{F}_2$ are presented in Table 3.

Table 3 – Expressions of τ_1, τ_2 and τ_3 for each element of the class $\mathcal{F}_2$.

Model	τ_1	τ_2	τ_3
$\mathcal{PL}$	r	1	1
$\mathcal{PS}_h$	$r\theta^{-1}$	θ^{-1}	θ^{-1}
$\mathcal{PS}_u$	$r+1$	$(r+1)(r+2)\theta^{-1}$	$1+2\theta^{-1}$

Source: Elaborated by the author.

Using Equation (2.1), one can obtain the moments about origin for each element of $\mathcal{F}_2$ accordingly the values τ_i. Then, in Table 4, we present the mean, the variance, and the index of dispersion of the $\mathcal{PL}$, the $\mathcal{PS}_h$ and the $\mathcal{PS}_u$ distributions. The way these measures are linked may be useful to evaluate how these models might behave when dealing with heterogeneous samples.

Table 4 – Some theoretical measures regarding each element of the class $\mathcal{F}_2$.

Model	Mean (μ)	Variance $\left(\sigma^2\right)$	Index of Dispersion (γ)
$\mathcal{PL}$	$\frac{\theta+2}{\theta(\theta+1)}$	$\frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2}$	$1+\frac{\theta^2+4\theta+2}{\theta(\theta+1)(\theta+2)}$
$\mathcal{PS}_h$	$\frac{\theta^2+2}{\theta(\theta^2+1)}$	$\frac{\theta^5+\theta^4+3\theta^3+4\theta^2+2\theta+2}{\theta^2(\theta^2+1)^2}$	$1+\frac{\theta^4+4\theta^2+2}{\theta\left(\theta^2+1\right)\left(\theta^2+2\right)}$
$\mathcal{PS}_u$	$\frac{\theta^2+2\theta+6}{\theta\left(\theta^2+\theta+2\right)}$	$\frac{\theta^5+4\theta^4+14\theta^3+28\theta^2+24\theta+12}{\theta^2(\theta^2+\theta+2)^2}$	$1+\frac{\theta^4+4\theta^3+18\theta^2+12\theta+12}{\theta\left(\theta^2+\theta+2\right)\left(\theta^2+2\theta+6\right)}$

Source: Elaborated by the author.

Firstly, it can be proved that $\mu \to \infty$ as $\theta \to 0$. Besides, after some straightforward algebra, it is possible to notice that the variances can be written as $\sigma^2 = \mu\,\gamma$, being the ratios involving parameter θ always positive. This implies that models $\mathcal{PL}$, $\mathcal{PS}_h$ and $\mathcal{PS}_u$ can accommodate overdispersion, that is, whichever $\theta \in \boldsymbol{\Theta}$ we have that $\sigma^2 > \mu$. Further, the indexes of dispersion are clearly greater than 1, also implying overdispersion since $\gamma = \sigma^2\mu^{-1}$. On the other hand, we have that $\gamma \to 1$ $(\sigma^2 \to \mu)$ as $\theta \to \infty$, that is, the elements of $\mathcal{F}_2$ have the property of equidispersion for large values of θ (small values of μ).

2.3 Zero-modified models

We are now interested in modeling a high/low amount of zeros observed beyond that generated by the original process, which we already supposed to account for overdispersion. There are some typical situations where zero modification may occur, and we list these cases in the following.

a) Not all members of the population are affected by the process, which causes inflation of zeros to occur due to the response of unaffected subjects being zero;

b) When zeros cannot be observed in the population (truncation at zero);

c) The occurrence of unavoidable problems during the sampling process may lead to an increase/decrease in the probability of a zero observation being selected, hence the zero inflation/deflation situation;

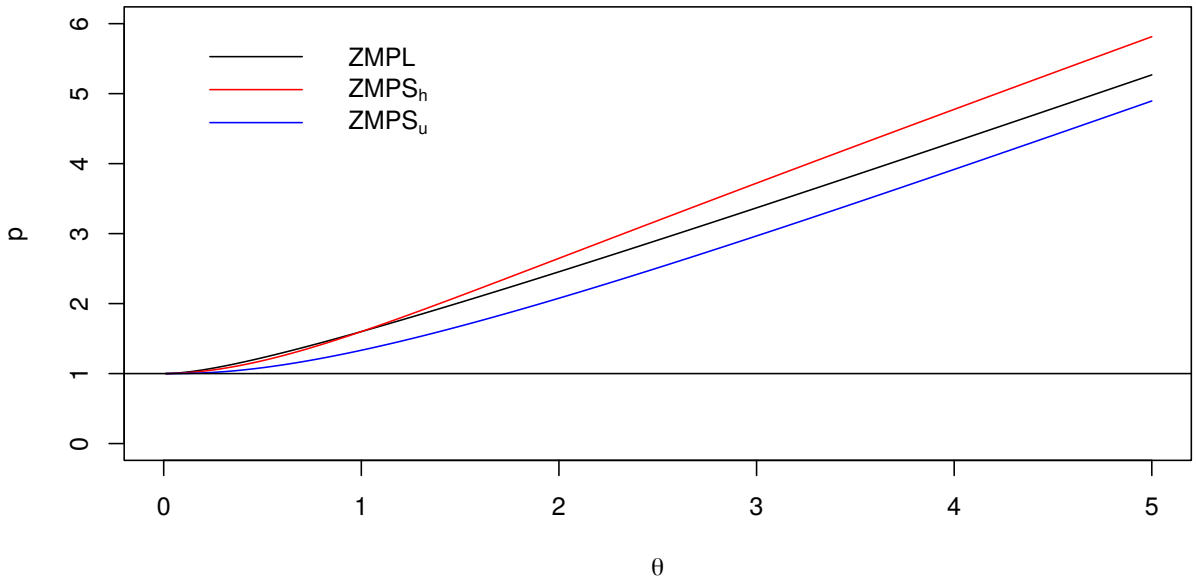
d) A combination of (a) and (b) causes a part of the population to be zero-truncated distributed while the other part is not affected and provides the zero observations.

In this section, we seek to introduce a class of biparametric models that can be seen as alternatives to model overdispersed count datasets when a high/low amount of zeros is observed beyond that which would be expected by the elements of $\mathcal{F}_2$. Thus, let $\mathcal{F}_3 = \{f_*(\cdot;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}^*\}$ be a class of pmfs that, for a discrete random variable Y defined on $\mathcal{Y}_0$, has elements of the form

$$f_*(y;\boldsymbol{\theta}) = (1-p)\,\delta_y + p f(y;\theta), \quad y \in \mathcal{Y}_0, \tag{2.2}$$

for $\boldsymbol{\theta} = (\theta, p) \in \boldsymbol{\Theta}^* = \boldsymbol{\Theta} \times \mathcal{A}_\theta$, where $\mathcal{A}_\theta = [0, [1 - f(0;\theta)]^{-1}]$, being $f(0;\theta)$ the pmf of an element of $\mathcal{F}_2$, evaluated at zero. Further, δ_y is the indicator function, so that $\delta_y = 1$ if $y = 0$ and $\delta_y = 0$ otherwise.

Figure 1 – Supremum of set $\mathcal{A}_\theta$ depending on each element of the class $\mathcal{F}_2$.



Source: Elaborated by the author.

The class $\mathcal{F}_3$ can be referred as the class of zero-modified models which have a set of $f(\cdot;\theta)$ as basis distributions. In our case, this set of pmfs is represented by $\mathcal{F}_2$. Hence, $\mathcal{F}_2 \subseteq \mathcal{F}_3$. In fact, we have three elements on $\mathcal{F}_3$, namely the zero-modified Poisson-Lindley $(\mathcal{ZMPL})$ distribution, the zero-modified Poisson-Shanker $(\mathcal{ZMPS}_h)$ distribution and the zero-modified Poisson-Sujatha $(\mathcal{ZMPS}_u)$ distribution.

From $\mathcal{A}_\theta$, one can notice that model (2.2) is not a mixture distribution typically adopted to fit zero-inflated datasets, since p (zero modification parameter) can assume values greater than 1. However, for all values of $p \in \mathcal{A}_\theta$, Equation (2.2) corresponds to a properly pmf since $f_*(y;\boldsymbol{\theta}) \geqslant 0$ for all $y \in \mathcal{Y}_0$ and $\sum_{y\in\mathcal{Y}_0} f_*(y;\boldsymbol{\theta}) = 1$ whatever the $f(y;\theta) \in \mathcal{F}_2$. Figure 1 presents the upper bounds of $\mathcal{A}_\theta$ accordingly each element of $\mathcal{F}_3$ using selected values for parameter θ.

Note that the higher the value of θ, the higher the value of the upper bound of $\mathcal{A}_\theta$. In addition, it is clear that the upper bounds under the $\mathcal{ZMPS}_u$ distribution are always lower than those obtained under the $\mathcal{ZMPL}$ and $\mathcal{ZMPS}_h$ distributions.

If a random variable Y is distributed according to an element of $\mathcal{F}_3$, then its mean, its variance and its index of dispersion are given, respectively, by

$$\mu^* = p\mu, \qquad \left(\sigma^2\right)^* = p\left[\sigma^2 + (1-p)\mu^2\right] \qquad \text{and} \qquad \gamma^* = \gamma + (1-p)\,\mu, \tag{2.3}$$

where the measures μ, σ^2 and γ are presented in Table 4 according to each possible choice in $\mathcal{F}_2$. The term $(1-p)\mu$ in γ^* represents the overdispersion caused by a modification on the zero frequency, regarding the elements of $\mathcal{F}_2$.

The elements of $\mathcal{F}_3$ may be considered as interesting alternatives to the usual zero-modified Poisson ($\mathcal{ZMP}$) model since each element of $\mathcal{F}_2$ can accommodate several levels of overdispersion, an issue that the $\mathcal{P}$ distribution generally fails to handle. In this context, the parameter p plays a major role in controlling the frequency of zeros, and it has a natural interpretation in terms of the proportions of either inflation or deflation at zero. The following statements describe the effect of parameter p on Equation (2.2).

i) If $p = 0$ then $f_*(0;\boldsymbol{\theta}) = 1$, $\mathcal{F}_3$ contains only degenerate distributions with all mass at zero;

ii) If $p = 1$ then $f_*(0;\boldsymbol{\theta}) = f(0;\theta)$, the elements of $\mathcal{F}_3$ has a proportion of zeros equal than the elements of $\mathcal{F}_2$. In this case, we can also conclude that $\mathcal{F}_3 = \mathcal{F}_2$ since for all $y \in \mathcal{Y}_0$, $f_*(y;\boldsymbol{\theta}) = f(y;\theta)$;

iii) If $p \in (0,1)$ then $f_*(0;\boldsymbol{\theta}) > f(0;\theta)$, the elements of $\mathcal{F}_3$ has a proportion of zeros greater than the elements of $\mathcal{F}_2$;

iv) If $p \in (1, \sup_{\theta\in\boldsymbol{\Theta}} \mathcal{A}_\theta)$ then $f_*(0;\boldsymbol{\theta}) < f(0;\theta)$, the elements of $\mathcal{F}_3$ has a proportion of zeros lower than the elements of $\mathcal{F}_2$;

v) If $p = \sup_{\theta\in\boldsymbol{\Theta}} \mathcal{A}_\theta$ then $f_*(0;\boldsymbol{\theta}) = 0$, the Equation (2.2) corresponds to the zero-truncated version of a particular element of $\mathcal{F}_2$, with general pmf given by

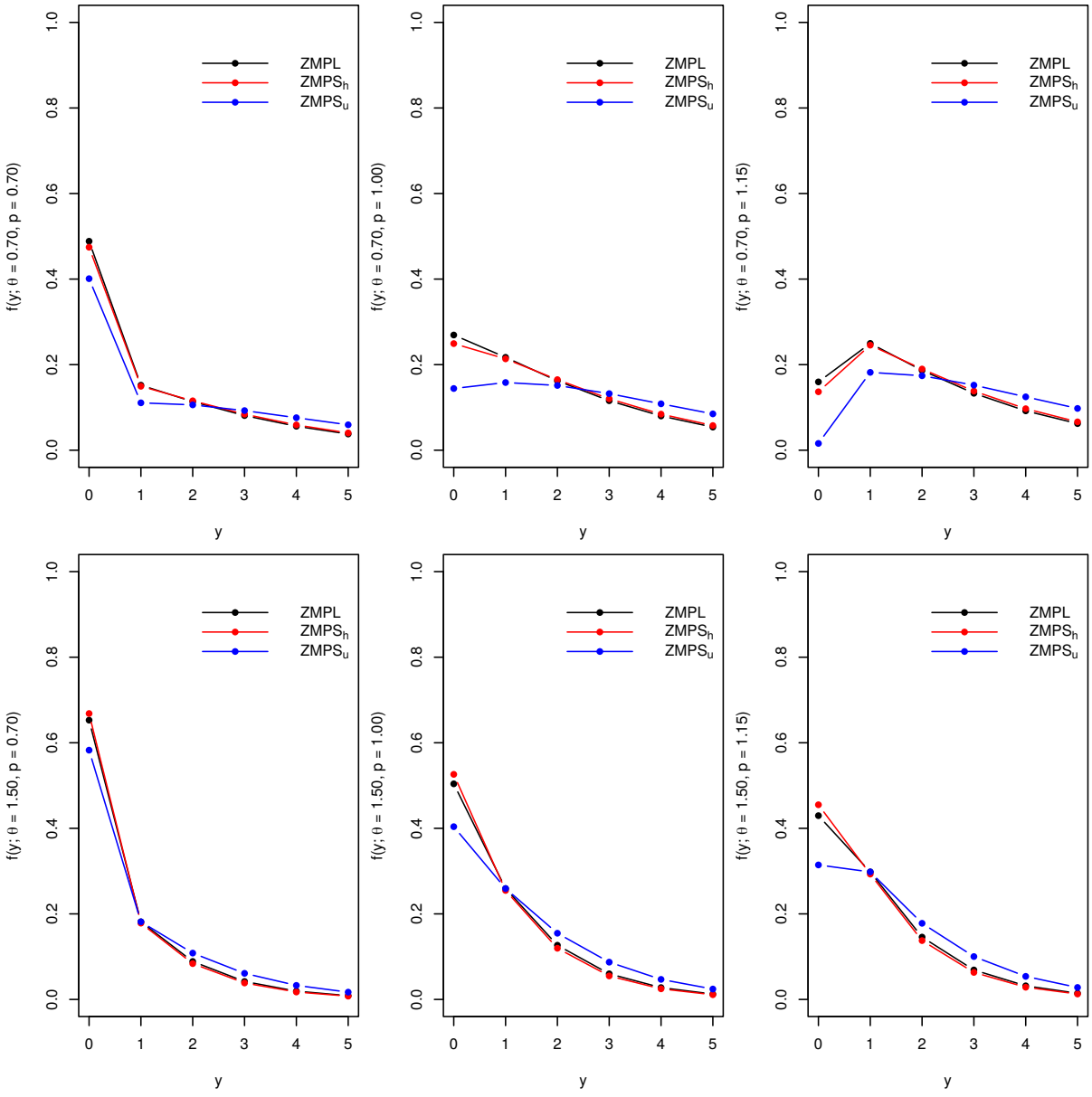
$$f^*(y;\theta) = \frac{f(y;\theta)}{1 - f(0;\theta)}, \quad y \in \mathcal{Y}_1, \tag{2.4}$$

for $\theta \in \boldsymbol{\Theta}$, $f(\cdot;\theta) \in \mathcal{F}_2$. See Ghitany, Al-Mutairi and Nadarajah (2008), Shanker and Fesshaye (2016) and Shanker (2017) for further details about the zero-truncated version of the $\mathcal{PL}$, the $\mathcal{PS}_h$ and the $\mathcal{PS}_u$ distributions.

The above statements can be easily checked through the proportion of additional/missing zeros, defined by

$$f_*(0;\boldsymbol{\theta}) - f(0;\theta) = (1-p) + pf(0;\theta) - f(0;\theta) = (1-p)\left[1 - f(0;\theta)\right].$$

Figure 2 – Behavior of each element of the class $\mathcal{F}_3$ for different values of θ and p (upper-panels: $\theta = 0.70$ and lower-panels: $\theta = 1.50$; upper/lower-left-panel: $p = 0.70$; upper/lower-middle-panel: $p = 1.00$ and upper/lower-right-panel: $p = 1.15$).



Source: Elaborated by the author.

Follows from the previous expression that (i) and (ii) are straightforward. As for (v),

$$f_*(0;\boldsymbol{\theta}) - f(0;\theta) = \left\{1 - [1 - f(0;\theta)]^{-1}\right\}[1 - f(0;\theta)] = -f(0;\theta),$$

hence $f_*(0;\boldsymbol{\theta}) = 0$. Statement (iii) follows from the fact that if $p \in (0,1)$ then $(1-p)[1 - f(0;\theta)] \in (0,1)$ since f is a probability measure. Therefore $f_*(0;\boldsymbol{\theta}) > f(0;\theta)$. For statement (iv), whichever $p > 1$, $(1-p) < 0$ and the result follows by the same argument for (iii). Hence, $f_*(0;\boldsymbol{\theta}) < f(0;\theta)$.

The case (iii) may be appropriate in situations (a), (c), and (d) as described at the

beginning of this section, and the case (iv) may be appropriate in situations (c) and (d). The index of dispersion γ^* can be analyzed in terms of the modification at zero in a given sample since, from Equation (2.3), one can conclude that $\gamma^* = \gamma$ in the standard case (ii), $\gamma^* > \gamma$ in the zero-inflated case (iii) and $\gamma^* < \gamma$ in the zero-deflated case (iv).

Table 5 – Value of some theoretical measures regarding each element of the class $\mathcal{F}_3$.

Model	Parameters		Measures			
	θ	p	μ^*	$(\sigma^2)^*$	γ^*	y_0^*
$\mathcal{ZMPL}$	0.70	0.70	1.59	5.28	3.32	0
	0.70	1.00	2.27	6.00	2.65	0
	0.70	1.15	2.61	6.02	2.31	1
	1.50	0.70	0.65	1.35	2.06	0
	1.50	1.00	0.93	1.66	1.78	0
	1.50	1.15	1.07	1.76	1.64	0
$\mathcal{ZMPS}_h$	0.70	0.70	1.67	5.57	3.33	0
	0.70	1.00	2.39	6.25	2.62	0
	0.70	1.15	2.75	6.20	2.26	1
	1.50	0.70	0.61	1.24	2.04	0
	1.50	1.00	0.87	1.55	1.77	0
	1.50	1.15	1.00	1.65	1.65	0
$\mathcal{ZMPS}_u$	0.70	0.70	2.47	9.42	3.81	0
	0.70	1.00	3.53	9.72	2.75	1
	0.70	1.15	4.06	9.02	2.22	1
	1.50	0.70	0.91	2.11	2.31	0
	1.50	1.00	1.30	2.50	1.92	0
	1.50	1.15	1.50	2.58	1.72	0

Source: Elaborated by the author.

Figure 2 depicts the pmf of each element of $\mathcal{F}_3$ for $\theta = 0.70$ and $\theta = 1.50$. To illustrate the behaviour of each model in terms of the modification at zero, we have considered $p = 0.70$ (zero-inflated case), $p = 1.00$ (standard case) and $p = 1.15$ (zero-deflated case). The case $\theta = 0.7$ and $p = 1.15$ (upper-right-panel) differs from the others only by the fact that, under such configuration, the mode (y_0^*) of the proposed models is equal to 1. This is expected in some situations where $p \geqslant 1$ and will never occur in zero-inflated cases. In Table 5, we present the values of the main theoretical measures (mean, variance, index of dispersion and mode) associated with the elements of $\mathcal{F}_3$ for the fixed set of parameters. One can notice that different values of p lead to different zero-modified distributions. For example, one can see that the higher the value of p, the lower the probability of observing zeros under the elements of $\mathcal{F}_3$. On the other hand, we can also note that Figure 2 highlight the similar behavior between the proposed zero-modified models, which is expected due to the functional form of the elements of $\mathcal{F}_2$. This

is not a problem at all since we are not interested in these models to be competitive with each other. Also, when dealing with real datasets, all these models can be implemented and estimated under the same computational complexity, and, according to an appropriate selection criterion, the best model can be chosen.

2.4 Hurdle models

The class of hurdle models was introduced and discussed by Mullahy (1986). The relevant feature of such models is the fact that the zero observations are treated separately from the positive ones. In the central formulation, a binary probability model determines whether a zero or a non-zero outcome occurs, and hence, an appropriated zero-truncated discrete distribution is chosen to describe the positive ones (SAFFARI; ADNAN; GREENE, 2012). On the other hand, this framework can also be considered in the cases where it is observed an incidence of zero-valued subjects within continuous positive samples.

In this section, we seek to introduce the hurdle version of the elements of $\mathcal{F}_2$. Firstly, one can notice that pmf (2.2) can be written as

$$\begin{aligned} f_*(y;\boldsymbol{\theta}) &= [1-p+pf(0;\theta)]\,\delta_y + pf(y;\theta)(1-\delta_y) \\ &= (1-\omega)\,\delta_y + \omega f^*(y;\theta) \quad y \in \mathcal{Y}_0, \end{aligned} \tag{2.5}$$

where $\omega = p[1-f(0;\theta)]$, $\boldsymbol{\theta}^* = (\theta,\omega)$ and $f^*(y;\theta)$ is the zero-truncated version of particular element of $\mathcal{F}_2$ with general pmf given by (2.4). Since $p \in \mathcal{A}_\theta$ then $\omega \in [0,1]$. The pmf (2.5) can be seen as a hurdle version of the elements of $\mathcal{F}_2$, where the probability of $Y=0$ is $(1-\omega)$ and the probability of $Y>0$ is $\omega\sum_{y\in\mathcal{Y}_1} f^*(y;\theta) = \omega$. It is worthwhile to mention that the hurdle configuration does not affect the general definition of the class $\mathcal{F}_3$.

The elements of $\mathcal{F}_3$, expressed through hurdle models, contain its corresponding zero-truncated versions as one of its components, which differs from the traditional mixture representation of zero-inflated distributions. Moreover, this representation can be interpreted as a superposition of two processes, that is, one that produces positive observations from a zero-truncated distribution and another that produces only zero-valued observations with probability $(1-\omega)$. Therefore, Equation (2.5) cannot be considered a 2-component mixture model.

By the hurdle representation of zero-modified models, only the positive observations are required to estimate parameter θ. This fact is well discussed by Conceição *et al.* (2017a) for the class of zero-modified Power Series distributions, and we extend this result by asserting that the zero-truncated version of a particular element of $\mathcal{F}_3$ is always a zero-truncated version of the correspondent element on $\mathcal{F}_2$ and they have the same parameter θ. For example, the zero-truncated version of the $\mathcal{ZMPL}$ distribution is equivalent to the zero-truncated Poisson-Lindley distribution for all $\theta \in \boldsymbol{\Theta}$. This can be easily checked using Equation (2.5) because, since the probability of Y being positive is ω, if we exclude the value zero from the domain and divide the

right-hand side of (2.5) by ω, then we get $Y \sim f^*(y;\theta)$. On the other hand, such representation allows us to obtain a closed-form solution for the maximum likelihood estimator (MLE) of parameter ω, which is given by the proportion of non-zeros in the dataset (see Appendix A, Section A.1). Also, it is easy to see that equation $s(\theta,\omega) = \omega[1 - f(0;\theta)]^{-1}$ is injective (one-to-one) into $\boldsymbol{\Theta} \times [0,1]$ and therefore, the invariance principle ensures that parameter p can be estimated using such an equation.

Under a Bayesian perspective, it will be shown in the next section that, for an appropriate *prior* specification, the *posterior* distribution of ω has known closed-form. Besides, inference procedures about parameter p are also required since we are often interested in identifying the kind of zero modification (inflation or deflation) is present in a given count dataset.

2.5 Inference

2.5.1 Maximum likelihood

Let $\boldsymbol{Y} = (Y_1,\ldots,Y_n)$ be a random sample of size n taken from any element of $\mathcal{F}_3$ and $\boldsymbol{y} = (y_1,\ldots,y_n)$ its observed values. Considering model (2.5), the likelihood function of vector $\boldsymbol{\theta}^*$ is given by

$$\mathcal{L}(\boldsymbol{\theta}^*;\boldsymbol{y}) = (1-\omega)^{n_0}\,\omega^{n_+}\,\frac{\prod_{i=1}^{n}[f(y_i;\theta)]^{1-\delta_{y_i}}}{[1-f(0;\theta)]^{n_+}}, \tag{2.6}$$

where $n_+ = n - n_0$, being n_0 the number of zero-valued observations in the sample and $f(\cdot;\theta) \in \mathcal{F}_2$. Now, the corresponding log-likelihood function is given by

$$\ell(\boldsymbol{\theta}^*;\boldsymbol{y}) = n\log(\omega) - n_0\log\left(\frac{\omega}{1-\omega}\right) - n_+\log[1-f(0;\theta)] + \sum_{i=1}^{n}(1-\delta_{y_i})\log[f(y_i;\theta)].$$

One can notice that the hurdle configuration leads to orthogonality between parameters θ and ω. Also, it is easy to see that all terms in the log-likelihood function depending on θ take into account only the positive values of the sample vector $\boldsymbol{y}$. Denoting by $\boldsymbol{y}^+$ the vector of positive observations from $\boldsymbol{y}$, $\{y_j^+, j = 1,\ldots,n_+\}$, the log-likelihood function of parameter θ, based on the assumption that y_j^+ comes from the zero-truncated version of a particular element of $\mathcal{F}_2$, is given by

$$\ell_n(\theta;\boldsymbol{y}^+) = -n_+\log[1-f(0;\theta)] + \sum_{j=1}^{n_+}\log\left[f\left(y_j^+;\theta\right)\right]. \tag{2.7}$$

Indeed $\ell_n(\theta;\boldsymbol{y}^+) = \ell_{n_+}(\theta;\boldsymbol{y})$, since each y_j present in (2.7) is generated by a zero-truncated distribution. Here, we are extending the fact that estimating the $\mathcal{P}$ parameter θ using the zero-truncated Poisson ($\mathcal{ZTP}$) distribution results in a loss of efficiency in the inference if there is no zero modification (DIETZ; BÖHNING, 2000; CONCEIÇÃO; ANDRADE; LOUZADA,

2014). Denoting by $\boldsymbol{y}^0$ the vector of zero observations from $\boldsymbol{y}$, $\{y_j^0, j = 1,\ldots,n_0\}$, the log-likelihood function of ω is given by

$$\ell_n\left(\omega;\boldsymbol{y}^0\right) = \ell_n\left(\omega;\boldsymbol{y}\right) = n\log\left(\omega\right) - n_0\log\left(\frac{\omega}{1-\omega}\right).$$

Since our approach relies on the Bayesian inference, the technical details about maximum likelihood estimation of parameters θ and ω were placed in Appendix A (Section A.1).

2.5.2 Prior distributions

We consider the typical case where no specialized information is available to justify the choice of an informative *prior* for the set of unknown model parameters. In other words, we specify a *prior* distribution such that, even for moderate sample sizes, the information provided by the data should dominate the *prior* information due to the vague nature of the *prior* knowledge. This can be done with noninformative *priors* such as *Jeffreys priors* (JEFFREYS, 1946). The *Jeffreys* and the reference *priors* are equivalent in 1-dimensional problems (SYVERSVEEN, 1998) and are optimal objective *priors* (BERGER; BERNARDO; SUN, 2015).

Formally, the *Jeffreys prior* for the vector $\boldsymbol{\theta}^*$ is defined as $\pi_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}^*) \propto \sqrt{\det \mathcal{I}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}^*)}$, where $\mathcal{I}_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}^*)$ stands for the Fisher information matrix of vector $\boldsymbol{\theta}^*$. However, due to the orthogonality between θ and ω, one can derive independent *priors* for the elements of $\boldsymbol{\theta}^*$. By considering Equation (A.1), the *Jeffreys prior* for parameter ω can be expressed as

$$\pi_\omega\left(\omega\right) \propto \omega^{-1/2}\left(1-\omega\right)^{-1/2}, \tag{2.8}$$

which is the *kernel* of a Beta density with hyperparameters equal to $1/2$. Now, the *Jeffreys prior* for parameter θ has the general form $\pi_\theta(\theta) \propto \sqrt{\mathcal{I}_\theta(\theta)}$, which clearly depends on the selected element of $\mathcal{F}_3$. For example, under the $\mathcal{ZMPL}$ distribution, the *prior* of θ takes the form

$$\pi_\theta\left(\theta\right) \propto \frac{1}{\sqrt{\theta^2+3\theta+1}}\left\{\theta^2\left[\left(\theta+1\right)^2\zeta\left[\left(\theta+1\right)^{-1},1,\theta\right] - \frac{\theta^3+5\theta^2+8\theta+2}{\theta\left(\theta+2\right)}\right] - \frac{\theta^4-\theta^3-8\theta^2-10\theta-2}{\theta^2\left(\theta^2+3\theta+1\right)}\right\}^{1/2},$$

where ζ stands for the Lerch-Phi function defined as $\zeta(z,a,v) = \sum_{k=0}^{\infty} z^k(k+v)^{-a}$ for $|z| < 1$. Further details on this function can be found at Bateman and Erdélyi (1953).

One of the most appealing properties of the *Jeffreys prior* is its invariance under reparameterizations. In this way, let us redefine the pmf (2.5) with $\theta = e^\eta$, $\eta \in \mathbb{R}$. This is particularly useful since it allows us to work in an unrestricted parametric space. In this case, the *Jeffreys prior* for the transformation $\eta = \log(\theta)$ can be derived as

$$\pi_\eta\left(\eta\right) \propto \pi_\theta\left(e^\eta\right)\left|e^\eta\right| = e^\eta\sqrt{\mathcal{I}_\theta\left(e^\eta\right)}. \tag{2.9}$$

2.5.3 Posterior distributions and estimation

A Bayesian approach for model (2.5) can be considered by writing the unnormalized joint *posterior* distribution of vector (η, ω) as

$$\pi(\eta, \omega; \boldsymbol{y}) \propto \exp\{\ell_{n_+}(\eta; \boldsymbol{y}) + \ell_n(\omega; \boldsymbol{y})\}\, \pi_\eta(\eta)\, \pi_\omega(\omega).$$

From this point of view, inference for the model parameters is wholly based on their marginal *posterior* distributions, which can be obtained by integrating the joint *posterior* distribution. In our case, however, we do not need to worry about the integration issue since η and ω are orthogonal. Therefore, the marginal *posterior* distributions take the following form

$$\pi_\eta(\eta; \boldsymbol{y}) \propto \exp\{\ell_{n_+}(\eta; \boldsymbol{y})\}\, \pi_\eta(\eta) \quad \text{and} \quad \pi_\omega(\omega; \boldsymbol{y}) \propto \exp\{\ell_n(\omega; \boldsymbol{y})\}\, \pi_\omega(\omega),$$

where

$$\pi_\omega(\omega; \boldsymbol{y}) \propto \omega^{(n_+ + 1/2) - 1} (1 - \omega)^{(n_0 + 1/2) - 1}, \tag{2.10}$$

which is the *kernel* of a Beta density with parameters $n_+ + 1/2$ and $n_0 + 1/2$. In this setting, any suitable Bayesian point estimator for parameter ω can be derived, regardless the elements of $\mathcal{F}_3$. For instance, the Bayesian estimator for ω with respect to quadratic loss is given by

$$\mathbb{E}(\omega; \boldsymbol{y}) = \frac{2n_+ + 1}{2(n+1)}. \tag{2.11}$$

Alternatively, one can obtain Bayesian estimators with respect to *modular* and *zero-one* losses for parameter ω. These estimators are obtained by computing, respectively, the conditional median and mode of the *posterior* distribution (2.10). Hence,

$$\mathbb{M}_d(\omega; \boldsymbol{y}) \approx \frac{6n_+ + 1}{2(3n+1)} \quad \text{and} \quad \mathbb{M}_o(\omega; \boldsymbol{y}) = \frac{2n_+ - 1}{2(n-1)},$$

for $n > n_+$. The approximation of the conditional median is due to the results provided by Kerman (2011). For $n_+ > 1$, the relative error of the approximation is less than 1% and quickly goes to zero as n_0 and n_+ increase. Besides, the *posterior* conditional variance of ω is given by

$$\mathbb{V}(\omega; \boldsymbol{y}) = \frac{(2n_+ + 1)(2n_0 + 1)}{4(n+1)^2(n+2)}.$$

Now, let the normalized marginal *posterior* distribution of parameter η be denoted by $\pi^*_\eta(\eta; \boldsymbol{y}) = B^{-1} \exp\{\ell_{n_+}(\eta; \boldsymbol{y})\}\, \pi_\eta(\eta)$, $B = \int_{\mathbb{R}} \exp\{\ell_{n_+}(\eta; \boldsymbol{y})\}\, \pi_\eta(\eta)\, \mathrm{d}\eta$. Notably, whichever the chosen element of $\mathcal{F}_3$, the integral B is analytically intractable over the unbounded convex set $\mathbb{R}$ and therefore, one can verify only numerically if the resulted *posterior* distribution is proper ($B < \infty$). We have considered the Laplace method (SMALL, 2010) for this task and we anticipate that no problems regarding the properness of $\pi^*_\eta(\eta; \boldsymbol{y})$ were found in the investigated cases.

To address the problem of generating pseudo-random values from $\pi^*_\eta(\eta; \boldsymbol{y})$ and make inference about parameter θ, we consider a rejection sampling scheme (VON NEUMANN,

1951; ROBERT; CASELLA, 2013). To implement the algorithm, one has to choose a candidate-generating distribution $\varphi_\eta(\eta)$, whose values are accepted or rejected accordingly to a prespecified rule. This rule is based on the choice of a constant $c > 1$ satisfying $\pi^*_\eta(\eta;\boldsymbol{y}) < c\,\varphi_\eta(\eta)$. Then, a generated value η^* from $\varphi_\eta(\eta)$ is accepted as being a drawn from the target density $\pi^*_\eta(\eta;\boldsymbol{y})$ iff $u < \pi^*_\eta(\eta^*;\boldsymbol{y})[c\,\varphi_\eta(\eta^*)]^{-1}$, where u is generated from a Uniform distribution defined on the unit interval.

In the Rejection Sampling algorithm, a candidate η^* is accepted with probability c^{-1}, which means that approximately one value is accepted out of c generated ones. The efficiency of the algorithm is closely related to the value c in the sense that, the lower this value, the more efficient is the algorithm. Therefore, the method builds an empirical distribution that will quickly converge to $\pi^*_\eta(\eta;\boldsymbol{y})$ if c is selected as being equal to $\pi^*_\eta(\alpha;\boldsymbol{y})[\varphi_\eta(\alpha)]^{-1}$ for $\alpha = \arg\max_{\eta\in\mathbb{R}}\{\pi_\eta(\eta;\boldsymbol{y})[\varphi_\eta(\eta)]^{-1}\}$. Here, we consider $\varphi_\eta(\eta) = \mathcal{N}(\tilde{\eta}, \nu)$, where $\tilde{\eta} = \arg\max_{\eta\in\mathbb{R}} \pi_\eta(\eta;\boldsymbol{y})$ and the tuning parameter ν is chosen in such a way that the acceptance rate is the highest possible, which indicates that $c\,\varphi_\eta(\eta)$ is a good envelope for the target density $\pi^*_\eta(\eta;\boldsymbol{y})$.

Algorithm 1 – Rejection Sampling

procedure REJSAM(M, ν)
  Set $k \leftarrow 1$ and $\tilde{\eta} \leftarrow \arg\max_{\eta\in\mathbb{R}} \pi_\eta(\eta;\boldsymbol{y})$
  Set $\alpha \leftarrow \arg\max_{\eta\in\mathbb{R}}\{\pi_\eta(\eta;\boldsymbol{y})[\varphi_\eta(\eta)]^{-1}\}$ and $c \leftarrow \pi^*_\eta(\alpha;\boldsymbol{y})[\varphi_\eta(\alpha)]^{-1}$
  while $k \leqslant M$ **do**
    Set $i \leftarrow 0$
    while $i = 0$ **do**
      Generate $\psi \sim \mathcal{N}(\tilde{\eta}, \nu)$ and $u \sim \mathcal{U}(0,1)$
      if $\log(u) < \log\left[\pi^*_\eta(\psi;\boldsymbol{y})\right] - \log\left[\varphi_\eta(\psi)\right] - \log(c)$ **then**
        Set $\eta^{(k)} \leftarrow \psi$ and $i \leftarrow 1$
      end if
    end while
    Set $k \leftarrow k + 1$
  end while
  return $\{\eta^{(k)}\}_{k=1}^{M}$
end procedure

Algorithm 1 can be used to generate pseudo-random values from $\pi^*_\eta(\eta;\boldsymbol{y})$ using a rejection sampling scheme. To run the algorithm, the requested sample size M, and the tuning parameter ν must be prespecified. After that, a summary of the marginal *posterior* distribution can be obtained. Recall that we are interested in $\theta = e^\eta$. Therefore, the Bayesian estimator for parameter θ with respect to quadratic loss is given by the *posterior* conditional mean, which can be approximated by the Monte Carlo (MC) estimator

$$\mathbb{E}(\theta;\boldsymbol{y}) = \frac{1}{M}\sum_{j=1}^{M} e^{\eta^{(j)}}, \tag{2.12}$$

and the *posterior* conditional variance of θ can be estimated by

$$\mathbb{V}(\theta;\boldsymbol{y}) = \frac{1}{M}\sum_{j=1}^{M}\left[e^{\eta^{(j)}} - \mathbb{E}(\theta;\boldsymbol{y})\right]^2. \tag{2.13}$$

Finally, the Bayesian estimator for the zero modification parameter (p) with respect to quadratic loss is given by

$$\mathbb{E}(p;\boldsymbol{y}) = \frac{1}{M}\sum_{j=1}^{M}\frac{\omega^{(j)}}{1-f(0,e^{\eta^{(j)}})}, \tag{2.14}$$

where $\omega^{(j)}$ is a pseudo-random value drawn from the Beta distribution (2.10). The *posterior* conditional variance of p can be computed analogously to (2.13).

In the next section, we present the results obtained in three different simulation studies involving the class of proposed zero-modified models. The application of such models using real datasets is illustrated in Section 2.7. The Bayesian procedures were conducted by considering the *Jeffreys priors* (2.8) and (2.9). To run Algorithm 1, we have kept fixed $M = 5{,}000$ and $\nu = 1.0$, which have always provided acceptance rates greater than 95%. In this setting, the requested computational time to generate 5,000 pseudo-random values from the marginal *posterior* distribution of parameter η was always less than 1 minute in a machine equipped with an Intel® Core™ i7-6500U CPU at 2.50GHz, with 8 GB DDR3 SDRAM. All computations were performed using the `R` environment (R Development Core Team, 2017).

2.5.4 Posterior predictive distribution

Under a Bayesian approach, one can obtain the *posterior* predictive distribution (ppd) as being the distribution of possible unobserved values conditioned on the observed ones. In our case, the pmf of any observation $y^* \in \mathcal{Y}_0$ can be obtained as

$$\begin{aligned}\pi(y^*;\boldsymbol{y}) &= \int_{\mathbb{R}}\int_0^1 f_*(y^*;\eta,\omega)\,\pi(\eta,\omega;\boldsymbol{y})\,\mathrm{d}\omega\,\mathrm{d}\eta \\ &= \frac{\mathcal{B}e\left(n_+ - \delta_{y^*} + 3/2, n_0 + \delta_{y^*} + 1/2\right)}{\mathcal{B}e\left(n_+ + 1/2, n_0 + 1/2\right)}\int_{\mathbb{R}}\left[\frac{f(y^*;\eta)}{1-f(0;\eta)}\right]^{1-\delta_{y^*}}\pi(\eta;\boldsymbol{y})\,\mathrm{d}\eta,\end{aligned}$$

where $\mathcal{B}e$ is the Beta function and $f(\cdot;\eta) \in \mathcal{F}_2$. One can notice that the ppd has no closed-form and therefore, an MC estimator for such quantity can be obtained by using pseudo-random generated values from $\pi^*_\eta(\eta;\boldsymbol{y})$. Hence,

$$\hat{\pi}(y^*;\boldsymbol{y}) = \frac{\mathcal{B}e\left(n_+ - \delta_{y^*} + 3/2, n_0 + \delta_{y^*} + 1/2\right)}{M\mathcal{B}e\left(n_+ + 1/2, n_0 + 1/2\right)}\sum_{j=1}^{M}\left\{\frac{f\left(y^*;\eta^{(j)}\right)}{1-f\left(0;\eta^{(j)}\right)}\right\}^{1-\delta_{y^*}}, \tag{2.15}$$

from which one can estimate the *posterior* probability of $Y = 0$ as

$$\hat{\pi}(0;\boldsymbol{y}) = \frac{\mathcal{B}e\left(n_+ + 1/2, n_0 + 3/2\right)}{\mathcal{B}e\left(n_+ + 1/2, n_0 + 1/2\right)}.$$

2.6 Simulation studies

The empirical properties of an estimation procedure can be evaluated through Monte Carlo simulations. We have performed three different simulation studies aiming to validate the proposed Bayesian approach. The first simulation study was conducted to assess the properties of the Bayesian estimators of parameters θ, ω, and p. Regarding the Bayesian comparison criteria (see Appendix A, Section A.2), the second simulation study was developed to estimate the probability of correct selection of the proposed distributions compared with the standard $\mathcal{ZMP}$ model when the true one is known. Finally, the third simulation study was performed to evaluate the loss of efficiency when considering only the positive observations in a given sample to estimate parameter θ, assuming that there is no zero modification. All the results obtained in each simulation study are available in Appendix A (Section A.3).

Table 6 – Actual parameter values for simulation of zero-modified artificial datasets.

Case	Scenario	Model	θ	p	ω	μ^*	y_0^*
Zero Inflation	1	$\mathcal{ZMPL}$	0.50	0.20	0.163	0.67	0
		$\mathcal{ZMPS}_h$	0.50	0.20	0.169	0.72	0
		$\mathcal{ZMPS}_u$	0.50	0.20	0.185	1.05	0
	2	$\mathcal{ZMPL}$	0.50	0.80	0.652	2.67	0
		$\mathcal{ZMPS}_h$	0.50	0.80	0.676	2.88	0
		$\mathcal{ZMPS}_u$	0.50	0.80	0.738	4.22	0
	3	$\mathcal{ZMPL}$	5.00	0.20	0.038	0.05	0
		$\mathcal{ZMPS}_h$	5.00	0.20	0.034	0.04	0
		$\mathcal{ZMPS}_u$	5.00	0.20	0.041	0.05	0
	4	$\mathcal{ZMPL}$	5.00	0.80	0.152	0.19	0
		$\mathcal{ZMPS}_h$	5.00	0.80	0.138	0.17	0
		$\mathcal{ZMPS}_u$	5.00	0.80	0.163	0.20	0
Zero Deflation	1	$\mathcal{ZMPL}$	2.50	1.40	0.482	0.72	0
		$\mathcal{ZMPS}_h$	2.50	1.40	0.439	0.64	0
		$\mathcal{ZMPS}_u$	2.50	1.40	0.558	0.90	0
	2	$\mathcal{ZMPL}$	2.50	2.40	0.826	1.23	1
		$\mathcal{ZMPS}_h$	2.50	2.40	0.753	1.09	1
		$\mathcal{ZMPS}_u$	2.50	2.40	0.956	1.54	1
	3	$\mathcal{ZMPL}$	6.00	1.40	0.225	0.27	0
		$\mathcal{ZMPS}_h$	6.00	1.40	0.205	0.24	0
		$\mathcal{ZMPS}_u$	6.00	1.40	0.238	0.29	0
	4	$\mathcal{ZMPL}$	6.00	2.40	0.385	0.46	0
		$\mathcal{ZMPS}_h$	6.00	2.40	0.351	0.41	0
		$\mathcal{ZMPS}_u$	6.00	2.40	0.408	0.49	0

Source: Elaborated by the author.

Table 6 presents the values of parameters θ and p that were considered for simulation in Studies 1 and 2. The respective mean (μ^*) and mode (y_0^*) of the zero-modified models are also presented. The value of ω in each scenario is obtained as $p[1 - f(0;\theta)]$. In order to evaluate

the behavior of the proposed methodology, we have considered quite different zero-inflated and zero-deflated situations. For the zero-inflated (zero-deflated) case, the samples were generated considering that $p \in (0,1)$ ($p \in (1, \sup_{\theta \in \boldsymbol{\Theta}} \mathcal{A}_\theta)$). Hence, the parameters were chosen by taking into account that zero-inflated (zero-deflated) samples naturally have the proportion of zeros higher (lower) than expected, and therefore, the variable Y has to be generated with θ close to zero (not even close to zero). For example, one can notice that the probability of generating a zero-valued observation is approximately 95% in the third scenario (zero-inflated case) and varies between 5% and 25% in the second scenario (zero-deflated case).

Algorithm 2 – Sequential-Search

1: **procedure** SEQSEA($\boldsymbol{\theta}^*$)
2: Generate $u \sim \mathcal{U}(0,1)$
3: Set $y \leftarrow 0$ and $k \leftarrow f_*(0;\boldsymbol{\theta}^*)$
4: **while** $u > k$ **do**
5: Set $y \leftarrow y+1$ and $k \leftarrow k + \omega f^*(y;\mu)$
6: **end while**
7: **return** y
8: **end procedure**

Algorithm 2 can be used to generate a single value from a discrete random variable distributed according to any element of $\mathcal{F}_3$, considering the hurdle version of the baseline distributions. The process to generate a random sample consists of running the algorithm as often as necessary, say n times. The sequential-search is a black-box type of algorithm and works with any computable probability vector. The main advantage of such a procedure is its simplicity. On the other hand, sequential-search algorithms may be slow as the while-loop may have to be repeated very often. More information on this algorithm can be found at Hörmann, Leydold and Derflinger (2013). For instance, under $\mathcal{F}_3$, the expected number of iterations, that is, the expected number of comparisons in the while condition, is given by $\omega[1-f(0;\theta)]^{-1}\mu + 1$, where μ is presented in Table 4 according to each possible choice in $\mathcal{F}_2$.

2.6.1 Study 1

The first simulation study was performed by generating $N = 500$ pseudo-random samples of sizes $n = 50, 100, 200,$ and 500 of a random variable Y distributed according to an element of $\mathcal{F}_3$, in the hurdle configuration, aiming to assess the properties of the Bayesian estimators (2.11), (2.12) and (2.14). In this way, we have assigned different values for θ and ω (and consequently p) accordingly Table 6.

Given N estimates of the model parameters, one can evaluate the performance of the aforementioned estimators using standard measurements as the bias (B), the variance (Var), the mean squared error (MSE), and the mean absolute percentage error (MAPE). Assuming

$\beta = \theta, \omega$ or p, we have computed the following quantities

$$\mathrm{B}\left(\hat{\beta}\right) = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\beta}_j - \beta\right), \ \mathrm{MSE}\left(\hat{\beta}\right) = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\beta}_j - \beta\right)^2, \ \text{and} \ \mathrm{MAPE}\left(\hat{\beta}\right) = \frac{1}{N}\sum_{j=1}^{N}\left|\frac{\hat{\beta}_j - \beta}{\beta}\right|.$$

The variance of $\hat{\beta}$ was estimated as the difference between the MSE and the square of the bias. Moreover, we have also estimated the coverage probability (CP) of the Bayesian credible intervals (BCIs) as follows

$$\mathrm{CP}_{\%}(\beta) = \frac{100}{N}\sum_{j=1}^{N}\delta_{\beta}^{j},$$

where δ_{β}^{j} assumes 1 if the j-th BCI contains the true value β and 0 otherwise. The BCI can be obtained by computing the percentiles of the *posterior* distribution of β. Besides, we have estimated the below noncoverage probability (BNCP) and the above noncoverage probability (ANCP) of the BCIs. These measures are computed analogously to the CP. The BNCP and the ANCP may be useful to highlight asymmetric features since they provide the estimated probabilities of finding the true value of β on the tails of the generated *posterior* distributions.

Based on the generated samples, we have summarized the behavior of the *posterior* distributions of parameters θ, ω, and p (see Appendix A, Section A.3). From the results obtained in this first simulation study, we conclude that the parameter estimates become more accurate since the estimated bias and MSEs decrease with increasing sample size in most cases. As expected, when zero-inflated samples are generated with $\theta = 5.00$ and $p = 0.20$ (Scenario 3), the majority of observations are zero-valued, and hence a little amount of positive observations remains available to estimate parameter θ accurately. In this case, for small n, our methodology does not adequately capture the real variability of $\hat{\theta}$ but, with increasing sample sizes, the biases are converging to more suitable values under these conditions. In this situation, we have also observed higher MAPE values for parameters θ and p even for large samples. This issue can lead to an inaccurate estimation process. For example, under the $\mathcal{ZMPL}$ model (zero-inflated case - Scenario 3, Table 48), for $n = 200$, we have observed a MAPE of approximately 115% for θ. Taking into account that the actual value of this parameter is 5.00, we have $\hat{\theta} = 2.15 \times 5.00 = 10.75$, which represents a considerable impact on the final result. In this case, however, the impact on parameter p can be considered softer, mainly when its actual value is small. Nevertheless, we have observed that the estimated coverage probabilities of the BCIs are always close to the nominal value of 95%, although small degrees of asymmetry have been observed in some cases. Therefore, our first simulation study indicates the feasibility of the proposed Bayesian approach to the use of any element of $\mathcal{F}_3$ in the analysis of overdispersed and zero-modified real datasets, excepting the cases where a substantial amount of zeros is observed (e.g., higher than $0.95n$ out of n values) in samples with very low mean (e.g., lower than 0.05).

2.6.2 Study 2

The second simulation study was conducted to evaluate the reliability of the Bayesian comparison criteria, presented in Appendix A (Section A.2), to perform correct model selection when the true one is known. The idea is to generate $N = 500$ pseudo-random samples of sizes $n = 50, 100, 200,$ and 500 of a random variable Y distributed according to an element of $\mathcal{F}_3$, in the hurdle configuration. Then, the true model is fitted along with the standard $\mathcal{ZMP}$ model, and the best fit is considered according to the selection criteria presented in Appendix A (Section A.2). Through this procedure, it was expected the comparison criteria to be able to identify the mechanism responsible for generating the data. We have considered this procedure for different specifications of parameters θ and ω accordingly Table 6.

For the purpose of this study, let ξ_{M_k} be the value of any statistic computed from the *posterior* distribution of any parametric vector $\boldsymbol{\theta}$ given model M_k $(k = 1, 2, \ldots, K)$, K finite. In our case, we have $K = 2$. Also, ξ_{M_k} is the value of any comparison criteria under consideration (deviance information criterion (DIC), expected Akaike information criterion (EAIC), expected Bayesian information criterion (EBIC), and log-marginal pseudo-likelihood (LMPL)). Without loss of generality, if the true model is M_1, the probability of correct selection based on N Monte Carlo simulations can be recovered by considering the following expression

$$\hat{p}_{CS} = \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi}^{j},$$

where δ_{ξ}^{j} assumes 1 if in the j-th simulation run, we get $|\xi_{M_1}| < |\xi_{M_2}|$ and 0 otherwise. Conversely, $(1 - \hat{p}_{CS})$ denotes the relative number of times that model M_2 was selected incorrectly. In each step of this procedure, model M_1 refers to a particular element of $\mathcal{F}_3$ and M_2 is always the $\mathcal{ZMP}$ model.

From the obtained results, we conclude that when applied in the comparison between any element of $\mathcal{F}_3$ and the standard $\mathcal{ZMP}$ model, the lower the value of θ the better the performance of the Bayesian selection criteria (see Tables 83-88). As mentioned in the previous simulation study, the third scenario of the zero-inflated case is quite problematic, and hence, the comparison criteria have not behaved as expected. In this scenario, the correct selection probability, for $n = 500$, was estimated between 34% and 43% when the $\mathcal{ZMPL}$ model is the true one and is not greater than 42% in the case of $\mathcal{ZMPS}_h$ distribution. Also, when dealing with zero-deflated samples, it is more likely the criteria to find difficulties when selecting the actual model since the correct selection probability was always estimated lower than 64.50% on Scenarios 3 and 4, which consider $\theta = 6.00$. In general, we have observed that model $\mathcal{ZMPS}_u$ is correctly selected more often than $\mathcal{ZMPL}$ and $\mathcal{ZMPS}_h$ models. Further, it is worthwhile to mention that, regardless the sample is zero-inflated/deflated, the LMPL criterion presents better performance in all scenarios, and therefore, it is highly recommended its use in this type of comparison mainly when small samples are available.

2.6.3 Study 3

In Section 2.5, we have pointed out that estimating the $\mathcal{ZMP}$ parameter θ using the $\mathcal{ZTP}$ distribution results in a loss of efficiency in the inference if there is no zero modification. From the frequentist point of view, Dietz and Böhning (2000) have derived the asymptotic variance of the MLE $\hat{\theta}$ for this model, which naturally corresponds to the asymptotic variance of $\hat{\theta}$ under the $\mathcal{ZTP}$ distribution. In this context, it is essential to evaluate the loss of efficiency in the estimation procedure using a hurdle model when no zero modification occurs $(p = 1)$ since, in this case, the zero-valued observations are discarded to obtain $\hat{\theta}$. In such a way, the authors have derived the asymptotic relative efficiency (ARE), and they pointed out that the impact of misspecification on the parameter estimate becomes negligible when the data are generated with large mean. Through their derivation, this is the expected result since ARE goes to 1 as θ increases. Further, they had concluded that the MLE under the $\mathcal{ZTP}$ distribution is practically fully efficient when the actual parameter is higher than 8.00.

In our context, if we had considered the frequentist approach, very complicated forms would be derived for the ARE when the true model is an element of $\mathcal{F}_2$. Under a Bayesian point of view, one cannot obtain analytic expressions for the *posterior* variance of θ since the *posterior* distribution of η does not have closed-form whether the model is zero-modified or not. However, one can easily overcome this issue by considering the MC estimator (2.13) for the *posterior* conditional variance. Therefore, this third simulation study consists of generating $N = 500$ pseudo-random samples of sizes $n = 50, 100, 200$ and 500 of a random variable Y distributed according to an element of $\mathcal{F}_2$. In this procedure, we have defined a *grid* varying in $(0, 6]$, containing 1,000 different values for parameter θ. Then, each value is considered to generate an artificial dataset from any element of $\mathcal{F}_2$. In the following, the true model and its corresponding zero-modified version are fitted, and the Bayesian relative efficiency (BRE) is computed. Considering estimator (2.12), the BRE can be estimated as

$$\widehat{BRE} = \frac{\sum_{j=1}^{M} \left[e^{\eta^{(j)}} - \mathbb{E}(\theta; \boldsymbol{y}) \right]^2}{\sum_{j=1}^{M} \left[e^{\eta_*^{(j)}} - \mathbb{E}(\theta^*; \boldsymbol{y}) \right]^2},$$

being the *posterior* sum of squares provided by the fit of an element of $\mathcal{F}_2$ (numerator) and by the fit of the respective component on $\mathcal{F}_3$ (denominator) – the subscripts * on η, and θ were placed to make this differentiation clear. The pseudo-random generation of a single value from any element of $\mathcal{F}_2$ can be performed in a very similar way to that presented in Algorithm 2.

From the obtained results, we conclude that the BRE measure has provided an interesting tool to evaluate the loss of efficiency, which occurs when using elements of $\mathcal{F}_3$ to fit data generated from $\mathcal{F}_2$, that is, zero-modified models with $p = 1$. When such measure is compared with the true values of θ, it is possible to identify from which values such misspecification does not significantly impact its estimation. We have observed that the uncertainty on the BRE estimate is large for small sample sizes (see Figures 34-36). Within a 10% margin, we have

verified that under the $\mathcal{ZMPL}$ model, this procedure is highly efficient for $\theta < 0.18$ when $n = 500$ is considered. As for $\mathcal{ZMPS}_h$ and $\mathcal{ZMPS}_u$ models, the conclusion remains the same for $\theta < 0.21$ and $\theta < 0.31$. Further, since $\mu \to \infty$ as $\theta \to 0$, it can be seen that the full efficiency is attained for large values of μ, for example, $\mu > 9.00$ under $\mathcal{ZMPS}_u$ distribution.

2.7 Application to real data

The proposed zero-modified distributions are considered in the analysis of three real datasets. The goodness-of-fit of these models is compared with those obtained by some distributions provided in the literature. The first dataset was provided by the Department of Youth and Community Development (DYCD) of New York City (State of New York, US). The records are from 2011, and the sample contains several demographic information from all jurisdictions in the city. The variable of our interest is the number of inhabitants of an ethnicity other than American Indian, Pacific Islander, Hispanic Latino, Asian non-Hispanic, black non-Hispanic, and white non-Hispanic in each jurisdiction.

The second one comprises a set of companies that were targets of tender offers between 1978-1985 and were taken over within 52 weeks of the initial bid. This dataset is due to Jaggia and Thosar (1993), and the response variable is the number of takeover bids after the initial proposal received by the target company.

The last one relates to the Student's historical data on the number of *Haemocytometer* yeast cell counts per square (STUDENT, 1907; GOSSET, 1908), which were obtained in each of 400 regions of a 20×20 grid on a microscope slide.

Table 7 – Response variables and some descriptive statistics for each dataset.

Dataset	Variable	n	Mean	Variance	ID(%)	CV(%)	PZ(%)
1	Number of individuals of another ethnicity	236	0.6964	3.2884	479.06	264.18	77.11
2	Number of takeover bids	126	1.7381	2.0509	117.99	82.39	7.14
3	Number of yeast cells per square	400	0.6825	0.8137	119.23	132.17	53.25

Measures - ID: Index of Dispersion; CV: Coefficient of Variation; PZ: Proportion of Zeros.

Source: Elaborated by the author.

Table 7 presents some descriptive statistics for each response variable. The last column shows the observed proportion of zeros in each dataset. One can note that more than 75% of the observations are zero-valued in the first sample, while the second one presents a low amount of zeros. The third dataset is approximately divided by half into positive and zero-valued

observations. This feature can also be seen in Table 8, which presents the frequency distribution of each dataset. The sample mode of Datasets 1 and 3 is zero, and the mode of Dataset 2 is 1. Further, the initial analysis highlights the presence of overdispersion (see the index of dispersion and the coefficient of variation). These indications justify the choice of the proposed models to describe such data.

Table 8 – Frequency distribution of each dataset.

Dataset	0	1	2	3	4	5	$\geqslant 6$	Max.
1	182	20	11	7	4	3	9	17
2	9	63	31	12	6	1	4	10
3	213	128	37	18	3	1	0	5

Source: Elaborated by the author.

The methodology described in Section 2.5 was considered to estimate the proposed models. For comparison purposes, identical procedures were adopted to fit the $\mathcal{P}$, the $\mathcal{PL}$, the $\mathcal{PS}_h$, the $\mathcal{PS}_u$, the $\mathcal{NB}$ and the $\mathcal{ZMP}$ models. To fit the $\mathcal{P}$ distribution, we have considered the *Jeffreys prior* $\pi_\theta(\theta) \propto \theta^{-1/2}$ for parameter θ, which results in a Gamma *posterior* distribution with shape $n\bar{y} + 1/2$ and scale n^{-1}, being $\bar{y}$ the sample mean. Also, we have considered the $\mathcal{NB}$ distribution parameterized by the positive real-valued parameters μ (mean) and ϕ (dispersion) with a numerical approach for computation of the *Jeffreys prior* through the observed information matrix.

Table 9 – *Posterior* estimates of parameter θ and 95% BCIs from the proposed models.

Dataset	Model	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
1	$\mathcal{ZMPL}$	0.7426	0.7345	0.1022	0.5573	0.9456
	$\mathcal{ZMPS}_h$	0.7651	0.7596	0.0975	0.5757	0.9506
	$\mathcal{ZMPS}_u$	1.0170	1.0110	0.1192	0.7983	1.2537
2	$\mathcal{ZMPL}$	1.5266	1.5151	0.1781	1.1951	1.8809
	$\mathcal{ZMPS}_h$	1.4704	1.4582	0.1579	1.1951	1.8000
	$\mathcal{ZMPS}_u$	1.9265	1.9136	0.1925	1.5723	2.3052
3	$\mathcal{ZMPL}$	2.6774	2.6567	0.3129	2.0889	3.3042
	$\mathcal{ZMPS}_h$	2.4644	2.4467	0.2756	1.9586	3.0210
	$\mathcal{ZMPS}_u$	3.1305	3.1149	0.3134	2.5175	3.7338

Source: Elaborated by the author.

Table 9 presents the mean, the median, the standard deviation and the 95% BCIs obtained from the *posterior* distribution of θ through parameter η. Noticeably, the higher the value of θ,

Table 10 – Comparison criteria and adequacy of the fitted models.

Model	DIC	EAIC	EBIC	LMPL	χ^2 (p-value)	p_{B}
			Dataset 1			
$\mathcal{P}$	726.57	727.59	731.05	−365.19	260.83 (< 0.001)	0.000
$\mathcal{PL}$	550.31	551.31	554.77	−276.15	64.91 (< 0.001)	0.244
$\mathcal{PS}_h$	545.55	546.55	550.01	−273.61	61.96 (< 0.001)	0.236
$\mathcal{PS}_u$	553.12	554.10	557.57	−277.48	66.76 (< 0.001)	0.136
$\mathcal{NB}$	469.62	469.25	476.18	−234.70	0.70 (0.705)	0.506
$\mathcal{ZMP}$	498.62	500.60	507.53	−250.40	14.58 (0.001)	0.112
$\mathcal{ZMPL}$	465.35	467.34	474.27	−232.95	0.98 (0.613)	0.478
$\mathcal{ZMPS}_h$	465.56	467.55	474.48	−233.04	1.05 (0.592)	0.474
$\mathcal{ZMPS}_u$	466.96	468.94	475.87	−233.82	1.60 (0.449)	0.452
			Dataset 2			
$\mathcal{P}$	405.12	406.13	408.97	−202.65	27.39 (< 0.001)	0.662
$\mathcal{PL}$	443.80	444.79	447.62	−221.65	72.11 (< 0.001)	0.762
$\mathcal{PS}_h$	442.70	443.69	446.53	−221.09	71.79 (< 0.001)	0.850
$\mathcal{PS}_u$	438.74	439.78	442.61	−219.14	68.06 (< 0.001)	0.858
$\mathcal{NB}$	409.40	410.07	415.75	−204.69	30.08 (< 0.001)	0.934
$\mathcal{ZMP}$	390.02	392.05	397.72	−195.52	8.12 (0.044)	0.098
$\mathcal{ZMPL}$	372.01	374.05	379.72	−186.18	0.39 (0.942)	0.480
$\mathcal{ZMPS}_h$	371.86	373.89	379.56	−186.10	0.35 (0.950)	0.478
$\mathcal{ZMPS}_u$	372.54	374.60	380.28	−186.46	0.44 (0.932)	0.446
			Dataset 3			
$\mathcal{P}$	900.97	901.99	905.98	−450.58	10.11 (0.018)	0.162
$\mathcal{PL}$	907.30	908.27	912.26	−453.52	14.29 (0.003)	0.570
$\mathcal{PS}_h$	908.81	909.84	913.83	−454.29	15.74 (0.001)	0.566
$\mathcal{PS}_u$	906.88	907.88	911.87	−453.33	14.02 (0.003)	0.594
$\mathcal{NB}$	901.14	901.16	909.14	−450.36	3.30 (0.192)	0.544
$\mathcal{ZMP}$	899.20	901.18	909.16	−449.70	5.23 (0.073)	0.404
$\mathcal{ZMPL}$	896.86	898.85	906.84	−448.40	2.97 (0.227)	0.492
$\mathcal{ZMPS}_h$	896.94	898.94	906.93	−448.43	3.04 (0.219)	0.488
$\mathcal{ZMPS}_u$	896.71	898.74	906.73	−448.33	2.93 (0.231)	0.492

Source: Elaborated by the author.

the lower is the average number of counts under the respective baseline distributions. The *posterior* summary of ω is presented in Table 11. Since our main interest remains on the estimation of the zero modification parameter we present, in Table 12, the *posterior* summary of p. Based on the results of the first simulation study and given the sample size of each dataset,

there are indications that the coverage probability of the BCIs obtained for each parameter is close to the nominal value of 95%.

Table 10 presents the model selection summary using some measures that were considered for comparison purposes (see Appendix A, Section A.2). It is noteworthy that the lower values of DIC, EAIC, and EBIC are provided by the proposed models when fitted to the considered datasets. The conclusion remains the same when assessing the larger value of LMPL. The goodness-of-fit was assessed by the chi-square χ^2 measure obtained from the observed and expected frequencies. The associated p-values are also presented. For each dataset, we have grouped cells with frequencies lower or equal than 5, resulting in $(4-\kappa)$ df for Datasets 1 and 3 and $(5-\kappa)$ df for Dataset 2, being κ the number of estimated parameters. For instance, when considering the zero-modified models we are directly estimating $\kappa = 2$ parameters, and therefore, the df for Datasets 1 and 3 is 2 and the df for Dataset 2 is 3. When analyzing the computed measures, one can notice that the proposed models adhere better to the considered datasets. Further, we have computed the Bayesian p-value (p_{B}) measure to evaluate model suitability. One can notice the unreliability of the $\mathcal{P}$ distribution to fit Dataset 1. Besides, the fits provided by the $\mathcal{NB}$ and the $\mathcal{ZMP}$ models are highly questionable, mainly on the two first datasets. Notably, all measures of model fit evaluation have agreed to the same best model.

Table 11 – *Posterior* estimates of parameter ω and 95% BCIs from the proposed models.

Dataset	Mean	Median	Mode	Std. Dev.	95% BCI	
					Lower	Upper
1	0.2300	0.2292	0.2277	0.0273	0.1780	0.2859
2	0.9252	0.9274	0.9320	0.0233	0.8771	0.9663
3	0.4676	0.4675	0.4674	0.0249	0.4207	0.5197

Source: Elaborated by the author.

By the results presented in Table 12, one can notice that exists evidence that the first dataset is zero-inflated while the second and third ones are zero-deflated. As an indication, when looking at the 95% BCIs obtained for p, we emphasize the importance of considering zero-modified models to fit these datasets since, in cases where exists clear evidence of zero modification, this type of models are highly recommended. In addition, one can use the estimated values of θ and p to estimate the measures presented in Equation (2.3) and the expected number of zeros (n_0) can be estimated as $nf_*(0;\hat{\theta},\hat{p})$. The obtained estimates are presented in Table 13. Through these measures, one can better understand how the proposed models are adhering to the data since the nature of the observed counts is being well described in terms of its overdispersion level as well as the frequency and the average number of non-zero observations.

Figure 3 depicts the observed and the estimated expected frequencies for some fitted

models. For each dataset, we have compared the fits provided by the $\mathcal{ZMPL}$, the $\mathcal{ZMPS}_h$ and the $\mathcal{ZMPS}_u$ models with the $\mathcal{P}$, the $\mathcal{NB}$, and the $\mathcal{ZMP}$ distributions. One can notice that the estimates for n_0 obtained from the $\mathcal{P}$ distribution are much lower than the real one (Datasets 1 and 2) while those provided by zero-modified models are precisely equal to the true ones (see Table 13). Notably, the models $\mathcal{NB}$, $\mathcal{ZMP}$, and $\mathcal{ZMPL}$ have adhered better to Dataset 1 and the models $\mathcal{P}$, $\mathcal{ZMP}$ and $\mathcal{ZMPS}_u$ provided very reliable fits for Dataset 3. However, for Dataset 2, the $\mathcal{NB}$ and $\mathcal{ZMP}$ models have provided almost the same results, whereas the best fit was obtained by the $\mathcal{ZMPS}_h$ model, which also highlights the potentiality of the proposed models.

Table 12 – *Posterior* estimates of parameter p and 95% BCIs from the proposed models.

Dataset	Model	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
1	$\mathcal{ZMPL}$	0.3227	0.3204	0.0426	0.2412	0.4061
	$\mathcal{ZMPS}_h$	0.3198	0.3175	0.0429	0.2344	0.3999
	$\mathcal{ZMPS}_u$	0.3101	0.3076	0.0412	0.2291	0.3902
2	$\mathcal{ZMPL}$	1.8874	1.8754	0.1496	1.6049	2.1762
	$\mathcal{ZMPS}_h$	1.9244	1.9121	0.1618	1.6244	2.2467
	$\mathcal{ZMPS}_u$	1.8670	1.8571	0.1548	1.5706	2.1602
3	$\mathcal{ZMPL}$	1.4367	1.4259	0.1563	1.1693	1.7760
	$\mathcal{ZMPS}_h$	1.4723	1.4636	0.1607	1.1756	1.8041
	$\mathcal{ZMPS}_u$	1.4459	1.4366	0.1580	1.1589	1.7655

Source: Elaborated by the author.

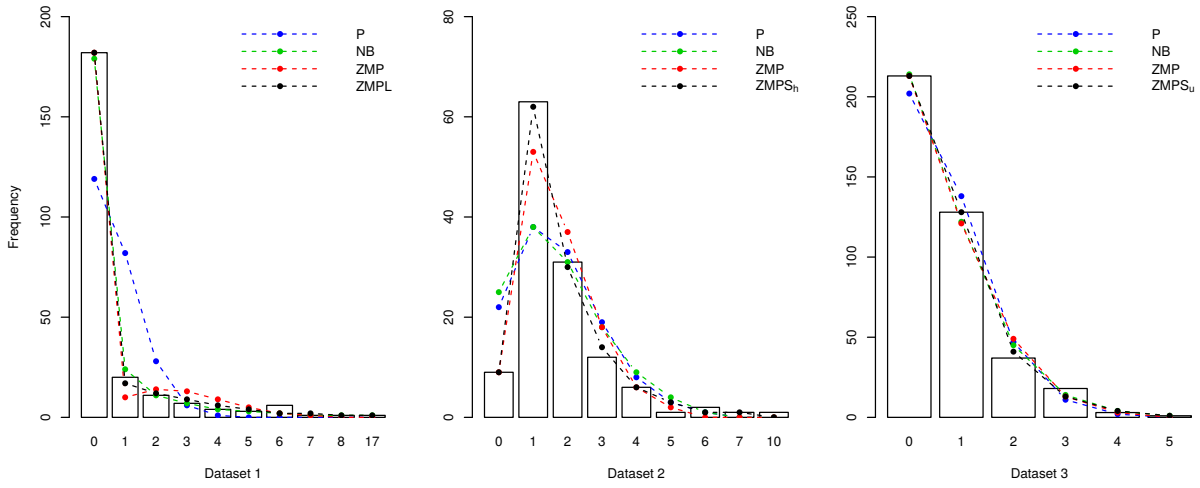
Table 13 – *Posterior* estimates of extra parameters from the proposed models.

Dataset	Model	$\hat{\mu}^*$	$\left(\hat{\sigma}^2\right)^*$	$\hat{\gamma}^*$	$\hat{n}_0$
1	$\mathcal{ZMPL}$	0.68	2.73	3.99	182
	$\mathcal{ZMPS}_h$	0.68	2.69	3.94	
	$\mathcal{ZMPS}_u$	0.68	2.60	3.81	
2	$\mathcal{ZMPL}$	1.73	1.65	0.96	9
	$\mathcal{ZMPS}_h$	1.72	1.66	0.96	
	$\mathcal{ZMPS}_u$	1.72	1.59	0.92	
3	$\mathcal{ZMPL}$	0.68	0.84	1.22	213
	$\mathcal{ZMPS}_h$	0.68	0.84	1.23	
	$\mathcal{ZMPS}_u$	0.68	0.83	1.22	

Source: Elaborated by the author.

In general, we have noticed that the zero-modified models are providing more realistic approaches for the considered datasets and that the proposed models are competitive with the $\mathcal{NB}$ distribution when dealing with overdispersed samples. This feature can be regarded as one of the most relevant achievements of the proposed models since they have to deal with non-zero observations using a single parameter (θ) while the $\mathcal{NB}$ distribution is highly competitive as it has two parameters (μ, ϕ) to perform the same task.

Figure 3 – *Posterior* expected frequencies under the fitted models.



Source: Elaborated by the author.

We conclude by pointing out that the use of a discrete distribution that accommodates a dispersion level beyond that one caused by the zero modification in a given dataset may be meaningful in the cases where $p = 1$, with overdispersion that is explained only by the theoretical model. In this way, there exists evidence that the proposed models adhere better to the three analyzed datasets and hence, can be considered as suitable options to model zero-inflated or zero-deflated count data in the presence of overdispersion.

2.8 Concluding remarks

In this capter, we have introduced a class of zero-modified Poisson mixture models as an alternative to analyze zero-inflated/deflated data in the presence of overdispersion. Through the hurdle approach, it was possible to write separable likelihood functions for the model parameters, which facilitates classical and Bayesian procedures. In this setup, a closed-form MLE can be derived only for parameter ω. Besides, we have shown that parameter θ of each model can be estimated using just the positive observations in a given sample. Through Monte Carlo simulations, we have assessed the properties of the Bayesian estimators, and we have obtained an approximation for the probability of correct selection of the proposed models compared with the standard $\mathcal{ZMP}$ distribution using the presented comparison criteria. Also, we have

estimated, via simulation, the loss of efficiency in the estimation of θ when a zero-modified model is considered in the case where no zero modification occurs $(p = 1)$. The proposed models were applied in the analysis of three real datasets. The response variables were detected as overdispersed and zero-modified, which justified the choice of this kind of model. Also, according to the comparison criteria, the proposed models presented better fits when compared with several other models provided in the literature. Further, using the Bayesian p-value, we have verified the suitability of the fits obtained by the proposed zero-modified models. Therefore, we conclude that the proposed class is an excellent addition to the set of models that can be used when analyzing overdispersed and zero-modified count data.

CHAPTER

3

THE ZERO-MODIFIED POISSON-SHANKER REGRESSION MODEL

3.1 Introduction

In this chapter, we propose the zero-modified Poisson-Shanker ($\mathcal{ZMPS}_h$) regression model as an alternative to model overdispersed count data exhibiting inflation or deflation of zeros in the presence of covariates. The zero modification has been incorporated using the zero-truncated version of the Poisson-Shanker ($\mathcal{PS}_h$) distribution. The $\mathcal{PS}_h$ distribution has been written as a hurdle model using a simple reparameterization of the probability function, which leads to the fact that the proposed model can be fitted without any previous information about the zero modification present in a given dataset. Bayesian procedures have been considered for estimation and inference. A simulation study has been presented to illustrate the performance of the developed methodology. The usefulness of the proposed model has been evaluated using a real dataset on fetal deaths notification data in Bahia State, Brazil. A sensitivity study to detect points which can influence the parameter estimates has been performed using Kullback-Leibler divergence measure. The randomized quantile residuals have been considered for the model validation issue. A general comparison of the proposed model with some well-known discrete distributions has been provided.

This chapter is organized as follows. In Section 3.2, we briefly present the $\mathcal{PS}_h$ distribution and some of its mathematical properties. In Section 3.3, we present the zero-truncated and introduce the zero-modified version of the $\mathcal{PS}_h$ distribution, demonstrating its flexibility to deal with zero-inflated/deflated data. In Section 3.4, we present the $\mathcal{ZMPS}_h$ distribution as a hurdle model. In Section 3.5, we present a regression framework based on the $\mathcal{ZMPS}_h$ distribution. In Section 3.6, the maximum likelihood function, the *prior* and the *posterior* densities for the unknown parameters, the method to evaluate the effect of influential points and the randomized quantile residuals theory are stated for the Bayesian inference procedures. In Section 3.7, a

simulation study is presented. In Section 3.8, an application of the proposed model is considered to evaluate its usefulness. Concluding remarks are addressed in Section 3.9.

3.2 The PS_h distribution

A random variable λ is said to have a Shanker $(\mathcal{S}_h)$ distribution if its probability density function can be written as

$$g(\lambda;\theta) = \frac{\theta^2}{\theta^2+1}(\theta+\lambda)e^{-\theta\lambda}, \quad \lambda \in \mathbb{R}_+,$$

for $\theta \in \mathbb{R}_+$.

The $\mathcal{PS}_h$ distribution is a probabilistic model that arises when the $\mathcal{S}_h$ distribution is chosen to describe the rate parameter λ of the Poisson $(\mathcal{P})$ distribution. In this case, a random variable Y is said to have $\mathcal{PS}_h$ distribution if it follows the stochastic representation

$$Y|\lambda \sim \mathcal{P}(\lambda) \quad \text{and} \quad \lambda \sim \mathcal{S}(\theta).$$

The unconditional distribution of the random variable Y can be denoted by $\mathcal{PS}_h(\theta)$. Let $\mathcal{Y}_0 = \{0,1,\ldots\}$ be the set of nonnegative integers. We completed the definition by stating that a random variable Y, defined on $\mathcal{Y}_0$, will have $\mathcal{PS}_h$ distribution if its probability mass function (pmf) can be written as

$$f(y;\theta) = \frac{\theta^2}{\theta^2+1}\left[\frac{\theta^2+\theta+y+1}{(\theta+1)^{y+2}}\right], \quad y \in \mathcal{Y}_0, \tag{3.1}$$

for $\theta \in \mathbb{R}_+$. Using the gamma integral, the above result can be easily proved by integrating $f(y|\lambda)g(\lambda;\theta)$ with respect to λ over $\mathbb{R}_+$, being $f(y|\lambda)$ the conditional pmf of a $\mathcal{P}$ variable.

The cumulative distribution function (cdf) of Y is given by

$$\mathrm{F}(y;\theta) = 1 - \left[\frac{\theta^3+\theta^2+\theta y+2\theta+1}{(\theta^2+1)(\theta+1)^{y+2}}\right], \quad y \in \mathcal{Y}_0.$$

From the results provided by Shanker (2016c), we have that the r-th factorial moment about the origin of the $\mathcal{PS}_h$ distribution is given by

$$\mu'_r = \frac{r!\left(\theta^2+r+1\right)}{\theta^r(\theta^2+1)},$$

which provides the moments about the origin. Thus, the mean and the variance are

$$\mu = \mu'_1 = \frac{\theta^2+2}{\theta(\theta^2+1)}, \quad \text{and} \quad \sigma^2 = \mu'_2 - \left(\mu'_1\right)^2 = \frac{\theta^5+\theta^4+3\theta^3+4\theta^2+2\theta+2}{\theta^2(\theta^2+1)^2}. \tag{3.2}$$

The variance term can be easily written as

$$\sigma^2 = \mu\left[1+\frac{\theta^4+4\theta^2+2}{\theta(\theta^2+1)(\theta^2+2)}\right] = \mu\gamma, \tag{3.3}$$

being the ratio involving the parameter θ always positive. This implies that the $\mathcal{PS}_h$ distribution is overdispersed, that is, whichever $\theta \in \mathbb{R}_+$ we have that $\sigma^2 > \mu$. Further, the useful index of dispersion (γ) is clearly greater than 1, also implying overdispersion since $\gamma = \sigma^2\mu^{-1}$. On the other hand, we have that $\gamma \to 1$ ($\sigma^2 \to \mu$) as $\theta \to \infty$, that is, the $\mathcal{PS}_h$ distribution has the property of equidispersion for large values of θ.

Now, let us reparameterize the pmf (3.1) in terms of the mean (μ). This can be useful since our interest is to define a regression framework based on the $\mathcal{PS}_h$ distribution. Since $\theta \in \mathbb{R}_+$, we have that

$$\theta = \frac{1}{3\mu}\left[1 + \frac{s(\mu)}{2} - \frac{2\left(3\mu^2 - 1\right)}{s(\mu)}\right], \tag{3.4}$$

where $s(\mu) = [12\mu\sqrt{3\left(4\mu^4 + 71\mu^2 + 8\right)} + 180\mu^2 + 8]^{1/3}$. Hence, if we denote $\theta = h(\mu)$, Equation (3.1) becomes

$$f(y;\mu) = \frac{h^2(\mu)}{h^2(\mu)+1}\left[\frac{h^2(\mu)+h(\mu)+y+1}{[h(\mu)+1]^{y+2}}\right], \quad y \in \mathcal{Y}_0,$$

for $\mu \in \mathbb{R}_+$. To complete the definition of the $\mathcal{PS}_h$ regression model we may consider a logarithmic link function that relates the mean to a linear predictor by $\log(\boldsymbol{\mu}) = \boldsymbol{x}^\intercal\boldsymbol{\beta}$, where $\boldsymbol{x} = (1, x_1, \ldots, x_q)$ is the vector of covariates and $\boldsymbol{\beta}^\intercal = (\beta_0, \beta_1, \ldots, \beta_q)$ is the vector of unknown parameters, both having length $q+1$.

3.3 The ZMPS$_h$ distribution

Let Y be a random variable defined on $\mathcal{Y}_0$. Thus, Y is said to have $\mathcal{ZMPS}_h$ distribution if its pmf can be written as

$$f_*(y;\mu,p) = (1-p)\delta_y + pf(y;\mu), \quad y \in \mathcal{Y}_0, \tag{3.5}$$

for $\mu \in \mathbb{R}_+$. The parameter p is subject to the condition (called p-condition) given by

$$0 \leqslant p \leqslant \frac{1}{1 - f(0;\mu)}, \tag{3.6}$$

being $f(0;\mu)$ the pmf of the $\mathcal{PS}_h$ random variable with mean μ, evaluated at zero. Further, δ_y is the indicator function, so that $\delta_y = 1$ if $y = 0$ and $\delta_y = 0$ otherwise. Note that (3.5) is not a mixture distribution typically fitted to zero-inflated data, since parameter p can assume values greater than 1. However, for all values of p between 0 and the boundary $[1 - f(0;\mu)]^{-1}$, the Equation (3.5) corresponds to a properly pmf since $f_*(y;\mu,p) \geqslant 0$ for all $y \in \mathcal{Y}_0$ and $\sum_{y\in\mathcal{Y}_0} f_*(y;\mu,p) = 1$.

Now, the corresponding cdf is given by

$$\mathrm{F}_*(y;\mu,p) = 1 - \frac{p\left[h^3(\mu) + h^2(\mu) + yh(\mu) + 2h(\mu) + 1\right]}{\left[h^2(\mu)+1\right]\left[h(\mu)+1\right]^{y+2}}, \quad y \in \mathcal{Y}_0. \tag{3.7}$$

The mean, the variance and the index of dispersion of Y are

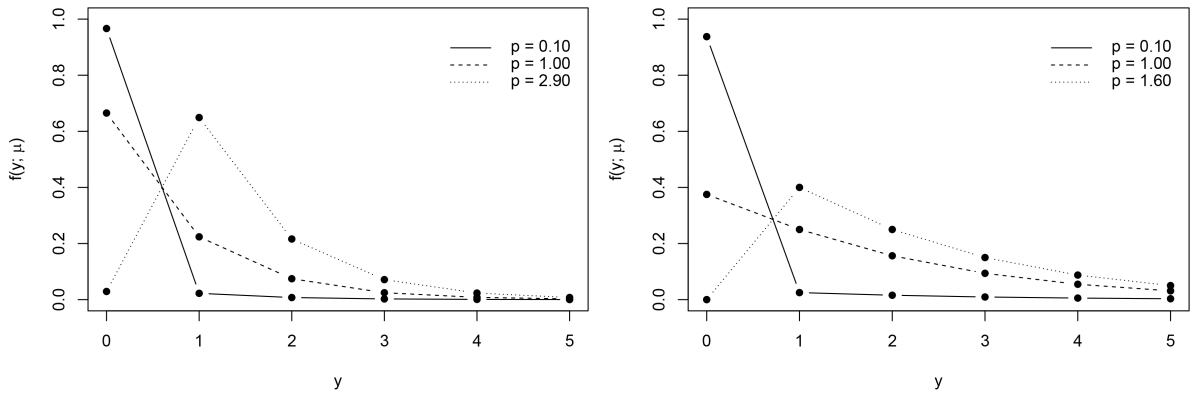
$$\mu_* = p\mu, \quad \sigma^2_* = p\left[\sigma^2 + (1-p)\mu^2\right] \quad \text{and} \quad \gamma_* = \gamma + (1-p)\,\mu,$$

where μ and σ^2 are given in Equation (3.2) and γ in Equation (3.3). The $\mathcal{ZMPS}_h$ distribution may be considered an interesting alternative to the usual zero-modified Poisson ($\mathcal{ZMP}$) model since the basis distribution of the former can accommodate several levels of overdisperson, issue that the $\mathcal{P}$ distribution generally fails in deal with.

Figure 4 depicts the pmf of the $\mathcal{ZMPS}_h$ distribution (3.5) for $\mu = 0.50$ (implying $0 \leqslant p \leqslant 2.99$), for $\mu = 1.50$ (implying $0 \leqslant p \leqslant 1.60$), for $\mu = 3.00$ (implying $0 \leqslant p \leqslant 1.24$) and for $\mu = 5.00$ (implying $0 \leqslant p \leqslant 1.11$). Notice that different values of p lead to different $\mathcal{ZMPS}_h$ distributions, as can also be seen by considering the proportion of additional or missing zeros, given by

$$f_*(0;\mu,p) - f(0;\mu) = (1-p) + pf(0;\mu) - f(0;\mu) = (1-p)[1 - f(0;\mu)].$$

Figure 4 – Behavior of the $\mathcal{ZMPS}_h$ distribution for different values of μ and p.



(a) $\mu = 0.50$ (b) $\mu = 1.50$

(c) $\mu = 3.00$ (d) $\mu = 5.00$

Source: Elaborated by the author.

One can notice that parameter p controls the frequency of zeros. When $p = 0$ we have $f_*(0;\mu,p) = 1$. In such case, (3.5) is a degenerate distribution with all mass at zero. For

all $0 < p < 1$ we have $(1-p)[1-f(0;\mu)] > 0$ and therefore, $f_*(0;\mu,p) > f(0;\mu)$. In such case, (3.5) corresponds to the pmf of the zero-inflated Poisson-Shanker $(\mathcal{ZIPS}_h)$ distribution which has a proportion of zero greater than the usual $\mathcal{PS}_h$ distribution. When $p = 1$ we have $f_*(0;\mu,p) - f(0;\mu) = 0$ and therefore, (3.5) is the pmf of the $\mathcal{PS}_h$ distribution. For all $1 < p < [1-f(0;\mu)]^{-1}$ we have $(1-p)[1-f(0;\mu)] < 0$ and therefore, $f_*(0;\mu,p) < f(0;\mu)$. In such case, (3.5) corresponds to the pmf of the zero-deflated Poisson-Shanker $(\mathcal{ZDPS}_h)$ distribution, which has a proportion of zero smaller than the usual $\mathcal{PS}_h$ distribution. Finally, when $p = [1-f(0;\mu)]^{-1}$ we have $f_*(0;\mu,p) = 0$ and therefore, (3.5) is the zero-truncated Poisson-Shanker $(\mathcal{ZTPS}_h)$ distribution, with pmf given by

$$f^*(y;\mu) = \frac{h^2(\mu)}{h^3(\mu)+h^2(\mu)+2h(\mu)+1}\left[\frac{h^2(\mu)+h(\mu)+y+1}{[h(\mu)+1]^y}\right](1-\delta_y), \quad y \in \mathcal{Y}_0, \tag{3.8}$$

for $\mu \in \mathbb{R}_+$.

3.4 Hurdle version of the PS_h distribution

The class of hurdle models was introduced by Mullahy (1986). The relevant feature of such models is that the zero outcomes are treated separately from the positive ones. In the main formulation, a binary probability model determines whether a zero or a non-zero outcome occurs, and hence, an appropriated zero-truncated discrete distribution is chosen to describe the positive values (SAFFARI; ADNAN; GREENE, 2012).

Let us define the hurdle version of the $\mathcal{PS}_h$ distribution. Firstly, the pmf (3.5) can be written as

$$\begin{aligned} f_*(y;\mu,p) &= [1-p+pf(0;\mu)]\delta_y + pf(y;\mu)(1-\delta_y) \\ &= [1-p(1-f(0;\mu))]\,\delta_y + pf(y;\mu)(1-\delta_y), \quad y \in \mathcal{Y}_0. \end{aligned} \tag{3.9}$$

Now, setting $\omega = p[1-f(0;\mu)]$, Equation (3.9) becomes

$$f_*(y;\mu,\omega) = (1-\omega)\delta_y + \omega f^*(y;\mu), \quad y \in \mathcal{Y}_0, \tag{3.10}$$

where $f^*(y;\mu)$ is the $\mathcal{ZTPS}_h$ distribution given by (3.8). Since $0 \leqslant p \leqslant [1-f(0;\mu)]^{-1}$ then $0 \leqslant p[1-f(0;\mu)] \leqslant 1$. Hence $0 \leqslant \omega \leqslant 1$.

The pmf (3.10) can be seen as a hurdle version of the $\mathcal{PS}_h$ distribution, where the probability of $Y = 0$ is $(1-\omega)$ and the probability of $Y > 0$ is ω. The $\mathcal{ZMPS}_h$ distribution parameterized by ω is denoted by $\mathcal{ZMPS}_h(\mu,\omega)$.

The $\mathcal{ZMPS}_h$ distribution expressed as a hurdle version of the $\mathcal{PS}_h$ model contains the $\mathcal{ZTPS}_h$ distribution as one of its components, which differs from the traditional mixture representation of zero-inflated distributions. Moreover, this representation of the $\mathcal{ZMPS}_h$ distribution can be interpreted as a superposition of two processes, that is, one that produces

positive observations from a $\mathcal{ZTPS}_h$ distribution and another one that produces only zero-valued observations with probability $(1-\omega)$.

The hurdle version of the $\mathcal{PS}_h$ distribution can be used to derive the maximum likelihood estimators for parameters μ and ω. Furthermore, such an approach allows us to use only the positive observations in a given dataset to estimate the parameter μ assuming that these observations come from a $\mathcal{ZTPS}_h$ distribution, while parameter ω can be estimated as the proportion of non-zeros in the dataset. The parameter p can be estimated using Equation $p = \omega[1-f(0;\mu)]^{-1}$ (by invariance principle). Indeed, inference procedures about parameter p are required since we may often be interested in identifying what kind of zero modification (inflation or deflation) is present in a given dataset.

3.5 The ZMPS$_h$ regression model

Let us suppose that we have a collection $(Y_1,\ldots,Y_n)$ of independent discrete random variables such that $Y_i|\boldsymbol{x}_i,\boldsymbol{z}_i,\boldsymbol{\beta} \sim \mathcal{ZMPS}_h(\mu_i,\omega_i)$ $(i=1,\ldots,n)$, where $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ is the full vector of fixed-effects. In this case, a regression model for count data based on the $\mathcal{ZMPS}_h$ distribution can be derived by rewriting Equation (3.10) as

$$f_*[y_i;\mu(\boldsymbol{x}_i),\omega(\boldsymbol{z}_i)] = [1-\omega(\boldsymbol{z}_i)]\,\delta_{y_i} + \omega(\boldsymbol{z}_i) f^*[y_i;\mu(\boldsymbol{x}_i)], \qquad y_i \in \mathcal{Y}_0. \tag{3.11}$$

To complete the definition of the $\mathcal{ZMPS}_h$ regression model we should consider that there are two known link functions, namely $g_1(\mu)$ and $g_2(\omega)$, so that $\mu(\boldsymbol{x}_i) = g_1^{-1}(\boldsymbol{x}_i^\top\boldsymbol{\beta}_1)$ and $\omega(\boldsymbol{z}_i) = g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)$, respectively. The $g_1(\boldsymbol{\mu}) = \boldsymbol{x}^\top\boldsymbol{\beta}_1$ is a differentiable link function, which relates the mean (μ) to a linear predictor with vector of parameters $\boldsymbol{\beta}_1^\top = (\beta_{10},\beta_{11},\ldots,\beta_{1q_1})$ and vector of covariates $\boldsymbol{x}_i^\top = (1,x_{i1},\ldots,x_{iq_1})$. The $g_2(\omega)$ is another appropriate link function, which relates ω to a linear predictor with vector of parameters $\boldsymbol{\beta}_2^\top = (\beta_{20},\beta_{21},\ldots,\beta_{2q_2})$ and vector of covariates $\boldsymbol{z}_i^\top = (1,z_{i1},\ldots,z_{iq_2})$. Here, q_1 and q_2 are the number of covariates used to model μ and ω, respectively.

An appropriate link function $g_2(\boldsymbol{\omega}) = \boldsymbol{z}^\top\boldsymbol{\beta}_2$, which relates ω to a linear predictor is given by

$$\text{logit}\,(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = \boldsymbol{z}_i^\top\boldsymbol{\beta}_2, \tag{3.12}$$

such that $0<\omega_i<1$. Further, the *probit* link function

$$\Phi^{-1}(\omega_i) = \boldsymbol{z}_i^\top\boldsymbol{\beta}_2, \tag{3.13}$$

is also appropriate for the desired purpose. Another very useful alternative is the *complementary log-log* link function defined as

$$\log[-\log(1-\omega_i)] = \boldsymbol{z}_i^\top\boldsymbol{\beta}_2, \tag{3.14}$$

which allows for a skewed specification. In addition, a more sophisticated approach considering power and reversal power link functions was proposed by Bazán *et al.* (2017) and can be applied in our context to provide more flexible relationships between the linear predictor and the parameter ω.

The $\mathcal{ZMPS}_h$ regression model has q_1+q_2+2 unknown parameters to be estimated, which correspond to the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The link functions for parameter ω, given by Equations (3.12), (3.13) and (3.14), exclude two specific cases of the random variable Y, namely when Y has a degenerate distribution at zero and when it has a zero-truncated distribution. It is worthwhile to mention that identifiability problems could occur if the same covariate was used to model both mean (μ) and the modification parameter (p) if we had considered a regression model derived from (3.5). From Equation (3.11), we can see $f_*[y_i;\mu(\boldsymbol{x}_i),\omega(\boldsymbol{z}_i)]$ as a mixture of a $\mathcal{ZTPS}_h$ and a distribution with point mass equals to 1 at zero, with mixing probabilities $\omega(\boldsymbol{z}_i)$. Unlike traditional approaches, model (3.11) can be fitted to both zero-inflated and zero-deflated data. Then, $\mathrm{P}(Y_i=0)=1-\omega(\boldsymbol{z}_i)$ and $\mathrm{P}(Y_i=y_i)=\omega(\boldsymbol{z}_i)f^*[y_i;\mu(\boldsymbol{x}_i)]$, for $y_i>0$. This parameterization makes model (3.11) separable into two parts, leading to orthogonality between the parameters in $\mu(\boldsymbol{x}_i)$ and $\omega(\boldsymbol{z}_i)$. This avoids nonidentifiability problems as well the use of the EM algorithm, typically used to fit mixture models. Regardless the model framework, we will adopt a full Bayesian approach for inference concerns. The likelihood function, the *prior* and the *posterior* densities are detailed in the next section.

3.6 Inference

Let $\boldsymbol{Y}=(Y_1,\ldots,Y_n)$ be a random sample of size n from the $\mathcal{ZMPS}_h$ distribution and $\boldsymbol{y}=(y_1,\ldots,y_n)$ its observed values. For each y_i we have associated vectors of covariates $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$. Considering the model (3.11), the likelihood function for the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ is given by

$$\mathcal{L}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})=\prod_{i=1}^{n}[1-\omega(\boldsymbol{z}_i)]^{\delta_{y_i}}\left[\frac{\omega(\boldsymbol{z}_i)\,f[y_i;\mu(\boldsymbol{x}_i)]}{1-f[0;\mu(\boldsymbol{x}_i)]}\right]^{1-\delta_{y_i}},$$

where the link function which relates μ to a linear predictor is $g_1[\mu(\boldsymbol{x}_i)]=\log[\mu(\boldsymbol{x}_i)]=\boldsymbol{x}_i^\top\boldsymbol{\beta}_1$ and an appropriate link function which relates $\omega(\boldsymbol{z}_i)$ to a linear predictor is $g_2[\omega(\boldsymbol{z}_i)]=\boldsymbol{z}_i^\top\boldsymbol{\beta}_2$. Now, the log-likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) &= \sum_{i=1}^{n}(1-\delta_{y_i})\left\{\log\left[\frac{f(y_i;\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}_1\})}{1-f(0;\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}_1\})}\right]+\log[g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)]\right\}+\\ &\quad\sum_{i=1}^{n}\delta_{y_i}\log[1-g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)]\\ &= \sum_{i=1}^{n}(1-\delta_{y_i})\left\{\log\left[\frac{f(y_i;\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}_1\})}{1-f(0;\exp\{\boldsymbol{x}_i^\top\boldsymbol{\beta}_1\})}\right]\right\}+\\ &\quad\sum_{i=1}^{n}\log[g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)]-\delta_{y_i}\log\left[\frac{g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)}{1-g_2^{-1}(\boldsymbol{z}_i^\top\boldsymbol{\beta}_2)}\right]\end{aligned}$$

$$= \ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{x}) + \ell_2(\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{z}).$$

One can notice that vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal and that $\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y})$ depends only on the positive values of $\boldsymbol{y}$. Denoting by $\boldsymbol{y}^+$ the vector of positive observations from $\boldsymbol{y}$, $\{y_j^+, j = 1,\ldots,n^+\}$ and $\{\boldsymbol{x}_j^+, j = 1,\ldots,n^+\}$ the vector of covariates associated with each y_j^+, the log-likelihood function for $\boldsymbol{\beta}_1$ based on the supposition that y_j^+ comes from a $\mathcal{ZTPS}_h$ distribution is given by

$$\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y}^+,\boldsymbol{x}^+) = \sum_{j=1}^{n^+} \log\left[\frac{f\left(y_j^+;\exp\{(\boldsymbol{x}_j^+)^\intercal\boldsymbol{\beta}_1\}\right)}{1-f\left(0;\exp\{(\boldsymbol{x}_j^+)^\intercal\boldsymbol{\beta}_1\}\right)}\right]. \tag{3.15}$$

Since $\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y}^+,\boldsymbol{x}^+) = \ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{x})$, the setup of a $\mathcal{ZMPS}_h$ regression model for $\mu(\boldsymbol{x})$ is equivalent to assume that the positive responses of observed vector $\boldsymbol{y}$ come from a $\mathcal{ZTPS}_h$ distribution. On the other hand, denoting by $\boldsymbol{y}^0$ the vector of zero observations from $\boldsymbol{y}$, $\{y_j^0, j = 1,\ldots,n_0\}$ and $\{\boldsymbol{z}_j^0, j = 1,\ldots,n_0\}$ the vector of covariates associated with each y_j^0, the log-likelihood function for $\boldsymbol{\beta}_2$ is given by

$$\ell_2(\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{z}) = \sum_{i=1}^{n}\log[g_2^{-1}(\boldsymbol{z}_i^\intercal\boldsymbol{\beta}_2)] - \sum_{j=1}^{n_0}\log\left[\frac{g_2^{-1}\left((\boldsymbol{z}_j^0)^\intercal\boldsymbol{\beta}_2\right)}{1-g_2^{-1}\left((\boldsymbol{z}_j^0)^\intercal\boldsymbol{\beta}_2\right)}\right]. \tag{3.16}$$

There are no closed-form for the maximum likelihood estimator (MLE) of the parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ and therefore, standard optimization algorithms such Newton-Raphson based methods may be used to obtain numerical estimates in the frequentist approach. By the maximum likelihood theory, a consistent estimator for the covariance matrix of $\hat{\boldsymbol{\beta}}_j$ $(j = 1,2)$, is given by the inverse of $\mathbb{E}[\mathcal{K}_j]$, being

$$\mathcal{K}_1 = -\frac{\partial^2\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{x})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1^\intercal} \quad \text{and} \quad \mathcal{K}_2 = -\frac{\partial^2\ell_2(\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{z})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2^\intercal},$$

the Hessian matrices. In our context, however, the computation of the expected value respect to $\boldsymbol{Y}$ is unfeasible, and therefore, the covariance matrices can be approximated only numerically by evaluating the Hessian matrices at $\boldsymbol{\beta}_j = \hat{\boldsymbol{\beta}}_j$ and using available observations.

3.6.1 Prior and Posterior distributions

A general class of *prior* densities, the so-called information matrix *prior*, which uses the Fisher information matrix similarly to a precision matrix for a Normal distribution up to a scalar variance factor was proposed by Zellner (1986) for application in the standard Gaussian linear model. This setup is well-known as *g-prior* method. Such an approach was extended for the class of generalized linear models by Gupta and Ibrahim (2009) in the context of high dimensional data. In this method, the information matrix captures the *prior* covariance between parameters via Fisher information, which seems an attractive specification since this matrix plays a major role in the determination of large sample covariance in both Bayesian and frequentist inference.

We shall consider, for the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, two independent multivariate Normal *prior* distributions. Let $\overline{q}_j = q_j + 1$ $(j = 1,2)$. Hence,

$$\boldsymbol{\beta}_j \sim \mathcal{N}_{\overline{q}_j}\left[\boldsymbol{\beta}_j^p, \tau_j \boldsymbol{\Sigma}\left(\boldsymbol{\beta}_j^p\right)\right],$$

where $\boldsymbol{\Sigma}(\boldsymbol{\beta}_j^p)$ is equivalent to $\boldsymbol{\mathcal{K}}_j^{-1}$ evaluated at $\boldsymbol{\beta}_j^p$ and $\tau_j \in \mathbb{R}_+$ is assumed known.

It can be proved that, if the data is discrete, then the use of a proper *prior* distribution (multivariate Normal in our case) avoids the *posterior* density to be improper. The *prior* vectors $\boldsymbol{\beta}_1^p$ and $\boldsymbol{\beta}_2^p$ can be chosen arbitrarily if no specialized information is available. The covariance structure is obtained from the observed information, which means that an additional computational effort is necessary to compute such matrix. On the other hand, if we choose constant precision matrices, not depending on $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, then the information matrix *priors* will always be Gaussian *priors*. As $\tau_1 \to \infty$ and $\tau_2 \to \infty$, such *priors* converge to *Jeffreys priors* given by $\pi(\boldsymbol{\beta}_1) \propto |\boldsymbol{\Sigma}(\boldsymbol{\beta}_1^p)|^{-1/2}$ and $\pi(\boldsymbol{\beta}_2) \propto |\boldsymbol{\Sigma}(\boldsymbol{\beta}_2^p)|^{-1/2}$, respectively. These densities can be seen as particular cases of *hyper-g prior* density proposed by Bové and Held (2011), where τ_1 and τ_2 should be estimated with model parameters.

The Bayesian approach for the model (3.11) can be performed by writing the unnormalized joint *posterior* density of the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ as

$$\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{x}) + \ell_2(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{z})\}\, \pi_1(\boldsymbol{\beta}_1)\, \pi_2(\boldsymbol{\beta}_2), \tag{3.17}$$

and, due to orthogonality between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, we have

$$\pi(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{x}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{x})\}\, \pi_1(\boldsymbol{\beta}_1) \quad \text{and} \quad \pi(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{z}) \propto \exp\{\ell_2(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{z})\}\, \pi_2(\boldsymbol{\beta}_2), \tag{3.18}$$

where ℓ_1 and ℓ_2 are given by (3.15) and (3.16), respectively.

From the Bayesian point of view, inference for the parameters is based on their marginal *posterior* densities, which can be obtained by integrating the joint *posterior* distributions in (3.18). Clearly, these densities have unknown forms mainly due to the complexity of the respective likelihood functions. In this case, Bayesian estimates for each element of $\boldsymbol{\beta}_j$ can be obtained by applying the Random-walk Metropolis algorithm, which is an iterative procedure of a broad class of MCMC methods. Through this procedure, $\overline{q}_j$ chains can be generated for $\boldsymbol{\beta}_j$. In fact, the dimensionality issue will depend on how much covariates will be taken under consideration to modeling the parameters of $\mathcal{ZMPS}$ model. Following Zelner's *g-prior* approach, we will consider a multivariate Normal specification for the proposal (candidate-generating) distribution in the algorithm. This density is used as the main term in the transition *kernel* when computing the acceptance probability (α). Thus, for any state $k \geqslant 0$, the MCMC simulations are performed proposing a candidate $\boldsymbol{\psi}_j$ for $\boldsymbol{\beta}_j$ by generating $\boldsymbol{\psi}_j | \boldsymbol{\beta}_j^* \sim \mathcal{N}_{\overline{q}_j}[\boldsymbol{\beta}_j^*, \nu_j \boldsymbol{\Sigma}(\boldsymbol{\beta}_j^*)]$, where $\boldsymbol{\beta}_j^* = \nu_j \boldsymbol{\beta}_j^{(k)} + (1 - \nu_j)\boldsymbol{\beta}_j^p$ and $\nu_j = \tau_j(1 + \tau_j)^{-1}$. Notice that transitions depend on the acceptance of pseudo-random vectors generated with mean given by the mixture between the actual state of the chain and the *priors* specification.

Algorithm 3 can be used to generate chains for the model parameters using the Random-walk Metropolis method. To run the algorithm, initial conditions $\boldsymbol{\beta}_1^{(0)}$ and $\boldsymbol{\beta}_2^{(0)}$ are needed and $\boldsymbol{\beta}_1^p$ and $\boldsymbol{\beta}_2^p$ must be specified. The parameters τ_1 and τ_2 are chosen by monitoring the acceptance rate of the algorithm. For a specific asymptotic Gaussian environment, it is well-known that the optimal acceptance rate is 23.40%, but, in our case, we consider rates reaching 40% as reasonable. This procedure generates a sample of size $N+1$ for each parameter. The convergence of the algorithm can be monitored by the Geweke criterion (GEWEKE, 1992). After convergence, some generated samples are discarded as burn-in. The procedure to decrease the correlation between and within generated chains is the usual approach of getting thinned steps. The final sample is supposed to have a size of M. A summary of the *posterior* densities can be written considering the MCMC estimates.

Algorithm 3 – Random-walk Metropolis

procedure METHAS($N, \boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}, \boldsymbol{\beta}_1^p, \boldsymbol{\beta}_2^p, \tau_1, \tau_2$)
  Set $k \leftarrow 0$
  Set $\nu_1 \leftarrow \tau_1(1+\tau_1)^{-1}$ and $\nu_2 \leftarrow \tau_2(1+\tau_2)^{-1}$
  while $k < N$ **do**
    Set $\boldsymbol{\beta}_1^* \leftarrow \nu_1 \boldsymbol{\beta}_1^{(k)} + (1-\nu_1)\boldsymbol{\beta}_1^p$ and $\boldsymbol{\beta}_2^* \leftarrow \nu_2 \boldsymbol{\beta}_2^{(k)} + (1-\nu_2)\boldsymbol{\beta}_2^p$
    Generate $\boldsymbol{\psi}_1 \sim \mathcal{N}_{\bar{q}_1}[\boldsymbol{\beta}_1^*, \nu_1 \boldsymbol{\Sigma}(\boldsymbol{\beta}_1^*)]$ and $\boldsymbol{\psi}_2 \sim \mathcal{N}_{\bar{q}_2}[\boldsymbol{\beta}_2^*, \nu_2 \boldsymbol{\Sigma}(\boldsymbol{\beta}_2^*)]$
    Set $\alpha_1 \leftarrow \min\{1, \exp[\ell_1(\boldsymbol{\psi}_1; \boldsymbol{y}, \boldsymbol{x}) - \ell_1(\boldsymbol{\beta}_1^{(k)}; \boldsymbol{y}, \boldsymbol{x})]\}$
    Set $\alpha_2 \leftarrow \min\{1, \exp[\ell_2(\boldsymbol{\psi}_2; \boldsymbol{y}, \boldsymbol{z}) - \ell_2(\boldsymbol{\beta}_2^{(k)}; \boldsymbol{y}, \boldsymbol{z})]\}$
    Set $\boldsymbol{\beta}_1^{(k+1)} \leftarrow \boldsymbol{\beta}_1^*$ and $\boldsymbol{\beta}_2^{(k+1)} \leftarrow \boldsymbol{\beta}_2^*$
    Generate $u_1, u_2 \sim \mathcal{U}(0,1)$
    if $u_1 \leqslant \alpha_1$ and $u_2 \leqslant \alpha_2$ **then**
      Set $\boldsymbol{\beta}_1^{(k+1)} \leftarrow \boldsymbol{\psi}_1$ and $\boldsymbol{\beta}_2^{(k+1)} \leftarrow \boldsymbol{\psi}_2$
    end if
    Set $k \leftarrow k+1$
  end while
  return $\{\boldsymbol{\beta}^{(k)}\}_{k=1}^N$
end procedure

In Sections 3.7 and 3.8 we present, respectively, the results obtained in the simulation study and the application of the proposed model to a real dataset. To attain the numerical results, all computations were performed under the R environment (R Development Core Team, 2017).

3.6.2 Influential points

The computation of divergence measures between *posterior* distributions is an useful way to quantify influence. According to Csiszár (1967), the q-divergence measure between two densities π_1 and π_2 for a vector of parameters $\boldsymbol{\beta}$ is defined by

$$d_q(\pi_1, \pi_2) = \int_{\mathcal{D}} \pi_2(\boldsymbol{\beta})\, q\left[\frac{\pi_1(\boldsymbol{\beta})}{\pi_2(\boldsymbol{\beta})}\right] \mathrm{d}\boldsymbol{\beta},$$

where $\mathcal{D}$ is the cartesian product defining the domain of $\boldsymbol{\beta}$ and q is a convex function such that $q(1) = 0$. Each choice q leads us to different divergence measures. The well-known Kullback-Leibler (KL) divergence is obtained by letting $q(t) = -\log(t)$. A symmetric version of KL divergence is obtained when $q(t) = (t-1)\log(t)$ and the L_1-distance divergence is obtained when $q(t) = |t-1|$. See Cho *et al.* (2009) and Garay *et al.* (2015) for further details about the q-divergence.

Let $\boldsymbol{y}_{(-i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ be the response vector after removal of the i-th observation of $\boldsymbol{y}$. Also, let $\boldsymbol{x}_{(-i)}$ and $\boldsymbol{z}_{(-i)}$ be the vectors of covariates having the same form of $\boldsymbol{y}_{(-i)}$. For ease of notation, let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ be the vector containing all the parameters of the regression model (3.11). Using the full vector of observations, the *posterior* density of the vector $\boldsymbol{\beta}$ can be denoted by $\pi(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$, as stated in Equation (3.17). On the other hand, using the vector without the i-th observation, the *posterior* density of $\boldsymbol{\beta}$ can be denoted by $\pi(\boldsymbol{\beta}; \boldsymbol{y}_{(-i)}, \boldsymbol{x}_{(-i)}, \boldsymbol{z}_{(-i)})$. In this chapter, we will consider the KL divergence between $\pi(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$ and $\pi(\boldsymbol{\beta}; \boldsymbol{y}_{(-i)}, \boldsymbol{x}_{(-i)}, \boldsymbol{z}_{(-i)})$ to assess the local influence of the observation y_i. This measure can be evaluated as

$$\mathrm{KL}\left(\pi, \pi_{(-i)}\right) = \int_{\mathcal{D}} \pi\left(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}\right) \log\left[\frac{\pi\left(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}\right)}{\pi\left(\boldsymbol{\beta}; \boldsymbol{y}_{(-i)}, \boldsymbol{x}_{(-i)}, \boldsymbol{z}_{(-i)}\right)}\right] \mathrm{d}\boldsymbol{\beta}, \tag{3.19}$$

where the region of integration is $\mathcal{D} = \mathbb{R}^{\bar{q}_1} \times \mathbb{R}^{\bar{q}_2}$. The term involving the logarithm of the ratio of the two *posterior* densities can be rewritten as

$$\begin{aligned} \log\left[\frac{\pi\left(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}\right)}{\pi\left(\boldsymbol{\beta}; \boldsymbol{y}_{(-i)}, \boldsymbol{x}_{(-i)}, \boldsymbol{z}_{(-i)}\right)}\right] &= \log\left[f_*\left(y_i; \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}\right)\right] - \\ &\quad \log\left[\int_{\mathcal{D}} f_*\left(y_i; \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}\right) \pi\left(\boldsymbol{\beta}; \boldsymbol{y}_{(-i)}, \boldsymbol{x}_{(-i)}, \boldsymbol{z}_{(-i)}\right) \mathrm{d}\boldsymbol{\beta}\right], \end{aligned}$$

where the integral in the right side corresponds to the conditional predictive ordinate (CPO) density of the observation y_i. The (3.19) can be expressed as

$$\mathrm{KL}\left(\pi, \pi_{(-i)}\right) = \mathbb{E}_{\boldsymbol{\beta}}\left\{\log\left[f_*\left(y_i; \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}\right)\right]\right\} - \log\left(\mathrm{CPO}_i\right),$$

being the expected value obtained with respect to $\pi(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$. Thus, given the set $\{\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^M\}$ of generated values from the *posterior* density (3.17), one can estimate the effect of the observation y_i as

$$\widehat{\mathrm{KL}}\left(\pi, \pi_{(-i)}\right) = \frac{1}{M}\sum_{k=1}^{M} \log\left[f_*\left(y_i; \boldsymbol{x}_i, \boldsymbol{z}_i, \hat{\boldsymbol{\beta}}^{(k)}\right)\right] - \log\left(\widehat{\mathrm{CPO}}_i\right),$$

where the CPO can be estimated as

$$\widehat{\mathrm{CPO}}_i = \left[\frac{1}{M}\sum_{k=1}^{M} \frac{1}{f_*\left(y_i; \boldsymbol{x}_i, \boldsymbol{z}_i, \hat{\boldsymbol{\beta}}^{(k)}\right)}\right]^{-1}. \tag{3.20}$$

A measure of calibration for the KL divergence was proposed by McCulloch (1989). Denoted by ρ_i, this measure is derived from the solution of the equation

$$\begin{aligned} \mathrm{KL}\left(\pi, \pi_{(-i)}\right) &= \mathrm{KL}\left[\mathcal{B}\left(0.5\right), \mathcal{B}\left(\rho_i\right)\right] \\ &= -\frac{\log\left[4\rho_i\left(1-\rho_i\right)\right]}{2}, \end{aligned} \tag{3.21}$$

where $\mathcal{B}(\rho_i)$ denotes the Bernoulli distribution with probability of success ρ_i. This implies that describing results using the full *posterior* density instead of that one removing the i-th observation is equivalent to describe a not observed event as having probability ρ_i, when the correct probability is 0.50. Now, solving Equation (3.21) for ρ_i, we have that

$$\rho_i = \frac{1}{2}\left\{1 + \sqrt{1 - \exp\{-2\mathrm{KL}\left(\pi, \pi_{(-i)}\right)\}}\right\},$$

which implies that $0.50 \leqslant \rho_i \leqslant 1$. For $\rho_i \gg 0.50$, it can be considered that the i-th observation is an influential point, and its removal can provide a more realistic fit of the $\mathcal{ZMPS}_h$ regression model.

3.6.3 Residual analysis

The residual analysis is defined by a set of methods used to investigate the suitability of a regression model. Here, we consider an approach based on the randomized quantile residuals proposed by Dunn and Smyth (1996), which is a particular case of Cox and Snell crude residuals (COX; SNELL, 1968), to evaluate the suitability of the $\mathcal{ZMPS}_h$ regression model for a given dataset.

Let $\mathrm{F}_*(y;\mu,p)$ be the cdf given by (3.7). Since F_* is not continuous, a general definition of quantile residuals is required. In this way, let us assume $a_i = \lim_{y\uparrow y_i}\mathrm{F}_*(y;\hat{\mu}_i,\hat{p}_i)$ and $b_i = \mathrm{F}_*(y_i;\hat{\mu}_i,\hat{p}_i)$ for the i-th observation of $\boldsymbol{y}$. Hence, the randomized quantile residuals for each y_i can be defined as

$$r_i = \Phi^{-1}\left(u_i\right),$$

where $\Phi(\cdot)$ is the cdf of the standard Normal distribution and u_i is a Uniform random variable defined on $(a_i, b_i]$. The main feature of such type of residuals is that the r_i are exactly standard Normal, regardless of the sample variability in $\hat{\mu}_i$ and $\hat{p}_i$. After model fitting, the normality assumption can be verified by using graphic techniques (e.g., the Normal *QQ-plot*) as well as using some well-known adherence tests.

3.7 Simulation study

In this section we seek to evaluate the performance of the proposed Bayesian approach by performing a simulation study. The simulation process consists of generating 500 independent samples of sizes $n = 50, 100, 200, 500,$ and $1{,}000$ of a random variable Y having $\mathcal{ZMPS}_h$

distribution under the regression framework (3.11). For the entirely process, we consider a $n \times 2$ regression matrix $\boldsymbol{X}$ in which the first column consists of 1's and the second consists of a fixed variable generated from an Uniform distribution on the unit interval. We also consider $\boldsymbol{Z} = \boldsymbol{X}$. Moreover, we assign different values for the vectors of parameters $\boldsymbol{\beta}_1^\top = (\beta_{10}, \beta_{11})$ and $\boldsymbol{\beta}_2^\top = (\beta_{20}, \beta_{21})$ in order to generate both zero-inflated and zero-deflated samples. To relate μ and ω with the their respective linear predictors, we choose the logarithm link function and the Equation (3.12), respectively. Each one of these cases are treated separately in the following subsections.

Algorithm 4 can be used to generate a single pseudo-random realization from the $\mathcal{ZMPS}_h$ distribution in the regression framework. The process to generate a random sample consists of running the algorithm as often as necessary, say n times. The sequential-search is a black-box type of algorithm and works with any computable probability vector. The main advantage of such a procedure is its simplicity. On the other hand, sequential-search algorithms may be slow as the while-loop may have to be repeated very often. More information on this algorithm can be found at Hörmann, Leydold and Derflinger (2013).

Algorithm 4 – Sequential-Search

1: **procedure** SEQSEA($\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$)
2: Generate $x, u \sim \mathcal{U}(0,1)$
3: Set $\mu \leftarrow \exp\{\beta_{10} + \beta_{11} x\}$ and $\omega \leftarrow [1 + \exp\{-(\beta_{20} + \beta_{21} x)\}]^{-1}$
4: Set $y \leftarrow 0$ and $k \leftarrow (1 - \omega)$
5: **while** $u > k$ **do**
6: Set $y \leftarrow y + 1$ and $k \leftarrow k + \omega f(y; \mu)$
7: **end while**
8: **return** y
9: **end procedure**

Under the $\mathcal{ZMPS}_h$ distribution, the expected number of iterations (NI), that is, the expected number of comparisons in the while condition, is given by

$$\begin{aligned} \mu_{\text{NI}} = \mu_* + 1 &= p\mu + 1 \\ &= \frac{\omega(\theta+1)^2(\theta^2+2)}{\theta(\theta^3+\theta^2+2\theta+1)} + 1 \\ &= \frac{\omega[h(\mu)+1]^2[h^2(\mu)+2]}{h(\mu)[h^3(\mu)+h^2(\mu)+2h(\mu)+1]} + 1, \end{aligned}$$

where $h(\mu)$ is given by the Equation (3.4).

To apply the Bayesian approach in each case, we consider the Random-walk Metropolis algorithm for MCMC sampling. A total of $N = 50,000$ values were generated for each parameter, considering a burn-in of 20% of the size of the chain. Then, the chains were diagnosed using the Geweke convergence method, revealing their stationarity. To obtain pseudo-independent samples

from the *posterior* distributions given in (3.18), one out every 10 generated values were kept, resulting in chains of size $M = 4,000$ for each parameter. We consider the *posterior* mean as a Bayesian estimator and evaluate its performance by its bias (B), its variance (Var), its mean squared error (MSE), and its mean absolute percentage error (MAPE). Also, we compute the coverage probability (CP) of the Bayesian Credible Intervals (BCIs) obtained for each parameter. Using the generated samples and assuming $\phi = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the measures of interest are given by

$$\mathrm{B}\left(\hat{\phi}\right) = \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\phi}_j - \phi\right), \quad \mathrm{MSE}\left(\hat{\phi}\right) = \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\phi}_j - \phi\right)^2 \quad \text{and} \quad \mathrm{MAPE}\left(\hat{\phi}\right) = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{\hat{\phi}_j - \phi}{\phi}\right|.$$

The variance of $\hat{\phi}$ can be estimated as the difference between the MSE and the square of the bias. Moreover, the coverage probability of the BCIs is computed as follows

$$\mathrm{CP}_{\%}\left(\phi\right) = \frac{100}{M}\sum_{j=1}^{M}\delta_j\left(\phi\right),$$

where $\delta_j(\phi)$ assumes 1 if the j-th BCI contains the true value ϕ and 0 otherwise. In addition, we estimated the below and the above non-coverage probability of the BCIs, denoted by BNCP and ANCP, respectively. These measures are computed analogously to the CP. The BNCP and the ANCP may be useful measures to determine asymmetric features since they provide the probabilities of finding the true value of the parameter ϕ on the tails of the generated *posterior* density.

The *priors* were chosen in order to ensure that parameter p provides zero inflation or zero deflation depending on the case. This was made by applying the *prior* values on the linear predictors of μ and ω. Also, the estimation procedures were performed using $\tau_1 = \tau_2 = 5$, which provided acceptance rates of approximately 40%. Indeed, higher values of τ_1 and τ_2 provide lower acceptance rates in the Random-walk Metropolis algorithm, resulting in higher variances of the Bayesian estimates. However, the point estimates of the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ (*posterior* expected values) were often close to the true values.

We noticed in the simulation study that the parameter estimates become more accurate, that is, the estimated bias and mean squared errors decrease with increasing sample size. Although high MAPE values were obtained for some parameters, this does not compromise accuracy in estimation. For example, in Table 14 (Scenario 2), for $n = 1,000$, we observed a MAPE of approximately 22% for β_{20}. Taking into account that the true value of this parameter is -0.5, we have $\hat{\beta}_{11} = 1.22 \times (-0.5) = -0.61$, which does not represent such a huge impact on the final result. Moreover, the estimated coverage probability of the credible intervals (Table 16), although not reach the 95% expected, also improves as the sample size increases. In Table 15 (Scenario 2),

Table 14 – Empirical properties of the Bayesian estimators using zero-inflated samples.

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	β_{10}	−0.0145	0.0595	0.0597	6.3957
	β_{11}	0.0066	0.3036	0.3036	87.4904
	β_{20}	0.3090	1.1218	1.2173	36.6655
	β_{21}	−0.5021	2.7551	3.0073	49.4452
100	β_{10}	−0.0088	0.0441	0.0442	5.5116
	β_{11}	0.0022	0.1594	0.1594	64.1307
	β_{20}	0.0933	0.3495	0.3582	22.9741
	β_{21}	−0.1264	0.9148	0.9308	30.3051
200	β_{10}	−0.0062	0.0148	0.0149	3.2738
	β_{11}	0.0043	0.0635	0.0635	41.0374
	β_{20}	0.0527	0.1435	0.1463	14.7141
	β_{21}	−0.0793	0.3915	0.3978	19.6125
500	β_{10}	−0.0007	0.0051	0.0051	1.8934
	β_{11}	−0.0039	0.0204	0.0204	22.5212
	β_{20}	0.0168	0.0546	0.0549	9.3431
	β_{21}	−0.0135	0.1389	0.1391	11.9998
1,000	β_{10}	0.0014	0.0030	0.0030	1.4645
	β_{11}	−0.0043	0.0110	0.0110	17.1100
	β_{20}	0.0140	0.0273	0.0275	6.5848
	β_{21}	−0.0101	0.0658	0.0659	8.1402
			Scenario 2		
50	β_{10}	−0.0400	0.2006	0.2022	14.0490
	β_{11}	−0.0157	0.9723	0.9726	50.3510
	β_{20}	0.0398	0.4241	0.4257	102.0528
	β_{21}	−0.3012	1.9653	2.0561	111.7261
100	β_{10}	−0.0133	0.1085	0.1087	10.0876
	β_{11}	−0.0529	0.3491	0.3519	31.1519
	β_{20}	0.0161	0.2714	0.2716	83.3754
	β_{21}	−0.1197	0.9490	0.9633	78.0873
200	β_{10}	−0.0189	0.0373	0.0377	6.1515
	β_{11}	−0.0016	0.1422	0.1422	19.9918
	β_{20}	0.0106	0.0882	0.0884	47.5292
	β_{21}	−0.0604	0.3518	0.3554	47.7006
500	β_{10}	−0.0039	0.0132	0.0133	3.6626
	β_{11}	−0.0271	0.0450	0.0457	11.5067
	β_{20}	−0.0083	0.0365	0.0365	30.6836
	β_{21}	−0.0058	0.1273	0.1274	28.6547
1,000	β_{10}	0.0030	0.0072	0.0072	2.6809
	β_{11}	−0.0277	0.0248	0.0256	8.3709
	β_{20}	0.0051	0.0189	0.0189	22.0800
	β_{21}	−0.0153	0.0645	0.0647	20.6595

Source: Elaborated by the author.

Table 15 – Empirical properties of the Bayesian estimators using zero-deflated samples.

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	β_{10}	−0.1875	0.6136	0.6487	60.4564
	β_{11}	0.1005	1.9180	1.9281	216.2502
	β_{20}	0.0826	0.5250	0.5319	111.3466
	β_{21}	−0.0093	2.1304	2.1305	115.2306
100	β_{10}	−0.0758	0.2932	0.2990	41.6854
	β_{11}	0.0277	0.7570	0.7578	136.3924
	β_{20}	−0.0164	0.2562	0.2565	78.1795
	β_{21}	0.1120	0.8945	0.9070	73.7165
200	β_{10}	−0.0554	0.1154	0.1185	27.2801
	β_{11}	0.0442	0.3145	0.3164	89.2993
	β_{20}	0.0053	0.0979	0.0980	50.0958
	β_{21}	0.0340	0.3561	0.3573	47.6602
500	β_{10}	−0.0282	0.0394	0.0401	15.9299
	β_{11}	0.0243	0.1017	0.1023	50.7572
	β_{20}	0.0016	0.0343	0.0343	29.0094
	β_{21}	0.0164	0.1282	0.1284	27.3856
1,000	β_{10}	−0.0139	0.0215	0.0217	11.4679
	β_{11}	0.0128	0.0543	0.0545	37.1224
	β_{20}	0.0079	0.0201	0.0202	22.6280
	β_{21}	−0.0005	0.0695	0.0695	20.6529
			Scenario 2		
50	β_{10}	−0.4934	2.3634	2.6068	53.6098
	β_{11}	0.1431	8.5660	8.5865	403.6466
	β_{20}	0.1726	0.5138	0.5436	57.2098
	β_{21}	−0.2159	1.6698	1.7164	204.9714
100	β_{10}	−0.2428	0.9398	0.9987	35.0369
	β_{11}	0.1028	2.3097	2.3203	228.8899
	β_{20}	0.0163	0.2716	0.2719	40.5817
	β_{21}	0.0333	0.8099	0.8110	143.6478
200	β_{10}	−0.1359	0.3124	0.3309	21.7213
	β_{11}	0.0710	0.9130	0.9181	151.6749
	β_{20}	0.0200	0.1089	0.1093	26.0174
	β_{21}	−0.0058	0.3208	0.3209	90.7417
500	β_{10}	−0.0522	0.0834	0.0861	11.5717
	β_{11}	0.0301	0.2271	0.2280	76.0097
	β_{20}	0.0055	0.0364	0.0365	15.2508
	β_{21}	0.0052	0.1059	0.1059	51.9439
1,000	β_{10}	−0.0204	0.0473	0.0477	8.5717
	β_{11}	0.0076	0.1336	0.1337	57.5964
	β_{20}	0.0132	0.0199	0.0201	11.4299
	β_{21}	−0.0104	0.0544	0.0545	37.7689

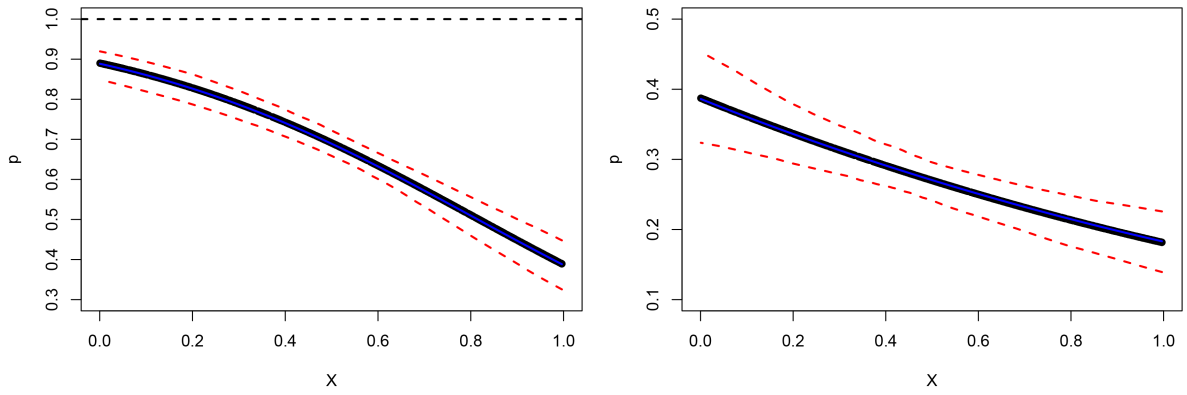
Source: Elaborated by the author.

Table 16 – Coverage probabilities (%) of the BCIs using zero-inflated samples.

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	3.40	92.40	4.20	4.40	90.60	5.00
	β_{11}	4.60	93.00	2.40	4.40	91.20	4.40
	β_{20}	8.00	90.40	1.60	5.20	92.20	2.60
	β_{21}	2.20	91.20	6.60	2.80	92.40	4.80
100	β_{10}	3.40	91.80	4.80	4.40	92.80	2.80
	β_{11}	4.40	91.80	3.80	3.00	93.80	3.20
	β_{20}	4.80	92.80	2.40	5.00	92.60	2.40
	β_{21}	3.40	91.60	5.00	3.80	91.60	4.60
200	β_{10}	3.40	92.80	3.80	3.60	92.40	4.00
	β_{11}	3.00	92.60	4.40	2.80	94.20	3.00
	β_{20}	3.80	93.40	2.80	3.60	95.00	1.40
	β_{21}	2.60	94.40	3.00	2.40	92.20	5.40
500	β_{10}	3.20	93.80	3.00	3.00	93.40	3.60
	β_{11}	3.80	92.20	4.00	2.20	93.40	4.40
	β_{20}	4.60	91.80	3.60	3.20	93.80	3.00
	β_{21}	4.60	91.60	3.80	4.00	92.00	4.00
1,000	β_{10}	3.40	93.80	2.80	3.00	93.40	3.60
	β_{11}	2.60	93.80	3.60	2.00	92.80	5.20
	β_{20}	3.40	92.00	4.60	3.40	93.20	3.40
	β_{21}	2.80	94.40	2.80	3.60	92.00	4.40

Source: Elaborated by the author.

Figure 5 – *Posterior* estimates of parameter p using zero-inflated samples ($n = 1,000$).



(a) Scenario 1 (b) Scenario 2

Source: Elaborated by the author.

also for $n = 1,000$, we observed a MAPE of approximately 58% for β_{11}. Since the true value, in this case, is 0.5, we have $\hat{\beta}_{11} = 1.58 \times 0.5 = 0.79$, which would also not have much impact on

the model fit. Therefore, our simulation study indicates the feasibility of the proposed Bayesian approach to the use of the $\mathcal{ZMPS}_h$ regression model in the analysis of overdispersed and zero-modified real datasets.

Table 17 – Coverage probabilities (%) of the BCIs using zero-deflated samples.

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	2.60	92.20	5.20	2.60	92.80	4.60
	β_{11}	4.20	92.80	3.00	4.60	90.80	4.60
	β_{20}	6.20	89.60	4.20	5.80	92.80	1.40
	β_{21}	3.20	92.20	4.60	2.40	92.00	5.60
100	β_{10}	3.80	92.40	3.80	3.80	88.60	7.60
	β_{11}	4.80	91.60	3.60	6.80	88.60	4.60
	β_{20}	2.60	94.00	3.40	3.80	91.80	4.40
	β_{21}	4.80	93.00	2.20	4.40	91.40	4.20
200	β_{10}	2.80	93.40	3.80	4.20	90.20	5.60
	β_{11}	3.60	93.00	3.40	5.20	90.40	4.40
	β_{20}	3.60	92.80	3.60	4.20	91.40	4.40
	β_{21}	2.60	93.20	4.20	3.80	92.20	4.00
500	β_{10}	2.80	93.20	4.00	2.60	93.20	4.20
	β_{11}	3.80	93.40	2.80	4.80	92.40	2.80
	β_{20}	2.40	93.60	4.00	2.80	93.40	3.80
	β_{21}	4.20	92.20	3.60	4.80	93.00	2.20
1,000	β_{10}	4.00	91.20	4.80	3.80	91.00	5.20
	β_{11}	3.60	93.20	3.20	4.60	91.80	3.60
	β_{20}	4.80	91.60	3.60	2.80	94.20	3.00
	β_{21}	4.60	91.20	4.20	2.80	94.20	3.00

Source: Elaborated by the author.

3.7.1 Zero-inflated artificial data

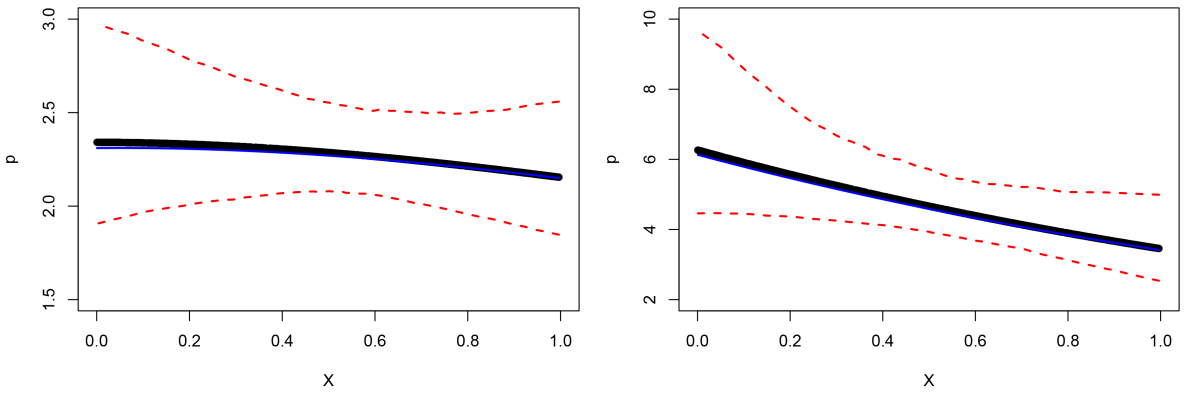
For the zero-inflated case, the samples were generated from a $\mathcal{ZMPS}_h$ distribution considering that $0 < p < 1$. Here, the parameters were chosen by taking into account that zero-inflated samples have, naturally, proportion of zeros greater than expected and therefore, the variable Y was generated with mean (μ) not even close to zero. Thus, in the first scenario, we assign the values $\boldsymbol{\beta}_1^\intercal = (3.0, -0.5)$ and $\boldsymbol{\beta}_2^\intercal = (2.0, -2.5)$ to the parameter vectors and perform the simulation process. In the following, the values were changed to $\boldsymbol{\beta}_1^\intercal = (2.5, 1.5)$ and $\boldsymbol{\beta}_2^\intercal = (-0.5, -1.0)$ and the procedure was repeated once again. The *prior* setup for these scenarios is $(\boldsymbol{\beta}_1^p)^\intercal = (1.0, 0.0)$ and $(\boldsymbol{\beta}_2^p)^\intercal = (0.0, 1.0)$. Figure 5 depicts the *posterior* estimates obtained for parameter p using zero-inflated samples considering $n = 1{,}000$ for each scenario. The real values are represented by the blue straight lines and the 95% BCIs are represented by the red dashed lines. The estimated values for each generated observation are represented by the

filled black dots. In the adopted scenarios, the degree of zero inflation noticeably increases as x approaches to 1.

3.7.2 Zero-deflated artificial data

For the zero-deflated case, the samples were generated from a $\mathcal{ZMPS}_h$ distribution considering that $1 < p < [1 - f(0;\mu)]^{-1}$. Here, the parameters were chosen by taking into account that zero-deflated samples have, naturally, proportion of zeros smaller than expected and therefore, the variable Y was generated with mean (μ) close to zero. Thus, in the first scenario,we assign the values $\boldsymbol{\beta}_1^\intercal = (-1.0, 0.5)$ and $\boldsymbol{\beta}_2^\intercal = (0.5, 1.0)$ to the parameter vectors and perform the simulation process. In the following, the values were changed to $\boldsymbol{\beta}_1^\intercal = (-2.0, 0.5)$ and $\boldsymbol{\beta}_2^\intercal = (1.0, -0.5)$ and the procedure was repeated once again. The *prior* setup for these scenarios is $(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$ and $(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$. Figure 6 depicts the *posterior* estimates obtained for parameter p using zero-deflated samples considering $n = 1{,}000$ for each scenario. This representation has the same characteristics of Figure 5. Table 17 presents the estimated coverage probabilities of the BCIs considering the zero-deflated case.

Figure 6 – *Posterior* estimates of parameter p using zero-deflated samples $(n = 1{,}000)$.



(a) Scenario 1 (b) Scenario 2

Source: Elaborated by the author.

3.8 Fetal deaths data analysis

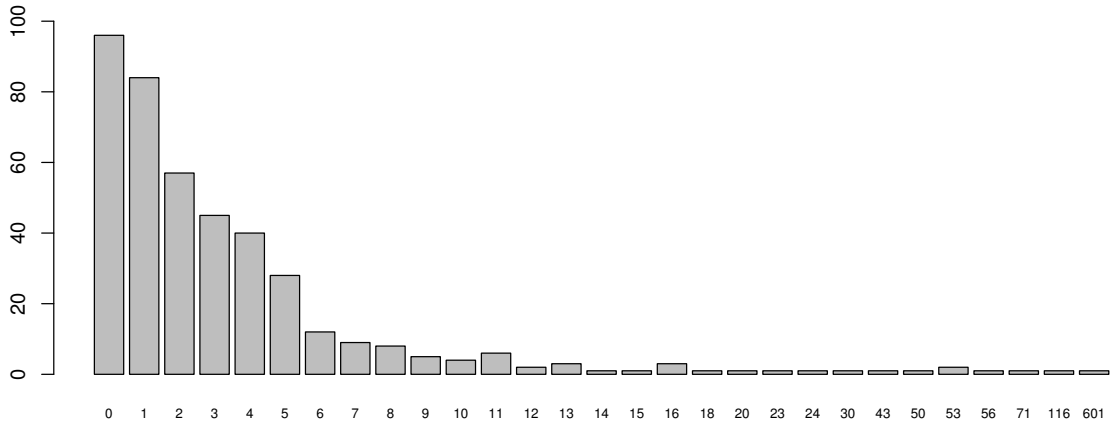
The application of the $\mathcal{ZMPS}_h$ regression model will be considered using a real dataset of the number of fetal deaths notifications reported in all the cities of Bahia State in the year of 2014 (data extracted from DATASUS repository (DATASUS, 2016)). Fetal death is defined as the death of a fetus before the expulsion or complete extraction of the mother's body, regardless of the gestation period. The death is verified when, after separation, the fetus does not breathe or present any sign of life, such as heartbeats, umbilical cord pulsations, effective movements, or voluntary contraction muscles. The main focus is to study the relationship of the mentioned

counts with the Human Development Index (HDI) assigned for each city. The HDI is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable, and have a decent standard of living. The HDI is computed as a geometric mean of normalized indices for each of these three dimensions.

According to the available information, Bahia State had 417 cities having HDI varying between 0.486 and 0.759 in 2014. The smaller index was assigned to Itapicuru and the upper one to the capital of Bahia State, Salvador. Cities having an HDI between 0.50 and 0.79 are considered as developing, and those ones with an HDI above 0.79 are considered developed. Excluding Itapirucu, the other 416 cities have an HDI within the range that ranks the developing cities. Indeed, the HDIs of these cities are comparable to those in some places in Africa and Central America (CONCEIÇÃO; ANDRADE; LOUZADA, 2013).

Now, let us characterize the number of fetal deaths notifications as the response variable (Y). From the observed dataset, there exists clear evidence that Y is overdispersed since the mean and the standard deviation are 5.47 and 30.62, respectively. The range of Y is 601, and its coefficient of variation is approximately 560%. The frequency of zeros is 96 (around 23% of the entire sample), which naively indicates zero inflation. The first and third quartiles of Y are respectively 1 and 4, showing that the number of notifications is very low for at least 75% of the cities in Bahia State. Moreover, we observed that only 29 out of 417 cities registered more than ten notifications over 2014. Figure 7 presents a barplot to illustrate the frequency distribution of the variable Y. A large amount of zeros in the dataset is a direct consequence of the lack of notifications of fetal deaths.

Figure 7 – Frequency distribution of the response variable.



Source: Elaborated by the author.

To fit the $\mathcal{ZMPS}_h$ regression model with HDI as a covariate, we adopted a similar procedure to that one used in the previous section. The logarithm link function was considered to relate μ with the linear predictor $\beta_{10} + \beta_{11}\text{HDI}$. To relate the parameter ω with the linear

predictor $\beta_{20} + \beta_{21}$HDI, we choose the link function given by (3.12). In this framework, the parameters β_{11} and β_{21} represent the effect of the HDI on the mean (μ) and the probability of zeros $(1-\omega)$, respectively. We considered the Random-walk Metropolis algorithm, generating a chain of size $N = 50,000$ for each parameter whereby the first 10,000 values were discarded as burn-in. The stationarity of each chain was evaluated with the Geweke criterion for the diagnostic of convergence. To obtain the pseudo-independent samples, one out every 10 generated values were kept, resulting in chains of size $M = 4,000$ for each parameter. For the *prior* distributions, we took $(\boldsymbol{\beta}_1^p)^\intercal = (1.0, 0.0)$, $(\boldsymbol{\beta}_2^p)^\intercal = (0.0, 1.0)$ and $\tau_1 = \tau_2 = 5$. Acceptance rates in the Random-walk Metropolis algorithm were at approximately 39%, and similar results were obtained to those observed in the simulation study, respect to the variability of the parameter estimates. For comparison purposes, identical procedures were adopted to fit the $\mathcal{P}$, the $\mathcal{PS}_h$ and the $\mathcal{ZMP}$ regression models using the same dataset.

Table 18 – *Posterior* parameter estimates and 95% BCIs from $\mathcal{ZMPS}_h$ fitted model (full dataset).

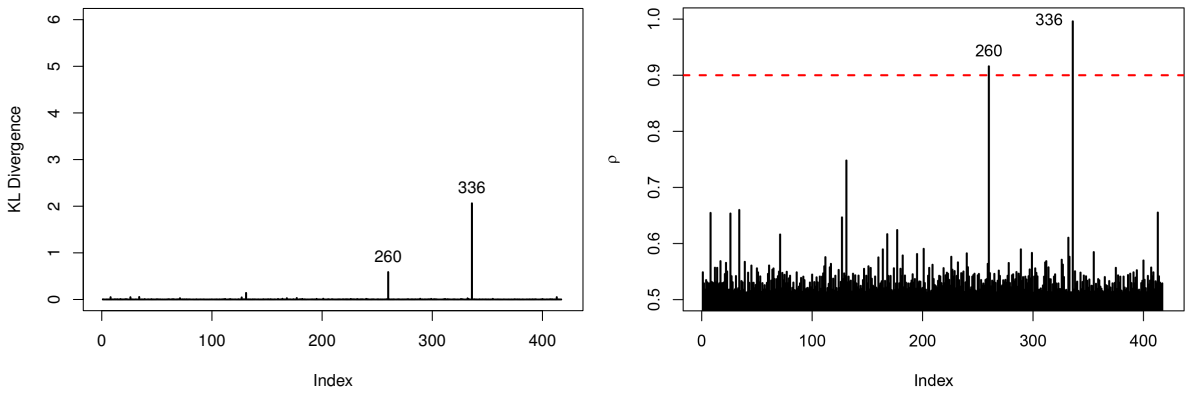
Parameter	Mean	Median	Std. Dev.	95% BCI	
				Lower	Upper
β_{10}	−10.9117	−10.9148	0.5786	−12.0842	−9.8028
β_{11}	20.1269	20.1224	0.9285	18.3625	22.0067
β_{20}	−5.6071	−5.5846	1.7747	−9.1571	−2.2021
β_{21}	11.5801	11.5178	3.0312	5.8298	17.6172

Source: Elaborated by the author.

Table 18 presents the mean, median, standard deviation obtained from the *posterior* distribution generated by the MCMC method for the $\mathcal{ZMPS}_h$ regression model. The 95% BCIs were estimated empirically from the generated samples. Given the sample size $(n = 417)$, the coverage probability is indeed high for the parameters. One can note that the BCIs obtained for β_{11} and β_{21} do not contain the value zero, which constitute the HDI as a relevant covariate to describe the number of notifications of fetal deaths. We verified that the assumption of normality for the generated chains is quite reasonable even in the presence of slightly heavy tails on the estimated marginal *posterior* densities of the parameters of the $\mathcal{ZMPS}_h$ regression model. In addition, there exists evidence of symmetry since *posterior* mean and median are very close to each other.

An analysis to verify the existence of influential points is presented in Figure 8. Figure 8a depicts the Kullback-Leibler divergence used to evaluate the effect of each observation on the parameter estimates. We consider an observation whose distance has a calibration exceeding 0.90 as an influential point. Based on Figure 8b, we observed the existence of two influential points, 260 and 336, corresponding to the cities Monte Santo and Salvador, respectively. In order to assess the influence of each observation, the inferential process was repeated considering three

Figure 8 – Sensitivity analysis for diagnostic of influential points.



(a) Kullback-Leibler divergence (b) Calibration (ρ)

Source: Elaborated by the author.

cases: removing only observation 260, removing only observation 336, and finally removing both the observations. The *posterior* summary for each case and variation percentage regarding the *posterior* summary obtained from the full dataset is presented in Table 20. After analyzing the parameter estimates in each case, it can be observed that the removal of observation 260 had a relative impact on the estimates of β_{10} and β_{11}, while the removal of observation 336 had a higher impact on these parameters. As the parameter μ was estimated using only the positive observations, we have that when more than zero notifications were registered, its expected value will be smaller than 4.00 on the cities with HDI $\leqslant 0.600$.

Table 19 – Comparison criteria for the fitted models.

Model	Full Dataset				Without Observation 260			
	DIC	EAIC	EBIC	LMPL	DIC	EAIC	EBIC	LMPL
$\mathcal{P}$	8754.11	3508.61	3516.67	−1809.86	8463.78	3392.44	3400.50	−1747.46
$\mathcal{PS}_h$	4825.27	1933.59	1941.66	−968.59	4734.94	1897.36	1905.42	−949.69
$\mathcal{ZMP}$	8429.38	3262.43	3278.57	−1699.91	8154.24	3146.39	3162.52	−1630.67
$\mathcal{ZMPS}_h$	4806.67	1924.61	1940.74	−962.59	4712.00	1886.59	1902.71	−942.89
Model	Without Observation 336				Without Observations 260 and 336			
	DIC	EAIC	EBIC	LMPL	DIC	EAIC	EBIC	LMPL
$\mathcal{P}$	6630.75	2658.44	2666.50	−1340.70	6412.46	2571.11	2579.17	−1296.09
$\mathcal{PS}_h$	4715.61	1889.51	1897.58	−944.61	4634.80	1857.10	1865.16	−928.02
$\mathcal{ZMP}$	6262.54	2459.07	2475.19	−1238.65	6069.23	2378.47	2394.58	−1197.78
$\mathcal{ZMPS}_h$	4712.26	1886.61	1902.73	−942.02	4627.73	1852.65	1868.76	−924.48

Source: Elaborated by the author.

Table 19 shows the DIC (Deviance Information Criterion), the EAIC (Expected Akaike Criterion, and the EBIC (Expected Bayesian Information Criterion) that were considered for comparison purposes. The main details about these measures are provided by Carlin and Louis

(1997). In addition, we computed the log-marginal pseudo-likelihood (LMPL) given by

$$\text{LMPL} = \sum_{i=1}^{n} \log\left(\widehat{\text{CPO}}_i\right),$$

where $\widehat{\text{CPO}}_i$ is the estimate of the conditional predictive ordinate (3.20) for the i-th observation in the sample. This measure also provides a tool to compare models being that larger values of LMPL indicate a better fit.

Table 20 – *Posterior* parameter estimates and 95% BCIs from $\mathcal{ZMPS}_h$ model (without influential points).

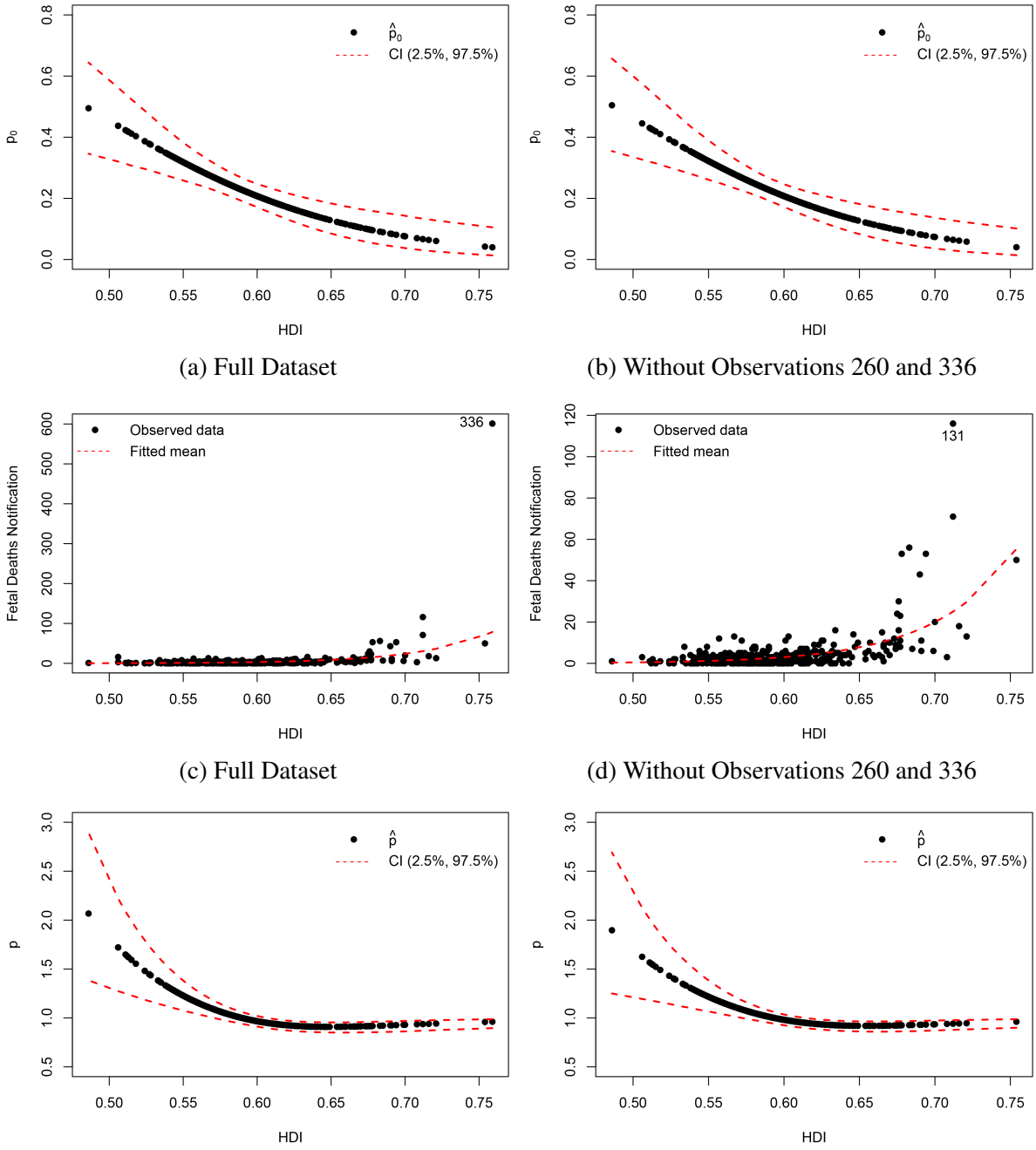
Removed Observation	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
{260}	β_{10}	−11.6423 (−6.70%)	−11.6500 (−6.74%)	0.6015 (3.96%)	−12.8374	−10.4662
	β_{11}	21.2509 (5.58%)	21.2665 (5.69%)	0.9664 (4.08%)	19.3859	23.2099
	β_{20}	−5.8668 (−4.63%)	−5.8316 (−4.42%)	1.8242 (2.79%)	−9.5516	−2.4080
	β_{21}	12.0156 (3.76%)	11.9457 (3.72%)	3.1186 (2.88%)	6.1113	18.3027
{336}	β_{10}	−9.3870 (13.97%)	−9.3763 (14.10%)	0.6410 (10.78%)	−10.7274	−8.1930
	β_{11}	17.5754 (−12.68%)	17.5569 (−12.75%)	1.0365 (11.63%)	15.6537	19.7368
	β_{20}	−5.5793 (0.50%)	−5.5974 (−0.23%)	1.7784 (0.21%)	−9.1765	−2.1177
	β_{21}	11.5285 (−0.45%)	11.5529 (0.30%)	3.0398 (0.28%)	5.6979	17.6675
{260, 336}	β_{10}	−10.2057 (6.47%)	−10.1962 (6.58%)	0.6877 (18.86%)	−11.5829	−8.8892
	β_{11}	18.8568 (−6.31%)	18.8472 (−6.36%)	1.1100 (19.55%)	16.7107	21.0679
	β_{20}	−5.8109 (−3.63%)	−5.7760 (−3.43%)	1.8239 (2.77%)	−9.4078	−2.3535
	β_{21}	11.9198 (2.93%)	11.8592 (2.96%)	3.1139 (2.73%)	6.0586	18.0467

Source: Elaborated by the author.

An important measure to be assessed is the probability of fetal deaths non-notification (p_0) for each city. In this context, it is worthwhile to mention that non-notification does not mean that the event of interest has not occurred in a particular city under investigation throughout the assessed year, but may indicate the inefficiency of the health system, which eventually does not

provides an adequate structure for monitoring all pregnant women (prenatal procedures). These probabilities were estimated and are presented in Figure 9, along with the *posterior* estimates for parameter p and the fitted mean, considering the $\mathcal{ZMPS}_h$ regression model. The presented estimates were obtained using the full dataset and that one without the influential points (260 and 336).

Figure 9 – *Posterior* estimates of parameters p_0, p, and μ.



(a) Full Dataset (b) Without Observations 260 and 336

(c) Full Dataset (d) Without Observations 260 and 336

(e) Full Dataset (f) Without Observations 260 and 336

Source: Elaborated by the author.

Based on Figures 9a and 9b one can notice that the estimated probabilities of fetal deaths

non-notification (p_0) are always smaller than 0.80. Also, it can be observed that the probability of non-notified cases of fetal deaths is very low in cities with high HDI. Figures 9c and 9d ilustrate the fitted average number of notifications per city. Noticeably, the uncertainty about the estimated mean is higher in the most developed cities. In addition, in Figures 9e and 9f we have observed that the estimates of p are really close to 1 for higher HDIs, ensuring that for these values, we can assume that the $\mathcal{PS}_h$ model $(p = 1)$ may fit well to such a subset. However, this result would require dividing the dataset into two subsets, since for most values of HDI, we have $\hat{p} > 1$, which characterizes the sample as having been generated by a zero-deflated distribution (the $\mathcal{ZDPS}_h$ in this case). This result highlights the importance of considering zero-modified models, which can be fitted without any previous information about the zero modification present in the dataset.

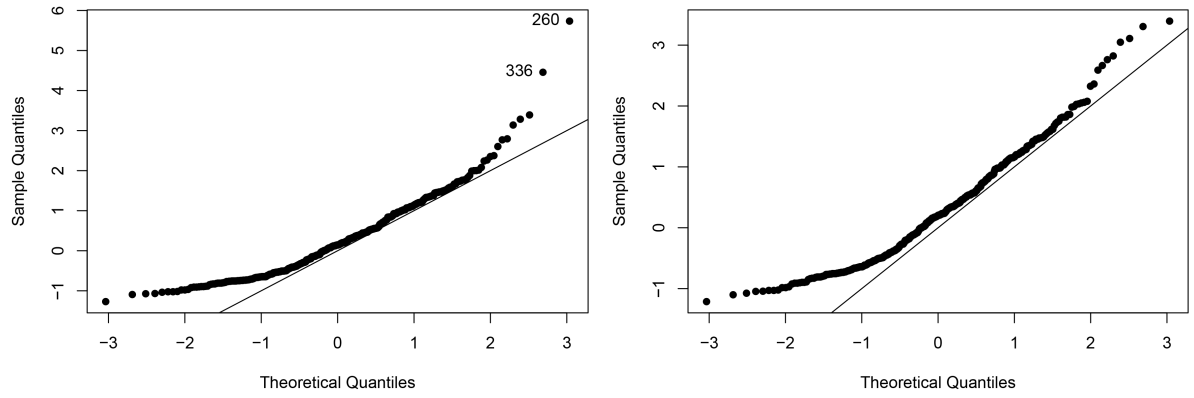
Table 21 presents the Bayesian estimates for the expected number of zeros (n_0), computed as $nf_*(0;\hat{\mu},\hat{p})$, where $\hat{\mu} = n^{-1}\sum_{i=1}^{n}\hat{\mu}_i$ and $\hat{p} = n^{-1}\sum_{i=1}^{n}\hat{p}_i$. The estimates for n_0, obtained from the $\mathcal{P}$ and the $\mathcal{PS}_h$ models, are much lower than the real one while those provided by zero-modified models are very close (or exactly) 96.

Table 21 – *Posterior* estimates of the expected number of zeros (n_0).

Model	Full Dataset	Without Obs. 260	Without Obs. 336	Without Obs. 260 and 336
$\mathcal{P}$	2	2	7	8
$\mathcal{PS}_h$	52	52	62	62
$\mathcal{ZMP}$	96	95	96	96
$\mathcal{ZMPS}_h$	96	96	96	96

Source: Elaborated by the author.

Figure 10 – Normal probability plot for the randomized quantile residuals.

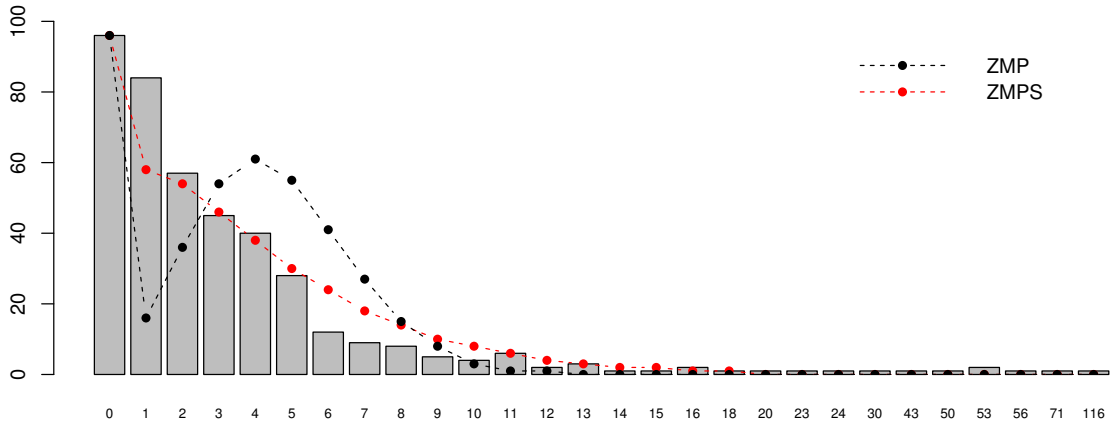


(a) Full Dataset (b) Without Observations 260 and 336

Source: Elaborated by the author.

Analyzing Figure 10 we observed a detachment of some quantile residual values on the lower tail with respect to the Normal straight line, which is expected when dealing with datasets with zero modification. This detachment highlights the inability of the model to describe very high counts. Such a feature can also be seen in Figure 11 when comparing the expected and the observed frequencies with values higher than 50 notifications.

Figure 11 – *Posterior* expected frequencies under zero-modified models.



Source: Elaborated by the author.

From the obtained results in this application, we conclude by pointing out that the use of a discrete distribution that accommodates an overdispersion level beyond that one caused by the zero modification in the dataset (e.g., $\mathcal{ZMP}$ and $\mathcal{ZMPS}_h$ models) may be necessary for the situation we get $p = 1$ (basis distribution) with data overdispersion explained only by the regression model.

3.9 Concluding remarks

In this chapter, we have introduced the $\mathcal{ZMPS}_h$ regression model that is useful to analyze zero-inflated/deflated data in the presence of overdispersion. Using the hurdle version of the $\mathcal{PS}_h$ distribution, it was possible to write separable likelihood functions for the parameter vectors, which led us to less complicated Bayesian procedures. In addition, we showed that the mean of the respective model could be estimated using only the positive observations. A particular case of q-divergence measure, the Kullback-Leibler divergence, was considered to evaluate local influence since such a task is very important to characterize the change in the frequency of zeros correctly. The proposed model was considered for the analysis of fetal death notifications, considering the HDI as a covariate. The response variable was found overdispersed and zero-inflated, which justifies the use of a zero-modified model. Model validation was addressed by using the randomized quantile residuals. From the Bayesian approach adopted to the $\mathcal{ZMPS}_h$ regression model, we conclude that the HDI is relevant to describe the mentioned counts. In

addition, following the DIC, the EAIC, the EBIC, and the LMPL criteria, the proposed model presented a better fit when compared with the $\mathcal{P}$, the $\mathcal{PS}_h$ and the $\mathcal{ZMP}$ regression models and therefore, can be as well considered when analyzing overdispersed and zero-modified real datasets.

CHAPTER

4

THE ZERO-MODIFIED POISSON-LINDLEY REGRESSION MODEL

4.1 Introduction

The primary goal of this chapter is to introduce the zero-modified Poisson-Lindley ($\mathcal{ZMPL}$) regression model as an alternative to model overdispersed count data exhibiting inflation or deflation of zeros in the presence of covariates. The zero modification is incorporated by considering that a zero-truncated process produces positive observations, and consequently, the proposed model can be fitted without any previous information about the zero modification present in a given dataset. A fully Bayesian approach based on the *g-prior* method has been considered for inference concerns. An intensive Monte Carlo simulation study has been conducted to evaluate the performance of the developed methodology and the maximum likelihood estimators (MLEs). The proposed model was considered for the analysis of a real dataset on the number of bids received by 126 US firms between 1978-1985, and the impact of choosing different *prior* distributions for the regression coefficients has been studied. A sensitivity analysis to detect influential points has been performed based on the Kullback-Leibler divergence. The randomized quantile residuals were considered for the task of model validation. Besides, a general comparison with some well-known regression models for discrete data has been presented.

This chapter is organized as follows. In Section 4.2, we briefly present the Poisson-Lindley ($\mathcal{PL}$) distribution and some of its main statistical properties. In Section 4.3, we introduce the $\mathcal{ZMPL}$ model, demonstrating its flexibility to deal with zero-inflated/deflated data. In Section 4.4, we present a regression framework based on the hurdle version of the $\mathcal{PL}$ distribution. In Section 4.5, we describe all the Bayesian procedures and methodologies that are considered for inferential purposes in this chapter. In Section 4.6, we discuss the results of a Monte Carlo simulation study and in Section 4.7, an application of the proposed model is exhibited. General comments and concluding remarks are addressed in Section 4.8.

4.2 The PL distribution

A random variable ψ is said to have Lindley $(\mathcal{L})$ distribution (LINDLEY, 1958) if its probability density function has the form

$$f(\psi;\theta) = \frac{\theta^2}{(\theta+1)}(\psi+1)e^{-\theta\psi}, \quad \psi \in \mathbb{R}_+,$$

where $\theta \in \mathbb{R}_+$ is the shape parameter.

The $\mathcal{PL}$ distribution is a probabilistic model that may arise when it is believed that the rate parameter (ψ) of the Poisson $(\mathcal{P})$ model behaves according to a $\mathcal{L}$ random variable into the subset of positive real numbers $\mathbb{R}_+$. In other words, a random variable Y is said to follow the $\mathcal{PL}$ law if the stochastic representation $Y|\psi \sim \mathcal{P}(\psi)$ and $\psi \sim \mathcal{L}(\theta)$ holds for all $\theta \in \mathbb{R}_+$. In this case, the $\mathcal{PL}$ distribution is defined by the equation

$$\mathrm{P}(Y=y;\theta) = \frac{\theta^2(\theta+y+2)}{(\theta+1)^{y+3}}, \quad y \in \mathcal{Y}_0, \tag{4.1}$$

for $\theta \in \mathbb{R}_+$ and $\mathcal{Y}_0 = \{0,1,\ldots\}$ is the set of nonnegative integers. Using the gamma integral, the Equation (4.1) can be easily obtained by integrating $\mathrm{P}(Y=y|\psi)f(\psi;\theta)$ respect to ψ over $\mathbb{R}_+$, in which $\mathrm{P}(Y=y|\psi)$ is the conditional probability mass function (pmf) of a $\mathcal{P}$ random variable.

The unconditional distribution of the random variable Y can be denoted by $\mathcal{PL}(\theta)$. The pmf (4.1) does not involve complicated expressions, and therefore, the probabilities can be easily computed over $\mathcal{Y}_0$ as

$$\mathrm{P}(Y=0;\theta) = \frac{\theta^2(\theta+2)}{(\theta+1)^3}.$$

From the results provided by Sankaran (1970), the mean and the variance of Y are given, respectively, by

$$\mu = \frac{\theta+2}{\theta(\theta+1)} \quad \text{and} \quad \sigma^2 = \frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2}, \tag{4.2}$$

for $\theta \in \mathbb{R}_+$. It can be easily shown that $\mu \propto \theta^{-1}$. Also, the expression of σ^2 can be straightforwardly rearranged as

$$\sigma^2 = \mu\left[1 + \frac{\theta^2+4\theta+2}{\theta(\theta+1)(\theta+2)}\right],$$

where the term inside the brackets correspond to the index of dispersion $(\mathrm{ID} = \sigma^2\mu^{-1})$. One can notice that the ratio involving θ is always positive. This implies that the $\mathcal{PL}$ distribution is always overdispersed, that is, $\{\theta \in \mathbb{R}_+ : \sigma^2 \leqslant \mu\} = \emptyset$. In addition, the index of dispersion is clearly greater than 1, also implying overdispersion. Conversely, we have that $\mathrm{ID} \to 1$ $(\sigma^2 \to \mu)$ as $\theta \to \infty$, that is, the $\mathcal{PL}$ distribution has the property of equidispersion for large values of θ.

Another useful measure to characterize a discrete distribution is the zero modification (ZM) index

$$\mathrm{ZM} = 1 + \mu^{-1}\log\left[\mathrm{P}(Y=0)\right],$$

which is defined based on the $\mathcal{P}$ distribution. This measure can be easily interpreted since $\text{ZM} > 0$ indicates zero inflation, $\text{ZM} < 0$ indicates zero deflation and $\text{ZM} = 0$ indicates no zero modification. For the $\mathcal{PL}$ distribution, the ZM index is given by

$$\text{ZM} = 1 + \frac{\theta(\theta+1)[2\log(\theta) + \log(\theta+2) - 3\log(\theta+1)]}{\theta+2},$$

for $\theta \in \mathbb{R}_+$. When analysing the ZM index more deeply, we have obtained that $\text{ZM} \to -1$ as $\theta \to \infty$ and $\text{ZM} \to 0$ as $\theta \to 0$. This implies that, besides the usual case $(\text{ZM} = 0)$, the $\mathcal{PL}$ distribution is suitable to deal with zero deflation, but is not indicated to model zero-inflated datasets.

Now, let us reparameterize the pmf (4.1) in terms of the mean μ. It can be particularly useful since our interest is to derive a regression model based on the $\mathcal{PL}$ distribution, in which the influence of fixed and random effects can be evaluated directly over the mean of the response variable. Since $\theta \in \mathbb{R}_+$, we have that

$$\theta = -\frac{(\mu-1) - \sqrt{(\mu-1)^2 + 8\mu}}{2\mu}, \tag{4.3}$$

and, if we denote $\theta = h(\mu)$, the pmf (4.1) can be rewritten as

$$\text{P}(Y = y; \mu) = \frac{h^2(\mu)[h(\mu) + y + 2]}{[h(\mu)+1]^{y+3}}, \quad y \in \mathcal{Y}_0, \tag{4.4}$$

for $\mu \in \mathbb{R}_+$.

4.3 The ZMPL distribution

In addition to the interest in the case where the equidispersion assumption on the $\mathcal{P}$ distribution is violated, we may also be interested in the cases where a high/low amount of zeros is observed beyond that generated by the original process, which we already supposed to account for overdispersion. There are some typical situations where zero modification may occur, and we list these cases in the following.

a) Not all members of the population are affected by the process, which causes inflation of zeros to occur due to the response of unaffected subjects being zero;

b) When zeros cannot be observed in the population (truncation at zero);

c) The occurrence of unavoidable problems during the sampling process may lead to an increase/decrease in the probability of a zero observation being selected, hence the zero inflation/deflation situation;

d) A combination of (a) and (b) causes a part of the population to be zero-truncated distributed while the other part is not affected and provides the zero observations.

In this section, the $\mathcal{ZMPL}$ model is introduced as an alternative to model overdispersed count datasets when a high/low amount of zeros is observed beyond what would be expected by the $\mathcal{PL}$ distribution. Thus, let Y be a discrete random variable defined on $\mathcal{Y}_0$. It can be stated that Y is distributed according to a $\mathcal{ZMPL}$ distribution if its pmf can be written as

$$\mathrm{P}_*(Y=y;\mu,p) = (1-p)\,\delta_y + p\,\mathrm{P}(Y=y;\mu)\,, \quad y \in \mathcal{Y}_0, \tag{4.5}$$

for $\mu \in \mathbb{R}_+$ and p is the zero modification parameter. Also, δ_y is an indicator function, so that $\delta_y = 1$ if $y = 0$ and $\delta_y = 0$ otherwise. For the class of zero-modified models, the parameter p is subject to the condition (the so-called p-condition) given by

$$0 \leqslant p \leqslant \mathrm{P}^{-1}(Y>0;\mu)\,, \tag{4.6}$$

where $\mathrm{P}(Y>0;\mu)$ is the probability of Y being positive under the $\mathcal{PL}$ distribution, given that its mean is μ. In this case, we have that

$$\mathrm{P}(Y>0;\mu) = \frac{h^2(\mu)+3h(\mu)+1}{[h(\mu)+1]^3}, \tag{4.7}$$

for $\mu \in \mathbb{R}_+$.

One can easily notice that model (4.5) is not a mixture distribution typically chosen to model zero-inflated data, since parameter p can assume values greater than 1. However, for all values of p between 0 and the boundary $\mathrm{P}^{-1}(Y>0;\mu)$, the Equation (4.5) corresponds to a proper pmf since $\mathrm{P}_*(Y=y;\mu,p) \geqslant 0$ for all $y \in \mathcal{Y}_0$ and the probabilities sum to 1 over $\mathcal{Y}_0$.

The mean and the variance of $Y \sim \mathcal{ZMPL}(\mu,p)$ are given, respectively, by

$$\mu_* = p\mu \quad \text{and} \quad \sigma^2_* = p\left[\sigma^2 + (1-p)\mu^2\right],$$

where μ and σ^2 are given in Equation (4.2). For the zero-modified case, the index of dispersion can be expressed as $\mathrm{ID}_* = \sigma^2\mu^{-1} + (1-p)\mu$. The term $(1-p)\mu$ represents the overdispersion caused by a modification on the zero frequency, regarding the $\mathcal{PL}$ distribution.

The $\mathcal{ZMPL}$ distribution may be considered an interesting alternative to the zero-modified Poisson ($\mathcal{ZMP}$) model, since the base distribution of the former can accommodate several levels of overdisperson, issue that the $\mathcal{P}$ distribution generally fails to deal with. Figure 12 depicts the behaviour of the $\mathcal{ZMPL}$ distribution for different values of p and for $\mu = 0.25$ (implying $p \in [0, 4.98]$), for $\mu = 0.50$ (implying $p \in [0, 2.96]$), for $\mu = 0.75$ (implying $p \in [0, 2.28]$) and for $\mu = 1.00$ (implying $p \in [0, 1.94]$). When looking at the pmf plots and the conditions regarding the zero modification parameter, one can notice that the behaviour of pmf (4.5) is highly affected by the value of p, as can also be seen by considering the proportion of additional/missing zeros

$$\mathrm{P}_*(Y=0;\mu,p) - \mathrm{P}(Y=0;\mu) \;=\; (1-p) + p\,\mathrm{P}(0;\mu) - \mathrm{P}(Y=0;\mu)$$

$$= \quad (1-p)\,\mathrm{P}(Y>0;\mu)\,. \tag{4.8}$$

The first interpretation one can take from Equation (4.8) is that parameter p plays the primary role in controlling the frequency of zeros, and therefore, it has a natural interpretation regarding the proportions of either inflation or deflation at zero. The following statements describe the effect of parameter p on Equation (4.5).

i) If $p=0$ then $\mathrm{P}_*(Y=0;\mu,p)=1$. This implies that Equation (4.5) is the pmf of a degenerate distribution with all mass at zero;

ii) If $p=1$ then $\mathrm{P}_*(Y=0;\mu,p)=\mathrm{P}(Y=0;\mu)$. This implies that Equation (4.5) is the pmf (4.7);

iii) If $p\in(0,1)$ then $(1-p)\mathrm{P}(Y>0;\mu)>0$. This implies that Equation (4.5) has a proportion of zeros greater than pmf (4.4), hence zero inflation;

iv) If $p\in[1,\mathrm{P}^{-1}(Y>0;\mu)]$ then $(1-p)\mathrm{P}(Y>0;\mu)<0$. This implies that Equation (4.5) has a proportion of zeros smaller than pmf (4.4), hence zero deflation;

v) If $p=\mathrm{P}^{-1}(Y>0;\mu)$ then $\mathrm{P}_*(Y=0;\mu,p)=0$. This implies that Equation (4.5) is the zero-truncated Poisson-Lindley ($\mathcal{ZTPL}$) distribution (GHITANY; AL-MUTAIRI; NADARAJAH, 2008), with pmf given by

$$\mathrm{P}^*(Y=y;\mu)=\frac{\mathrm{P}(Y=y;\mu)}{\mathrm{P}(Y>0;\mu)}\,(1-\delta_y)\,, \tag{4.9}$$

where the numerator is given by (4.4) and the denominator is given by (4.7). One can notice that the reparameterization of Equation (4.1) in terms of μ does not affect the general definition of the zero-truncated version of the $\mathcal{PL}$ model. See Shanker and Fesshaye (2016) for further details about the $\mathcal{ZTPL}$ distribution.

Given the value of p, one can easily identify the nature of the zero-valued observations under the phenomenon of interest. In this way, we have that the case (iii) may be appropriate in situations (a), (c) and (d) as described at the beginning of this section and the case (iv) may be appropriate in situations (c) and (d). Moreover, the index of dispersion can be investigated in terms of the modification at zero since $\mathrm{ID}_*=\mathrm{ID}$ in the standard case (ii), $\mathrm{ID}_*>\mathrm{ID}$ in the zero-inflated case (iii) and $\mathrm{ID}_*<\mathrm{ID}$ in the zero-deflated case (iv).

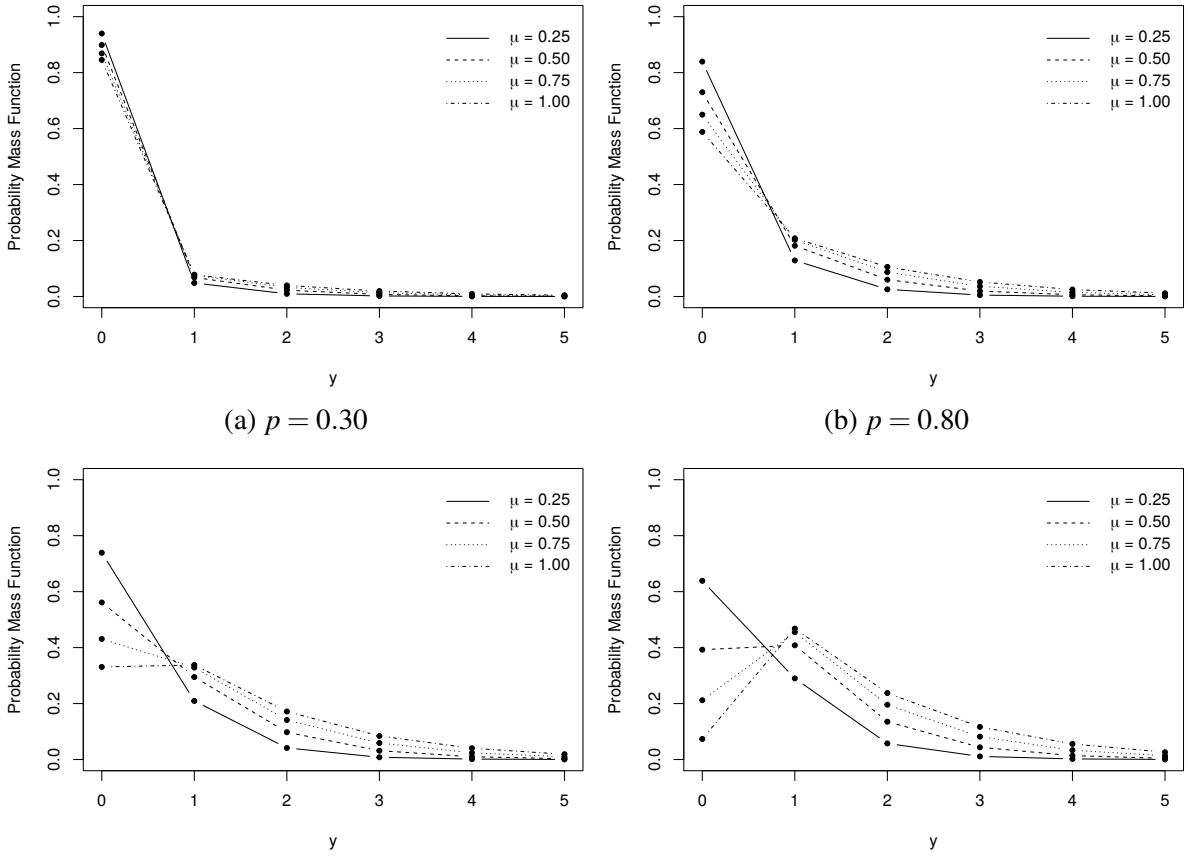
Now, let us rewritten pmf (4.5) as

$$\begin{aligned}\mathrm{P}_*(Y=y;\mu,p) &= [1-p+p\,\mathrm{P}(Y=0;\mu)]\,\delta_y+p\,\mathrm{P}(Y=y;\mu)\,(1-\delta_y)\\ &= [1-p\,\mathrm{P}(Y>0;\mu)]\,\delta_y+p\,\mathrm{P}(Y=y;\mu)\,(1-\delta_y)\,,\qquad y\in\mathcal{Y}_0,\end{aligned}$$

for $\mu\in\mathbb{R}_+$. Taking $\omega=p\,\mathrm{P}(Y>0;\mu)$, we have that

$$\mathrm{P}_*(Y=y;\mu,\omega)=(1-\omega)\,\delta_y+\omega\,\mathrm{P}^*(Y=y;\mu)\,,\qquad y\in\mathcal{Y}_0, \tag{4.10}$$

Figure 12 – Behavior of the $\mathcal{ZMPL}$ distribution for different values of μ and p.



(a) $p = 0.30$ (b) $p = 0.80$

(c) $p = 1.30$ (d) $p = 1.80$

Source: Elaborated by the author.

where $\mathrm{P}^*(Y = y;\mu)$ is given by Equation (4.9). By condition (4.6), we clearly have that $\omega \in [0,1]$.

The corresponding cumulative distribution function (cdf) of Y is given by

$$\mathrm{F}_*(y;\mu,\omega) = 1 - \omega\left[\frac{h^2(\mu) + (y+3)\,h(\mu) + 1}{[h^2(\mu) + 3h(\mu) + 1]\,[h(\mu)+1]^y}\right], \quad y \in \mathcal{Y}_0. \tag{4.11}$$

Equation (4.10) corresponds to the hurdle version of the $\mathcal{PL}$ distribution. Mullahy (1986) introduced the class of hurdle models, and the relevant feature of such class is that the zero-valued observations are treated separately from the positive ones. In the main formulation, a binary probability model determines whether a zero or a nonzero outcome occurs, and hence, an appropriated zero-truncated discrete distribution is chosen to describe the positive values (SAFFARI; ADNAN; GREENE, 2012). In this case, we have that the probability of $Y = 0$ is $1-\omega$, and the probability of $Y > 0$ is ω. The $\mathcal{ZMPL}$ distribution parameterized by ω can be denoted by $\mathcal{ZMPL}(\mu,\omega)$. Such representation can be visualized as a superposition of two random processes, that is, one that produces positive observations from the $\mathcal{ZTPL}$ distribution and another one that produces only zero-valued observations with probability $1-\omega$. Therefore, model (4.10) cannot be considered a 2-component mixture distribution.

By the hurdle representation of zero-modified models, only the positive observations are required to estimate the parameter μ. Conceição *et al.* (2017a) well discusses this fact for the class of zero-modified Power Series distributions, and here we extend this result by asserting that the zero-truncated version of the $\mathcal{ZMPL}$ distribution is equivalent to the $\mathcal{ZTPL}$ distribution and they have the same parameter μ. This can be easily checked using Equation (4.10) because if we exclude the value zero from $\mathcal{Y}_0$ and divide the right-hand side of (4.10) by the probability of Y being positive (ω), then we will get that $Y \sim \mathcal{ZTPL}(\mu)$. Besides, such a representation allows us to obtain a closed-form solution for the MLE of parameter ω, which is given by the proportion of nonzeros in the dataset. Also, it can be easily seen that, for any fixed $\mu \in \mathbb{R}_+$, the function $\omega \mathrm{P}^{-1}(Y > 0; \mu)$ maps from $[0,1]$ to $[0, \mathrm{P}^{-1}(Y > 0; \mu)]$ bijectively and therefore, the invariance principle ensures that parameter p can be estimated using such function. Indeed, inference procedures about parameter p are required since we are often interested in identifying the kind of zero modification (inflation or deflation) is present in a given dataset.

4.4 The ZMPL regression model

Let us suppose that we have a collection $(Y_1, \ldots, Y_n)$ of independent discrete random variables such that $Y_i | \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta} \sim \mathcal{ZMPL}(\mu_i, \omega_i)$ $(i = 1, \ldots, n)$. In this case, a regression model for count data based on the $\mathcal{ZMPL}$ distribution can be derived by rewriting Equation (4.10) as

$$\mathrm{P}_*(Y_i = y_i; \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}) = (1 - \omega_i)\, \delta_{y_i} + \omega_i \mathrm{P}^*(Y_i = y_i; \mu_i), \quad y_i \in \mathcal{Y}_0, \tag{4.12}$$

where $\boldsymbol{x}_i^\top = (1, x_{i1}, \ldots, x_{iq_1})$ and $\boldsymbol{z}_i^\top = (1, z_{i1}, \ldots, z_{iq_2})$ are related, respectively, to μ_i and ω_i and can include, for example, *dummy* variables, cross-level interactions and polynomials. The $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is the full vector of fixed-effects, being $\boldsymbol{\beta}_1^\top = (\beta_{10}, \ldots, \beta_{1q_1})$ and $\boldsymbol{\beta}_2^\top = (\beta_{20}, \ldots, \beta_{2q_2})$. Here, q_1 (q_2) denotes the number of covariates considered on the systematic component of a linear predictor for parameter μ (ω).

The full design matrices of model (4.12) can be written as $\boldsymbol{X} = (\boldsymbol{1}_n, \boldsymbol{X}_{n \times q_1})$ and $\boldsymbol{Z} = (\boldsymbol{1}_n, \boldsymbol{Z}_{n \times q_2})$, where $\boldsymbol{1}_n$ is the intercept column and the submatrices $\boldsymbol{X}_{n \times q_1}$ and $\boldsymbol{Z}_{n \times q_2}$ are defined in such a way that the vector $(x_{i1}, \ldots, x_{iq_1})$ is the i-th row of $\boldsymbol{X}_{n \times q_1}$ and the vector $(z_{i1}, \ldots, z_{iq_2})$ is the i-th row of $\boldsymbol{Z}_{n \times q_2}$. To complete model definition, one have to specify two monotonic, invertible and twice differentiable link functions, say g_1 and g_2, in which $\mu_i = g_1^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}_1)$ and $\omega_i = g_2^{-1}(\boldsymbol{z}_i^\top \boldsymbol{\beta}_2)$ are well defined on $\mathbb{R}_+$ and $(0,1)$, respectively. For this purpose, one can choose any suitable mappings g_1 and g_2 such that $g_1^{-1}\colon \mathbb{R} \to \mathbb{R}_+$ and $g_2^{-1}\colon \mathbb{R} \to (0,1)$. The logarithm link function, $\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_1$, is the natural choice for g_1. For g_2, the most usual choice is the *logit* link function,

$$\operatorname{logit}(\omega_i) = \log\left(\frac{\omega_i}{1 - \omega_i}\right) = \boldsymbol{z}_i^\top \boldsymbol{\beta}_2. \tag{4.13}$$

The *probit* link function,

$$\Phi^{-1}(\omega_i) = \boldsymbol{z}_i^\top \boldsymbol{\beta}_2, \tag{4.14}$$

is also appropriate for the requested purpose. Another possible choice for g_2 is

$$\log\left[-\log\left(1-\omega_i\right)\right] = \boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2, \tag{4.15}$$

which corresponds to the *complementary log-log* link function. Unlike the *logit* and *probit*, the *complementary log-log* transformation provides an asymmetric specification that can be useful when the probability of an outcome is very high/low. Further, a more sophisticated approach considering the power and the reversal power link functions was proposed by Bazán *et al.* (2017) and can be applied in our context to provide more flexible relationships between the linear predictor and parameter ω.

The $\mathcal{ZMPL}$ regression model has q_1+q_2+2 unknown parameters to be estimated, the components of the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The link functions (4.13)-(4.15) for parameter ω exclude the limit cases (i) and (v). Besides, it is worthwhile to mention that identifiability problems could occur if the same covariate were used to model both mean (μ) and the zero modification parameter (p) if we had considered a regression model derived from (4.10). Also, to ensure that the regression coefficients are identifiable, it is essential the covariates (within linear predictors) to be linearly independent. Unlike traditional approaches, the proposed model can be fitted to both zero-inflated/deflated datasets. In this case, given a set of covariates, the probability of a zero-valued count being observed for the i-th individual is given by $1-\omega_i$. The adopted parameterization makes model (4.12) separable into two parts due to orthogonality between parameters in the structural form of μ and ω. It also avoids nonidentifiability problems as well as the use of the EM algorithm, typically used to fit mixture models. Regardless of the model framework, in this chapter, we propose a fully Bayesian approach for estimation and inference procedures. The next section is dedicated to present the details of such an approach.

4.5 Inference

Let Y be a discrete random variable taking values on $\mathcal{Y}_0$. Suppose that a random experiment is carried out n times independently and, subject to $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ for each i, a vector $\boldsymbol{y}=(y_1,\ldots,y_n)$ of observed values from Y is obtained. Considering model formulation (4.12), if $Y_i|\boldsymbol{x}_i,\boldsymbol{z}_i,\boldsymbol{\beta}\sim\mathcal{ZMPL}(\mu_i,\omega_i)$ holds for all i, then the likelihood function of the vector $\boldsymbol{\beta}$ can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta};\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z}) &= \prod_{i=1}^{n}\omega_i\left(\frac{1-\omega_i}{\omega_i}\right)^{\delta_{y_i}}\left[\frac{\mathrm{P}(Y_i=y_i;\mu_i)}{\mathrm{P}(Y_i>0;\mu_i)}\right]^{1-\delta_{y_i}}\\ &= \prod_{i=1}^{n}\left\{g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)\left[\frac{1-g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)}{g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)}\right]^{\delta_{y_i}}\left[\frac{\mathrm{P}\left[Y_i=y_i;g_1^{-1}\left(\boldsymbol{x}_i^{\top}\boldsymbol{\beta}_1\right)\right]}{\mathrm{P}\left[Y_i>0;g_1^{-1}\left(\boldsymbol{x}_i^{\top}\boldsymbol{\beta}_1\right)\right]}\right]^{1-\delta_{y_i}}\right\},\end{aligned}$$

and the correspondent log-likelihood function is given by

$$\ell(\boldsymbol{\beta};\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z}) = \sum_{i=1}^{n}(1-\delta_{y_i})\log\left\{\frac{\mathrm{P}\left[Y_i=y_i;g_1^{-1}\left(\boldsymbol{x}_i^{\top}\boldsymbol{\beta}_1\right)\right]}{\mathrm{P}\left[Y_i>0;g_1^{-1}\left(\boldsymbol{x}_i^{\top}\boldsymbol{\beta}_1\right)\right]}\right\}+$$

$$
\begin{aligned}
& \sum_{i=1}^{n} \left\{ \log\left[g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)\right] - \delta_{y_i} \log\left[\frac{g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)}{1-g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)}\right] \right\} \\
= \; & \ell_1\left(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{X}\right) + \ell_2\left(\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{Z}\right). \qquad (4.16)
\end{aligned}
$$

For the $\mathcal{ZMPL}$ regression model, we will consider the log-linearity of the mean, that is, $g_1(\mu_i) = \log(\mu_i) = \boldsymbol{x}_i^{\top}\boldsymbol{\beta}_1$. The choice of g_2 is left open and the notation $\omega_i = g_2^{-1}(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2)$ will be used when necessary. From Equation (4.16), one can easily notice that the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal and that $\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{X})$ depends only on the positive values on $\boldsymbol{y}$. In this way, the log-likelihood function of $\boldsymbol{\beta}_1$ takes the form

$$
\begin{aligned}
\ell_1\left(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{X}\right) = & \sum_{k\in\mathcal{K}_1} \log\left[h\left(\exp\left\{\boldsymbol{x}_k^{\top}\boldsymbol{\beta}_1\right\}\right) + y_k + 2\right] + 2\sum_{k\in\mathcal{K}_1} \log\left[h\left(\exp\left\{\boldsymbol{x}_k^{\top}\boldsymbol{\beta}_1\right\}\right)\right] - \\
& \sum_{k\in\mathcal{K}_1} \log\left[h^2\left(\exp\left\{\boldsymbol{x}_k^{\top}\boldsymbol{\beta}_1\right\}\right) + 3h\left(\exp\left\{\boldsymbol{x}_k^{\top}\boldsymbol{\beta}_1\right\}\right) + 1\right] - \\
& \sum_{k\in\mathcal{K}_1} y_k \log\left[h\left(\exp\left\{\boldsymbol{x}_k^{\top}\boldsymbol{\beta}_1\right\}\right) + 1\right], \qquad (4.17)
\end{aligned}
$$

where $\mathcal{K}_1 = \{i : y_i > 0, y_i \in \boldsymbol{y}\}$ is the finite set of indexes regarding the positive observations of $\boldsymbol{y}$. Adopting this setup is equivalent to assuming that each positive element of $\boldsymbol{y}$ comes from a $\mathcal{ZTPL}$ distribution. Here, we are extending the fact that estimating the $\mathcal{P}$ parameter θ using the zero-truncated Poisson distribution results in a loss of efficiency in the inference if there is no zero modification (DIETZ; BÖHNING, 2000; CONCEIÇÃO; ANDRADE; LOUZADA, 2014). Now, the log-likelihood function of $\boldsymbol{\beta}_2$ can be written as

$$
\ell_2\left(\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{Z}\right) = \sum_{i=1}^{n} \log\left[g_2^{-1}\left(\boldsymbol{z}_i^{\top}\boldsymbol{\beta}_2\right)\right] - \sum_{k\in\mathcal{K}_2} \log\left[\frac{g_2^{-1}\left(\boldsymbol{z}_k^{\top}\boldsymbol{\beta}_2\right)}{1-g_2^{-1}\left(\boldsymbol{z}_k^{\top}\boldsymbol{\beta}_2\right)}\right], \qquad (4.18)
$$

where $\mathcal{K}_2 = \{i : y_i = 0, y_i \in \boldsymbol{y}\}$ is the finite set of indexes regarding the zero-valued observations of $\boldsymbol{y}$.

There are no closed-form solutions for the MLEs of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and therefore, nonlinear optimization algorithms or direct numerical search on the surface of log-likelihoods functions may be applied in order to obtain point estimates in the classical approach. By the maximum likelihood theory, a consistent estimator for the covariance matrix of $\hat{\boldsymbol{\beta}}_r$ $(r = 1,2)$, is given by the inverse of the Fisher information $\mathcal{I}_r = \mathbb{E}_Y[\mathcal{J}_r]$, where

$$
\mathcal{J}_1 = -\frac{\partial^2 \ell_1\left(\boldsymbol{\beta}_1;\boldsymbol{y},\boldsymbol{X}\right)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1^{\top}} \quad \text{and} \quad \mathcal{J}_2 = -\frac{\partial^2 \ell_2\left(\boldsymbol{\beta}_2;\boldsymbol{y},\boldsymbol{Z}\right)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2^{\top}},
$$

are the observed information matrices. In our context, however, the computation of the expected value respect to Y is unfeasible and therefore, a numerical approximation for the covariance matrices can be obtained by evaluating $\mathcal{J}_r^{-1}$ at $\boldsymbol{\beta}_r = \hat{\boldsymbol{\beta}}_r$ and using the observed vector $\boldsymbol{y}$.

4.5.1 Prior distributions

A Bayesian analysis starts by choosing suitable *prior* distributions for the set of unknown parameters. The *g-prior* (ZELLNER, 1986) is a common choice among Bayesian users of the

multiple linear regression model, mainly due to the fact of providing a closed-form *posterior* distribution for the regression coefficients. The *g-prior* is classified as an objective *prior* method which uses the inverse of the Fisher information matrix up to a scalar variance factor ($\tau \in \mathbb{R}_+$) to obtain the *prior* covariance structure of the multivariate Normal distribution. Such specification is indeed quite attractive since the Fisher information plays a major role in the determination of large sample covariance in both Bayesian and classical inference.

The problem of eliciting conjugate *priors* for generalized linear models (GLMs) was addressed by Chen and Ibrahim (2003). Their approach can be considered as a generalization of the original *g-prior* method, but its application is restricted for the class of GLMs since the proposed *prior* does not have closed-form for non-normal exponential families. As an alternative, Gupta and Ibrahim (2009) have proposed the information matrix *prior* as a way to assess the *prior* correlation structure between the regression coefficients, not including the intercept since the design matrix is centered in order to ensure that β_0 is orthogonal to the other coefficients. This method uses the Fisher information similarly to a precision matrix whose elements are shrunken by the factor τ, which is considered fixed ($\tau \geqslant 1$). Based on such approach, we will consider, for the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, two multivariate Normal *prior* distributions of the form

$$\boldsymbol{\beta}_1 \sim \mathcal{N}_{\overline{q}_1}\left(\boldsymbol{\beta}_1^0, \tau_1 \boldsymbol{\Sigma}_1^0\right) \quad \text{and} \quad \boldsymbol{\beta}_2 \sim \mathcal{N}_{\overline{q}_2}\left(\boldsymbol{\beta}_2^0, \tau_2 \boldsymbol{\Sigma}_2^0\right),$$

where $\overline{q}_r = q_r + 1$ and $\boldsymbol{\Sigma}_r^0$ refers to $\mathcal{J}_r^{-1}$ evaluated numerically at $\boldsymbol{\beta}_r^0$, and τ_r is assumed known. The vectors $\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^0$ can be chosen arbitrarily if no specialized information is available. It is worthwhile to mention that we are not considering centered design matrices in our approach. Hence, we are able to include β_{10} in the proposed *g-prior* but, in this case, the intercept is *a priori* correlated with the other coefficients $(\beta_{11}, \ldots, \beta_{1q_1})$. The same applies for β_{20} and $(\beta_{21}, \ldots, \beta_{2q_2})$.

4.5.2 Posterior distributions and estimation

After *prior* specifications, the following step in a Bayesian analysis consists in the obtaining of computable *posterior* densities for the unknown model parameters. For the $\mathcal{ZMPL}$ regression model (4.12), the unnormalized joint *posterior* distribution of the vector $\boldsymbol{\beta}$ can be expressed as

$$\pi(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{X}) + \ell_2(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{Z})\} \pi_1(\boldsymbol{\beta}_1) \pi_2(\boldsymbol{\beta}_2).$$

However, since $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal, we have that

$$\pi_1(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{X}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{X})\} \pi_1(\boldsymbol{\beta}_1) \tag{4.19}$$

and

$$\pi_2(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{Z}) \propto \exp\{\ell_2(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{Z})\} \pi_2(\boldsymbol{\beta}_2), \tag{4.20}$$

where ℓ_1 and ℓ_2 are given by (4.17) and (4.18), respectively. It can be proved that, if the data is discrete, then the use of a proper *prior* distribution (multivariate Normal in our case) avoids the *posterior* density to be improper.

From the Bayesian point of view, the parametric inference is based on the marginal *posterior* distributions, which can be obtained by integrating the joint *posterior* distributions given in (4.19). These densities have unknown forms, mainly due to the complexity of the respective likelihood functions. In this case, Bayesian estimates for each element of $\boldsymbol{\beta}_r$ can be obtained by applying iterative procedures within a broad class of MCMC (Markov Chain Monte Carlo) methods. Here we will consider the well-known Random-walk Metropolis (RwM) algorithm. Through this procedure, $q_r + 1$ chains can be generated for $\boldsymbol{\beta}_r$. The dimensionality issue will depend on how much covariates will be taken under consideration to describe the parameters of the $\mathcal{ZMPL}$ model. For the *posterior* distributions in (4.19), we will consider multivariate Normal specifications for the proposal (candidate-generating) densities in the algorithm. These distributions will be used as the main terms in the transition *kernels* when computing the acceptance probabilities. Hence, at any state $k > 0$, the MCMC simulation is performed by proposing a candidate $\boldsymbol{\psi}_r$ for $\boldsymbol{\beta}_r$ as

$$\boldsymbol{\psi}_r | \boldsymbol{\beta}_r^* \sim \mathcal{N}_{\overline{q}_r}\left[\boldsymbol{\beta}_r^*, \nu_r \boldsymbol{\mathcal{S}}_r^*\right],$$

where $\boldsymbol{\beta}_r^* = \nu_r \boldsymbol{\beta}_r^{(k-1)} + (1-\nu_r)\,\boldsymbol{\beta}_r^0$ and $\nu_r = \tau_r\,(\tau_r+1)^{-1}$. One can notice that transitions depends on the acceptance of pseudo-random vectors generated with mean given by the mixture between the actual state of the chains and the *priors* specification, which are shrunked by the factor $1-\nu_r$. In addition, at any state $k > 0$, the covariance matrix of the candidate vector $\boldsymbol{\psi}_r$ can be approximated numerically by evaluating $\boldsymbol{\mathcal{S}}_r^* = \boldsymbol{\mathcal{H}}_r^{-1}$ at $\boldsymbol{\beta}_r = \boldsymbol{\beta}_r^*$, where

$$\boldsymbol{\mathcal{H}}_1 = -\frac{\partial^2 \log\left[\pi_1\left(\boldsymbol{\beta}_1; \boldsymbol{y}, \boldsymbol{X}\right)\right]}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^\intercal} \quad \text{and} \quad \boldsymbol{\mathcal{H}}_2 = -\frac{\partial^2 \log\left[\pi_2\left(\boldsymbol{\beta}_2; \boldsymbol{y}, \boldsymbol{Z}\right)\right]}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^\intercal}.$$

Algorithm 5 can be used to generate chains for the regression coefficients using the RwM algorithm. To run the algorithm, initial conditions $\boldsymbol{\beta}_1^{(0)}$ and $\boldsymbol{\beta}_2^{(0)}$ are needed. For a specific asymptotic Gaussian environment, Roberts, Gelman and Gilks (1997) have shown that the optimal acceptance rate should be around 45% for 1-dimensional problems and asymptotically approaches to 23.40% in higher-dimensional problems (> 4). Here we are considering acceptance rates varying between 23.40% and 40% as quite reasonable since the proposed model will generally have at least two parameters to be estimated. Indeed, the higher the values of τ_1 and τ_2, the smaller the acceptance rates in the RwM algorithm, which results in smaller variability of the estimates. This procedure generates a sample of size N for each parameter. The convergence of the chains can be monitored by the Gelman-Rubin (GELMAN; RUBIN, 1992) and Geweke (GEWEKE, 1992) diagnostics. After convergence, some of the generated samples can be discarded as burn-in. The procedure to decrease the correlation between and within generated chains is the usual approach of getting thinned steps. The final sample is supposed to

have a size of M. A summary of the *posterior* distributions can be obtained through the MCMC estimates.

In the next section, we discuss the results of a Monte Carlo simulation study that was conducted to assess the performance of the proposed Bayesian methodology. In Section 4.7, the usefulness, and the competitivity of the proposed regression model are illustrated by using a real dataset. All computations were performed using the R environment (R Development Core Team, 2017).

Algorithm 5 – Random-walk Metropolis

procedure RWM($N, \boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}, \boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \tau_1, \tau_2$)
  Set $k \leftarrow 1$, $\nu_1 \leftarrow \tau_1 (1+\tau_1)^{-1}$, and $\nu_2 \leftarrow \tau_2 (1+\tau_2)^{-1}$
  while $k \leqslant N$ **do**
    Set $\boldsymbol{\beta}_1^{(k)} \leftarrow \boldsymbol{\beta}_1^{(k-1)}$ and $\boldsymbol{\beta}_2^{(k)} \leftarrow \boldsymbol{\beta}_2^{(k-1)}$
    Set $\boldsymbol{\beta}_1^* \leftarrow \nu_1 \boldsymbol{\beta}_1^{(k)} + (1-\nu_1)\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^* \leftarrow \nu_2 \boldsymbol{\beta}_2^{(k)} + (1-\nu_2)\boldsymbol{\beta}_2^0$
    Generate $\boldsymbol{\psi}_1 \sim \mathcal{N}(\boldsymbol{\beta}_1^*, \nu_1 \boldsymbol{\mathcal{S}}_1^*)$ and $\boldsymbol{\psi}_2 \sim \mathcal{N}(\boldsymbol{\beta}_2^*, \nu_2 \boldsymbol{\mathcal{S}}_2^*)$
    Set $\alpha_1 \leftarrow \log[\pi_1(\boldsymbol{\psi}_1; \boldsymbol{y}, \boldsymbol{X})] - \log[\pi_1(\boldsymbol{\beta}_1^{(k)}; \boldsymbol{y}, \boldsymbol{X})]$
    Set $\alpha_2 \leftarrow \log[\pi_2(\boldsymbol{\psi}_2; \boldsymbol{y}, \boldsymbol{Z})] - \log[\pi_2(\boldsymbol{\beta}_2^{(k)}; \boldsymbol{y}, \boldsymbol{Z})]$
    Generate $u_1, u_2 \sim \mathcal{U}(0,1)$
    if $\log(u_1) \leqslant \alpha_1$ and $\log(u_2) \leqslant \alpha_2$ **then**
      Set $\boldsymbol{\beta}_1^{(k)} \leftarrow \boldsymbol{\psi}_1$ and $\boldsymbol{\beta}_2^{(k)} \leftarrow \boldsymbol{\psi}_2$
    end if
    Set $k \leftarrow k+1$
  end while
  return $\{\boldsymbol{\beta}^{(k)}\}_{k=1}^N$
end procedure

4.6 Simulation study

The primary empirical properties of an estimation procedure can be evaluated through Monte Carlo simulations. We have performed an intensive simulation study aiming to validate the proposed Bayesian approach. For comparison purposes, the performance of the MLEs was also assessed. The simulation process was performed by generating 500 pseudo-random samples of sizes $n = 50, 100, 200$, and 500 of a variable Y following a $\mathcal{ZMPL}$ distribution under the regression framework (4.12). For the whole process, it was considered a $n \times 2$ design matrix $\boldsymbol{X} = (\boldsymbol{1}_n, \boldsymbol{X}_{n\times 1})$ in which $\boldsymbol{X}_{n\times 1}$ is a vector containing n generated values from an Uniform distribution on the unit interval. Here, we have fixed $\boldsymbol{Z} = \boldsymbol{X}$. Moreover, we have assigned different values for the vectors $\boldsymbol{\beta}_1^\intercal = (\beta_{10}, \beta_{11})$ and $\boldsymbol{\beta}_2^\intercal = (\beta_{20}, \beta_{21})$ in order to generate both zero-inflated and zero-deflated samples. We have considered two scenarios for each kind of zero modification and these cases are treated separately in the following subsections. The logarithm link function was considered for g_1. For g_2, we have considered the link function (4.13).

Algorithm 6 can be used to generate a single pseudo-random realization from the $\mathcal{ZMPL}$ distribution in the regression framework with covariate $\mathcal{U}(0,1)$ for μ and ω. The extension for the use of more covariates is straightforward. The process to generate a pseudo-random sample of size n consists of running the algorithm as often as necessary, say n^* times $(n^* \geqslant n)$. The sequential-search is a black-box type of algorithm (see Hörmann, Leydold and Derflinger (2013)) and works with any computable probability vector. The main advantage of such a procedure is its simplicity. On the other hand, sequential-search algorithms may be slow as the while-loop may have to be repeated very often.

Algorithm 6 – Sequential-Search

1: **procedure** SEQSEA($\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$)
2: Generate $x, u \sim \mathcal{U}(0,1)$
3: Set $\mu \leftarrow \exp\{\beta_{10} + \beta_{11} x\}$ and $\omega \leftarrow [1 + \exp\{-(\beta_{20} + \beta_{21} x)\}]^{-1}$
4: Set $k \leftarrow (1-\omega)$ and $y \leftarrow 0$
5: **while** $u > k$ **do**
6: Set $y \leftarrow y+1$ and $k \leftarrow k + \omega\, \mathrm{P}^*(Y = y; \mu)$
7: **end while**
8: **return** y
9: **end procedure**

For the $\mathcal{ZMPL}$ distribution, the expected number of iterations (NI), that is, the expected number of comparisons in the while condition, is given by

$$\mathbb{E}(\mathrm{NI}) = \mu_* + 1 = \frac{\omega\,[h(\mu)+1]^2\,[h(\mu)+2]}{h(\mu)\,[h^2(\mu)+3h(\mu)+1]} + 1,$$

where $h(\mu)$ is given by Equation (4.3).

To apply the proposed Bayesian approach in each case, we have considered the RwM algorithm for MCMC sampling. A total of $N = 50,000$ values were generated for each parameter, considering a burn-in of 20% of the size of the chain. Using trace plots and Geweke's diagnostic, the convergence of the chains was monitored, and their stationarity was revealed. To obtain pseudo-independent samples from the *posterior* distributions given in (4.19), one out every 10 generated values were kept, resulting in chains of size $M = 4,000$ for each parameter. The *priors* were chosen to ensure that parameter p provides zero inflation or zero deflation depending on the case. We classify these *priors* as being "vague" since the only information we have taken into account is the kind of zero modification present on the generated sample. We fixed $\tau_1 = \tau_2 = 5.0$, which has provided acceptance rates ranging between 30 and 35%. The *posterior* mean was considered as the Bayesian point estimator, and its performance was assessed by evaluating its bias (B), its mean squared error (MSE), and its mean absolute percentage error (MAPE). Also, the coverage probability (CP) of the 95% Bayesian credible intervals (BCIs) was estimated for each parameter. Using the generated samples and letting $\gamma = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the measures

of interest were obtained by

$$\mathrm{B}(\hat{\gamma}) = \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\gamma}_j - \gamma\right), \quad \mathrm{MSE}(\hat{\gamma}) = \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\gamma}_j - \gamma\right)^2 \tag{4.21}$$

and

$$\mathrm{MAPE}(\hat{\gamma}) = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{\hat{\gamma}_j - \gamma}{\gamma}\right|. \tag{4.22}$$

The variance of $\hat{\gamma}$ was estimated as the difference between the MSE and the square of the bias. Moreover, the CP of the BCIs was estimated as follow

$$\mathrm{CP}_{\%}(\gamma) = \frac{100}{M}\sum_{j=1}^{M}\delta_j(\gamma), \tag{4.23}$$

where $\delta_j(\gamma)$ assumes 1 if the j-th BCI contains the real value γ and 0 otherwise. Also, we have estimated the below noncoverage probability (BNCP) and the above noncoverage probability (ANCP) of the BCIs. These measures are computed analogously to CP. The BNCP and ANCP may be useful to determine asymmetrical features since they provide the probabilities of finding the real value of parameter γ on the tails of the generated *posterior* distribution.

For the classical approach, we have considered the Newton-Raphson optimization method to obtain numerical estimates, since no closed-form solution is available for the MLE of vector $\boldsymbol{\beta}$. The estimates were obtained using several initial values to guarantee convergence to the global maximum. Again, assuming $\gamma = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the bias, the MSE and the MAPE of $\hat{\gamma}_{\mathrm{MLE}}$ were estimated as previously stated in Equations (4.21) and (4.22). Besides, we were also interested in the computation of the coverage and noncoverage probabilities of the asymptotic confidence interval (ACI) of γ. The large sample approximation for the $100(1-\alpha)\%$ two-sided confidence interval of γ is given by

$$\hat{\gamma}_{\mathrm{MLE}} \pm z_{(1-\alpha/2)}\hat{\mathrm{SE}}\left(\hat{\gamma}_{\mathrm{MLE}}\right),$$

where $z_{(1-\alpha/2)}$ is the upper $(\alpha/2)$-th percentile of the standard Normal distribution. The standard error (SE) is estimated as the squared root of the variance of $\hat{\gamma}_{\mathrm{MLE}}$. Finally, the CP of the ACIs is estimated using (4.23) and the noncoverage probabilities were computed analogously.

The computed measures are presented in Tables 22-23 (*posterior* summaries) and Tables 24-25 (coverage and noncoverage probabilities). We have noticed that, as expected, the parameter estimates became more accurate with increasing sample sizes since the estimated biases and mean squared errors have decreased considerably as n increased. In general, the MLEs were found more biased regarding the *posterior* expected value. Although high MAPE values were obtained for some parameters, this does not compromise overall accuracy in estimation. For example, on Table 22 (Scenario 1), for $n = 200$, we have obtained a estimated MAPE value of approximately 55% for β_{11}, the Bayesian approach. Taking into account that the real value of this parameter is 0.50, we have that the estimates for β_{11} were ranging mostly between 0.23 and

0.78, which do not represent such a huge impact on the estimated mean (μ). Moreover, we have observed that the estimated CPs of the BCIs are converging to the nominal level of 95%, and the *posterior* distributions became more symmetric with increasing sample sizes. At this point, one can see that slightly better results were obtained for the ACIs, but the use of asymptotic results plays against the classical approach since they are valid only for large n $(n \to \infty)$. Therefore, by considering the predefined structure, our simulation study has provided several indications about the suitability of the proposed Bayesian approach to estimate the parameters of the $\mathcal{ZMPL}$ regression model.

Regarding the comparison between the estimation procedures, it is well-known that the Bayesian approach has advantages when specialized *prior* information is available for the phenomenon under investigation. One of the main concerns in this chapter is to provide the necessary tools for users interested in the application of the proposed model and who have this kind of information accessible. Nonetheless, our methodology can be applied using "vague" and noninformative *priors*. When "vague" *priors* are considered, Bayesian procedures and MLEs present similar results as those obtained in this section. On the other hand, when noninformative *priors* are selected, it can be theoretically proved that both approaches present equivalent results. Therefore, a comparison between Bayesian and classical approaches will only show great Bayesian advantage if the *priors* are specified by aggregating relevant information about the parameters to be estimated. Further, with small samples and good *prior* information, the Bayesian procedure is indeed more advantageous.

4.6.1 Zero-inflated artificial data

For the zero-inflated case, the samples were generated from the $\mathcal{ZMPL}$ distribution by considering that $p_i \in (0,1)$ for all i. Here, the regression coefficients were chosen by taking into account that zero-inflated samples has, naturally, proportion of zeros greater than expected and therefore, the variable Y was generated with mean (μ) not even close to zero. Then, for the first scenario we have considered $\boldsymbol{\beta}_1^\intercal = (1.5, 0.5)$ and $\boldsymbol{\beta}_2^\intercal = (1.0, -1.5)$ to perform the simulation. In the following, the procedure was repeated by considering $\boldsymbol{\beta}_1^\intercal = (2.5, 1.5)$ and $\boldsymbol{\beta}_2^\intercal = (-0.5, -1.0)$. For these scenarios, the "vague" *prior* setup is $(\boldsymbol{\beta}_1^0)^\intercal = (1.0, 0.0)$ and $(\boldsymbol{\beta}_2^0)^\intercal = (0.0, 1.0)$.

Figure 13 depicts the Bayesian estimates obtained for parameter p using zero-inflated samples with $n = 500$ for each scenario. The real values are represented by the straight blue lines and the 95% BCIs are represented by the red dashed lines. The filled black dots represent the estimated values for each generated observation.

4.6.2 Zero-deflated artificial data

For the zero-deflated case, the samples were generated from the $\mathcal{ZMPL}$ distribution by considering that $p_i \in [1, \mathrm{P}^{-1}(Y > 0; \mu_i)]$ for all i. Here, the regression coefficients were chosen

by taking into account that zero-deflated samples has, naturally, proportion of zeros smaller than

Table 22 – Summary of Bayesian and classical estimation using zero-inflated samples.

n	Parameter	Bias		Variance		MSE		MAPE (%)	
		Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
Scenario 1									
50	β_{10}	−0.0496	−0.0514	0.1201	0.1813	0.1225	0.1839	18.4958	19.0244
	β_{11}	0.0327	0.0148	0.5590	0.6374	0.5601	0.6376	119.3572	120.8523
	β_{20}	0.1516	0.0879	0.5012	0.7762	0.5242	0.7839	55.3431	55.7580
	β_{21}	−0.3017	−0.1640	1.7303	3.3837	1.8213	3.4106	68.5309	70.0988
100	β_{10}	−0.0254	−0.0343	0.0778	0.1357	0.0785	0.1369	14.5455	15.1311
	β_{11}	0.0193	0.0240	0.2704	0.3506	0.2708	0.3512	82.2737	83.4865
	β_{20}	0.0366	−0.0147	0.2353	0.5232	0.2366	0.5234	38.4661	40.5110
	β_{21}	−0.0511	0.0429	0.7113	2.3246	0.7139	2.3265	45.6203	48.1084
200	β_{10}	−0.0229	−0.0304	0.0299	0.0935	0.0305	0.0944	9.1091	9.9292
	β_{11}	0.0249	0.0294	0.1182	0.2163	0.1188	0.2172	54.5856	57.8233
	β_{20}	0.0342	−0.0059	0.0993	0.4401	0.1004	0.4401	24.9519	27.4913
	β_{21}	−0.0494	0.0357	0.2999	2.3711	0.3023	2.3724	29.5489	33.7693
500	β_{10}	−0.0065	−0.0194	0.0098	0.0780	0.0098	0.0784	5.2412	6.0263
	β_{11}	0.0013	0.0144	0.0337	0.1350	0.0337	0.1352	29.2489	32.1616
	β_{20}	0.0059	−0.0198	0.0335	0.3524	0.0353	0.3527	14.7641	17.4109
	β_{21}	−0.0033	0.0577	0.1043	2.0047	0.1043	2.0080	16.8862	20.9636
Scenario 2									
50	β_{10}	−0.0559	−0.0763	0.2565	0.3049	0.2596	0.3108	15.0565	15.4028
	β_{11}	0.0345	0.0284	1.2664	1.3063	1.2676	1.3071	54.2357	54.2181
	β_{20}	0.0140	−0.0185	0.4487	0.9193	0.4489	0.9196	104.6621	108.5400
	β_{21}	−0.2539	−0.1335	2.0746	2.9093	2.1391	2.9271	114.3879	114.9491
100	β_{10}	−0.0274	−0.0525	0.1300	0.1613	0.1308	0.1641	11.0662	11.0420
	β_{11}	0.0047	0.0321	0.4326	0.4518	0.4326	0.4528	34.5772	34.1142
	β_{20}	0.0030	−0.0412	0.2639	0.8869	0.2639	0.8886	81.9453	87.8518
	β_{21}	−0.1044	0.0113	0.8780	2.0218	0.8889	2.0220	74.6983	79.8030
200	β_{10}	−0.0332	−0.0526	0.0433	0.3165	0.0445	0.3193	6.7164	7.6425
	β_{11}	0.0358	0.0535	0.1627	0.4814	0.1640	0.4843	21.3950	23.3487
	β_{20}	−0.0061	−0.0393	0.0955	0.7542	0.0955	0.7557	48.9634	56.3251
	β_{21}	−0.0299	0.0217	0.3616	1.5927	0.3625	1.5932	47.6516	53.7506
500	β_{10}	−0.0116	−0.0309	0.0149	0.2812	0.0151	0.2821	3.9067	4.8219
	β_{11}	0.0051	0.0225	0.0519	0.3654	0.0520	0.3659	12.1947	13.7427
	β_{20}	−0.0035	−0.0573	0.0372	1.4334	0.0372	1.4367	31.1360	40.9934
	β_{21}	−0.0051	0.0658	0.1269	2.6745	0.1269	2.6788	28.5447	34.8024

Source: Elaborated by the author.

Table 23 – Summary of Bayesian and classical estimation using zero-deflated samples.

n	Parameter	Bias		Variance		MSE		MAPE (%)	
		Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
Scenario 1									
50	β_{10}	−0.1971	−0.1667	0.6135	0.7634	0.6524	0.7912	60.4995	59.4275
	β_{11}	0.1060	0.0973	1.9266	2.1244	1.9378	2.1338	216.5709	214.5939
	β_{20}	0.0941	0.0674	0.5204	0.6096	0.5292	0.6142	111.2615	110.3212
	β_{21}	−0.0105	−0.0356	2.1566	3.1154	2.1567	3.1167	116.0536	115.4830
100	β_{10}	−0.0773	−0.0813	0.2853	0.3727	0.2913	0.3794	41.2579	41.6971
	β_{11}	0.0226	0.0476	0.7260	0.8623	0.7265	0.8646	132.8333	133.6173
	β_{20}	−0.0188	−0.0356	0.2673	0.4108	0.2676	0.4121	79.8809	81.0349
	β_{21}	0.1173	0.1383	0.9410	2.1766	0.9548	2.1957	75.4413	78.0565
200	β_{10}	−0.0564	−0.0695	0.1131	0.3314	0.1163	0.3362	27.0471	29.4275
	β_{11}	0.0432	0.0645	0.3103	0.6435	0.3122	0.6476	88.5523	94.8071
	β_{20}	0.0058	−0.0218	0.0965	0.3417	0.0966	0.3421	49.6113	54.0982
	β_{21}	0.0297	0.0929	0.3488	2.5299	0.3497	2.5386	47.2081	53.7624
500	β_{10}	−0.0284	−0.0433	0.0394	0.1844	0.0402	0.1863	15.8450	17.6385
	β_{11}	0.0226	0.0434	0.1000	0.3317	0.1006	0.3336	49.7987	54.7585
	β_{20}	0.0026	−0.0195	0.0352	0.2622	0.0352	0.2626	29.4331	33.6326
	β_{21}	0.0155	0.0761	0.1299	2.3141	0.1301	2.3199	27.4785	34.1051
Scenario 2									
50	β_{10}	−0.5417	−0.3344	2.6364	2.1542	2.9298	2.2660	55.7422	49.1413
	β_{11}	0.1932	−0.0237	8.4866	8.4168	8.5239	8.4173	414.1736	397.8847
	β_{20}	0.1733	0.1010	0.5235	0.6591	0.5535	0.6693	57.8831	56.3176
	β_{21}	−0.2063	−0.1175	1.7803	3.3037	1.8228	3.3176	211.1536	212.7205
100	β_{10}	−0.2281	−0.1943	0.9002	0.9583	0.9522	0.9961	34.3945	33.8649
	β_{11}	0.0857	0.0934	2.1968	2.3735	2.2041	2.3822	222.6073	225.7564
	β_{20}	0.0325	−0.0097	0.2832	0.4358	0.2842	0.4359	41.4213	41.6895
	β_{21}	0.0074	0.0881	0.8323	2.4468	0.8324	2.4546	144.7771	150.7059
200	β_{10}	−0.1305	−0.1385	0.3088	0.9207	0.3258	0.9398	21.3299	22.6907
	β_{11}	0.0703	0.0971	0.9024	1.7183	0.9074	1.7277	149.3391	156.1515
	β_{20}	0.0219	−0.0181	0.1065	0.4449	0.1069	0.4452	25.6228	28.0340
	β_{21}	−0.0049	0.0909	0.3166	3.7195	0.3166	3.7277	90.1610	105.7611
500	β_{10}	−0.0438	−0.0675	0.0839	0.8042	0.0858	0.8087	11.5690	13.2514
	β_{11}	0.0073	0.0469	0.2345	1.1821	0.2346	1.1843	76.6633	84.2876
	β_{20}	0.0066	−0.0207	0.0373	0.3408	0.0373	0.3412	15.3647	17.9984
	β_{21}	0.0010	0.0765	0.1105	3.0407	0.1105	3.0465	52.7069	68.8181

Source: Elaborated by the author.

Table 24 – Coverage probabilities $(\%)$ of the BCIs/ACIs using zero-inflated samples.

n	Parameter	BNCP		CP		ANCP		BNCP		CP		ANCP	
		Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
		Scenario 1						Scenario 2					
50	β_{10}	2.60	1.40	93.40	95.80	4.00	2.80	4.40	2.40	90.20	93.60	5.40	4.00
	β_{11}	4.80	2.40	91.40	95.40	3.80	2.20	5.20	4.00	90.60	92.40	4.20	3.60
	β_{20}	6.60	2.40	91.00	95.40	2.40	2.20	5.80	2.80	90.80	95.40	3.40	1.80
	β_{21}	3.40	2.80	90.80	94.40	5.80	2.80	4.00	3.20	90.20	94.80	5.80	2.00
100	β_{10}	3.00	1.80	91.80	94.40	5.20	3.80	4.80	2.20	91.00	94.20	4.20	3.60
	β_{11}	5.60	4.20	90.40	93.20	4.00	2.60	4.60	2.80	92.00	95.60	3.40	1.60
	β_{20}	3.20	2.00	94.20	96.00	2.60	2.00	4.00	2.20	92.40	95.60	3.60	2.20
	β_{21}	3.80	2.40	92.80	96.60	3.40	1.00	3.00	3.60	92.60	94.80	4.40	1.60
200	β_{10}	3.80	1.60	91.00	94.60	5.20	3.80	3.40	2.40	91.60	94.00	5.00	3.60
	β_{11}	6.00	4.40	89.20	91.20	4.80	4.40	4.80	3.00	92.60	94.80	2.60	2.20
	β_{20}	4.60	3.00	92.80	94.40	2.60	2.60	3.80	2.40	93.40	95.60	2.80	2.00
	β_{21}	3.00	1.80	93.00	95.60	4.00	2.60	3.20	2.80	92.20	94.40	4.60	2.80
500	β_{10}	4.60	3.60	92.20	93.80	3.20	2.60	2.40	1.80	92.60	95.20	5.00	3.00
	β_{11}	3.20	2.80	93.00	94.40	3.80	2.80	4.20	3.00	92.40	94.80	3.40	2.20
	β_{20}	3.40	2.00	92.00	94.80	4.60	3.20	3.80	3.00	93.20	94.80	3.00	2.20
	β_{21}	4.20	4.00	91.80	93.80	4.00	2.20	3.80	2.60	92.20	94.60	4.00	2.80

Source: Elaborated by the author.

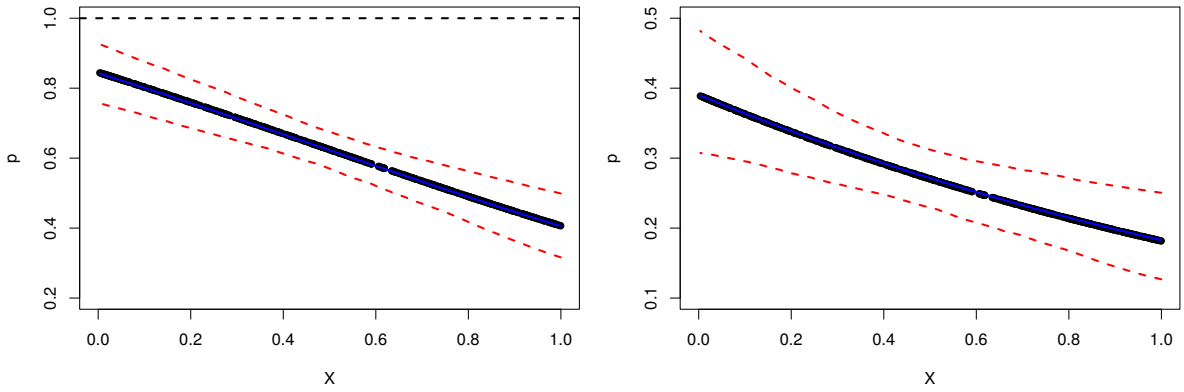
Table 25 – Coverage probabilities $(\%)$ of the BCIs/ACIs using zero-deflated samples.

n	Parameter	BNCP		CP		ANCP		BNCP		CP		ANCP	
		Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE	Bayes	MLE
		Scenario 1						Scenario 2					
50	β_{10}	2.80	1.80	92.40	96.20	4.80	2.00	3.00	2.40	91.80	97.60	5.20	0.00
	β_{11}	4.00	2.80	92.80	95.20	3.20	2.00	4.00	1.60	91.80	97.80	4.20	0.60
	β_{20}	6.40	1.80	89.60	96.00	4.00	2.20	6.40	2.00	91.40	96.20	2.20	1.80
	β_{21}	3.40	1.40	91.40	95.80	5.20	2.80	3.00	1.60	90.40	95.80	6.60	2.60
100	β_{10}	3.40	2.40	92.80	95.60	3.80	2.00	2.80	2.60	91.00	96.60	6.20	0.80
	β_{11}	4.00	2.80	92.60	95.00	3.40	2.20	6.20	2.20	90.00	95.40	3.80	2.40
	β_{20}	3.20	1.60	92.60	95.00	4.20	3.40	4.60	1.40	91.20	94.80	4.20	3.80
	β_{21}	5.60	3.20	92.00	95.60	2.40	1.20	4.00	3.40	91.60	94.00	4.40	2.60
200	β_{10}	2.40	2.00	94.40	96.20	3.20	1.80	4.00	2.60	90.20	94.00	5.80	3.40
	β_{11}	3.40	2.20	93.80	95.60	2.80	2.20	5.80	3.40	90.40	95.00	3.80	1.60
	β_{20}	3.40	2.00	93.00	94.40	3.60	3.60	4.40	2.60	92.00	94.00	3.60	3.40
	β_{21}	2.40	1.80	93.60	95.20	4.00	3.00	3.20	3.00	92.40	93.80	4.40	3.20
500	β_{10}	3.00	2.80	93.80	94.40	3.20	2.80	2.60	2.40	93.40	95.40	4.00	2.20
	β_{11}	3.80	3.20	93.40	95.20	2.80	1.60	3.80	3.40	93.20	94.60	3.00	2.00
	β_{20}	2.60	1.40	93.20	95.80	4.20	2.80	2.60	1.80	93.20	95.20	4.20	3.00
	β_{21}	4.40	2.80	91.80	93.80	3.80	3.40	5.20	3.20	92.00	94.80	2.80	2.00

Source: Elaborated by the author.

expected and therefore, the variable Y was generated with mean (μ) close to zero. Then, for the third scenario we have considered $\boldsymbol{\beta}_1^\intercal = (-1.0, 0.5)$ and $\boldsymbol{\beta}_2^\intercal = (0.5, 1.0)$ to perform the simulation. In the following, the procedure was repeated by considering $\boldsymbol{\beta}_1^\intercal = (-2.0, 0.5)$ and $\boldsymbol{\beta}_2^\intercal = (1.0, -0.5)$. For these scenarios, the "vague" *prior* setup is $(\boldsymbol{\beta}_1^0)^\intercal = (0.0, -1.0)$ and $(\boldsymbol{\beta}_2^0)^\intercal = (2.0, 0.0)$. Figure 14 depicts the Bayesian estimates obtained for parameter p using zero-deflated samples with $n = 500$ for each scenario. Such representation has the same characteristics of Figure 13.

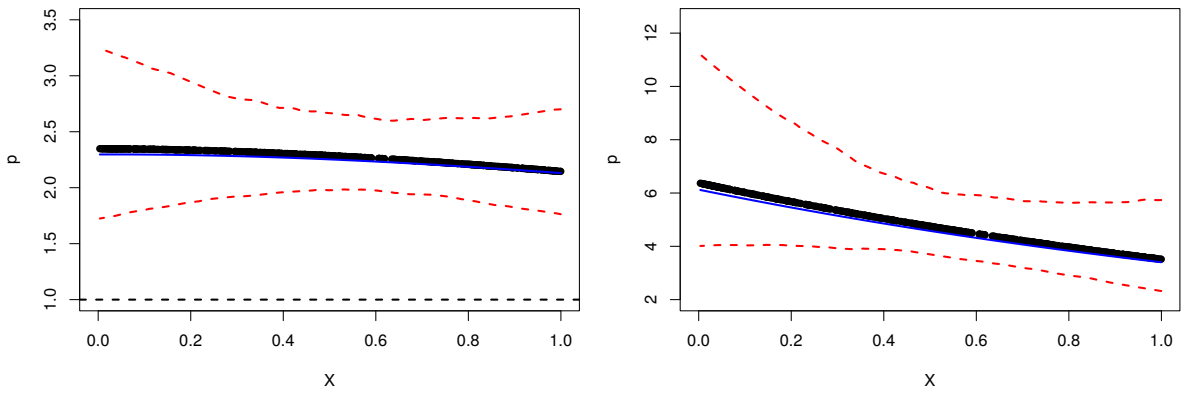
Figure 13 – *Posterior* estimates of parameter p using zero-inflated samples $(n = 500)$.



(a) Scenario 1 (b) Scenario 2

Source: Elaborated by the author.

Figure 14 – *Posterior* estimates of parameter p using zero-deflated samples $(n = 500)$.



(a) Scenario 1 (b) Scenario 2

Source: Elaborated by the author.

4.7 Takeover bids data analysis

In this section, the $\mathcal{ZMPL}$ regression model is considered for the analysis of a real dataset obtained from Jaggia and Thosar (1993). The sample consists of 126 US firms that

were targets of tender offers between 1978-1985 and which were taken over within a period of 52 weeks. In this study, the response variable is the number of additional bids after the initial bid received by the target firms. Also, a set of explanatory variables regarding target management actions and firm-specific characteristics was observed. The authors have analyzed the data by fitting a Poisson regression model, and they have verified that the *white knight* is one of the covariates associated with additional bids. The *white knight* is a management action of inviting a friendly third-part to enter the bidding. The authors pointed out that, when inviting a friendly bidder, the management is indicating that may cede at least some control of the firm, and therefore, the entry (or potential entry) of at least one additional bidder is expected to spur the auction process.

Let us characterize the number of bids as the response variable (Y) and the *white knight* as a covariate (x). The independent variable x was coded as 0 (no additional bidder) and 1 (additional bidder). From the observed dataset, there exists evidence that Y is overdispersed since its mean is 1.74, and its variance is 2.05. Also, the range of Y is 10, and its coefficient of variation is approximately 118%. In this study, 75 out of 126 firms have at least one additional bidder invited for the process. The average number of bids was 1.18 when no additional bidder was invited and 2.12 otherwise. The absolute frequency of zeros is 9 (about 7% of the entire sample), which naively indicates a zero deflation. Such characteristic is evidenced when fitting model (4.10) using the full dataset. This procedure was performed strictly for descriptive purposes. The model was fitted using a simpler version of Algorithm 5 since covariate x was not used. From its *posterior* summary, we have estimated Equation (4.8) as $(1-1.87)\mathrm{P}(Y>0;0.93)\times 100 \approx -43\%$, suggesting that exist approximately 7 missing zeros, reinforcing our suspect that the sample is zero-deflated. Moreover, we had noticed that when the sample was observed, 6 firms had received no further bids when no additional bidder was invited. As would be expected, the firm which has received the highest number of additional bids (10) had at least one additional bidder invited.

Table 26 – *Posterior* parameter estimates and 95% BCIs/HPDIs from $\mathcal{ZMPL}$ fitted model.

Parameter	Mean	Median	SD	ESS	95% BCI		95% HPDI	
					Lower	Upper	Lower	Upper
β_{10}	−0.8367	−0.8332	0.2563	2368.97	−1.3684	−0.3446	−1.3620	−0.3391
β_{11}	0.9354	0.9319	0.2972	2293.12	0.3665	1.5421	0.3806	1.5529
β_{20}	2.0608	2.0458	0.4055	2406.81	1.3147	2.9033	1.3266	2.9131
β_{21}	0.8104	0.8140	0.5965	2400.04	−0.3397	1.9806	−0.3064	2.0073

Source: Elaborated by the author.

To fit the $\mathcal{ZMPL}$ regression model with the *white knight* as a covariate, we have adopted a similar procedure to that one used in the previous section. Since x is a dummy variable, we ha-

Table 27 – Sensitivity analysis to evaluate the effect of different *prior* specifications.

Priors	Parameter	Mean	$\hat{\mu}_*$	$\hat{p}$	$\hat{n}_0$
$\tau_1 = \tau_2 = 2.0$					
$(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$ $(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$	β_{10} β_{11} β_{20} β_{21}	-0.64 0.56 2.04 0.57	1.63	2.15	11
$(\boldsymbol{\beta}_1^p)^\intercal = (4.0, -4.5)$ $(\boldsymbol{\beta}_2^p)^\intercal = (4.0, -1.0)$	β_{10} β_{11} β_{20} β_{21}	-0.07 0.09 2.25 0.93	1.81	1.83	8
$(\boldsymbol{\beta}_1^p)^\intercal = (8.5, -9.5)$ $(\boldsymbol{\beta}_2^p)^\intercal = (10.0, -7.0)$	β_{10} β_{11} β_{20} β_{21}	-1.05 0.98 2.09 1.10	1.75	2.54	9
$\tau_1 = \tau_2 = 5.0$					
$(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$ $(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$	β_{10} β_{11} β_{20} β_{21}	-0.84 0.94 2.06 0.81	1.84	2.22	10
$(\boldsymbol{\beta}_1^p)^\intercal = (4.0, -4.5)$ $(\boldsymbol{\beta}_2^p)^\intercal = (4.0, -1.0)$	β_{10} β_{11} β_{20} β_{21}	-0.58 0.72 2.15 1.10	1.85	2.03	8
$(\boldsymbol{\beta}_1^p)^\intercal = (8.5, -9.5)$ $(\boldsymbol{\beta}_2^p)^\intercal = (10.0, -7.0)$	β_{10} β_{11} β_{20} β_{21}	-1.08 1.18 2.09 1.17	1.95	2.47	8
$\tau_1 = \tau_2 = 8.0$					
$(\boldsymbol{\beta}_1^p)^\intercal = (0.0, -1.0)$ $(\boldsymbol{\beta}_2^p)^\intercal = (2.0, 0.0)$	β_{10} β_{11} β_{20} β_{21}	-0.92 1.07 2.06 0.93	1.93	2.27	9
$(\boldsymbol{\beta}_1^p)^\intercal = (4.0, -4.5)$ $(\boldsymbol{\beta}_2^p)^\intercal = (4.0, -1.0)$	β_{10} β_{11} β_{20} β_{21}	-0.75 0.93 2.13 1.16	1.92	2.13	8
$(\boldsymbol{\beta}_1^p)^\intercal = (8.5, -9.5)$ $(\boldsymbol{\beta}_2^p)^\intercal = (10.0, -7.0)$	β_{10} β_{11} β_{20} β_{21}	-1.08 1.23 2.09 1.19	2.03	2.45	8

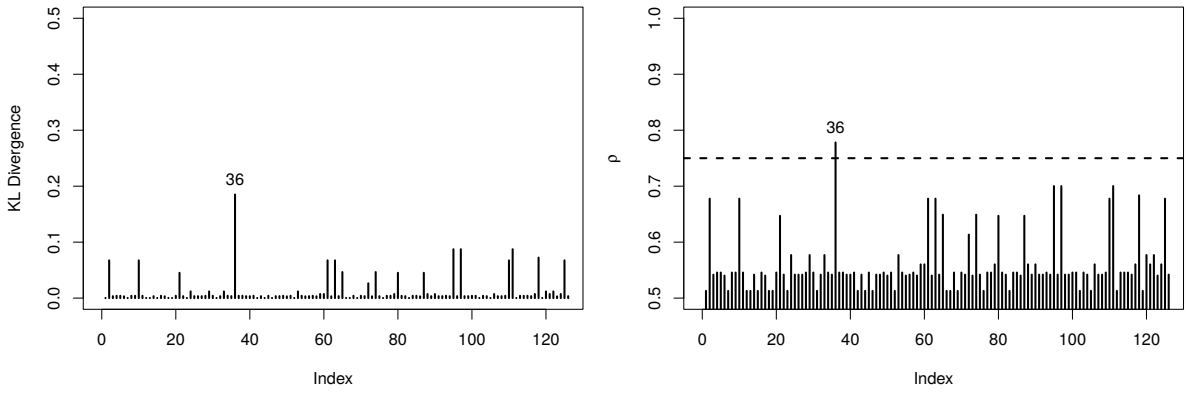
Source: Elaborated by the author.

ve fixed 0 (no additional bidder) as the baseline for estimation purposes. The logarithm link function was considered to relate μ_i with the linear predictor $\beta_{10}+\beta_{11}x_i$.

To relate parameter ω_i with $\beta_{20}+\beta_{21}x_i$, we choose the link function given by (4.13). In this framework, the parameter β_{11} represents the indirect effect of the invitation of at least one additional bidder on the mean (μ_*) and parameter β_{21} indicates the direct effect of such invitation on the probability of zeros $(1-\omega)$. We have considered the RwM algorithm, generating a chain of size $N = 50,000$ for each parameter whereby the first 10,000 values were discarded as burn-in. The stationarity of each chain was checked through the Geweke criterion for diagnosing convergence. To obtain the pseudo-independent samples, we have kept one out of every 10 generated values, resulting in chains of size $M = 4,000$ for each parameter.

Table 26 presents the mean, the median and the standard deviation obtained from the *posterior* distribution of $\boldsymbol{\beta}$. To obtain the full descriptive summary we have arbitrarily selected the *priors* $(\boldsymbol{\beta}_1^0)^\intercal = (0.0,-1.0)$, $(\boldsymbol{\beta}_2^0)^\intercal = (2.0,0.0)$ and $\tau_1 = \tau_2 = 5.0$. In this framework, the acceptance rates in the RwM algorithm were at approximately 35%. In addition, we have calculated the number of effectively pseudo-independent draws (effective sample size - ESS) from the *posterior* distribution. The 95% BCIs were estimated empirically from the generated samples and the 95% highest *posterior* density intervals (HPDIs) were also computed.

Figure 15 – Sensitivity analysis for diagnostic of influential points.



(a) Kullback-Leibler divergence (b) Calibration (ρ)

Source: Elaborated by the author.

A sensitivity analysis was performed to evaluate the behaviour of the Bayesian estimators under distinct *prior* specifications. We have established three scenarios where the model parameters were estimated considering different choices for $\boldsymbol{\beta}_r^0$ and τ_r $(r = 1,2)$. The results are displayed on Table 27. The Bayesian estimates for the parameters μ_*, p and n_0 (expected number of zeros) are also presented. Here, parameters μ_* and p were estimated as functions of the fitted $\mathcal{ZMPL}$ model, that is, $\hat{\mu}_* = n^{-1}\sum_{i=1}^n \hat{\mu}_i\hat{p}_i$, where $\hat{p}_i = \hat{\omega}_i \mathrm{P}^{-1}(Y > 0;\hat{\mu}_i)$, and naturally, $\hat{p} = n^{-1}\sum_{i=1}^n \hat{p}_i$. Since $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal, the results can be directly combined by taking $\tau_1 \neq \tau_2$. Obviously, we cannot decide on the *prior* distributions based on the *posterior* results,

but we can investigate whether the *prior* specifications are influential. In this way, we have observed that the estimator of β_{20} is less sensitive regarding the *prior* choice, with variance smaller than 0.005 between estimates. For $\tau_j = 2.0$, the estimators of β_{10} and β_{11} are more sensitive, implying higher variability between estimates of parameter p. In addition, when large values are selected for τ_r, one can notice that the lower the *prior* choices for β_{11} and β_{21} the larger the estimates of μ_*. Amidst these features, it is worthwhile to mention that, as τ_r increases, the *prior* specification tends to have lower impact on the final fit $(\hat{\mu}_*, \hat{p}, \hat{n}_0)$ and, in general, such impact can be considered negligible, even for quite distinct *priors*.

Table 28 – Comparison criteria for the fitted models.

Model	Full Dataset				Without Observation 36			
	DIC	EAIC	EBIC	LMPL	DIC	EAIC	EBIC	LMPL
$\mathcal{P}$	1011.60	409.04	414.71	−195.53	959.15	388.06	393.72	−185.45
$\mathcal{NB}$	988.68	402.04	405.71	−195.87	943.12	383.81	387.47	−185.97
$\mathcal{PL}$	1104.06	446.03	451.70	−291.00	1075.44	434.58	440.24	−213.58
$\mathcal{ZMP}$	941.45	385.37	387.04	−183.65	884.90	362.77	364.42	−172.56
$\mathcal{ZMPL}$	920.02	376.81	378.49	−178.43	882.32	361.71	363.37	−171.40

Source: Elaborated by the author.

Table 29 – *Posterior* parameter estimates and 95% BCIs/HPDIs from $\mathcal{ZMPL}$ fitted model (without influential point).

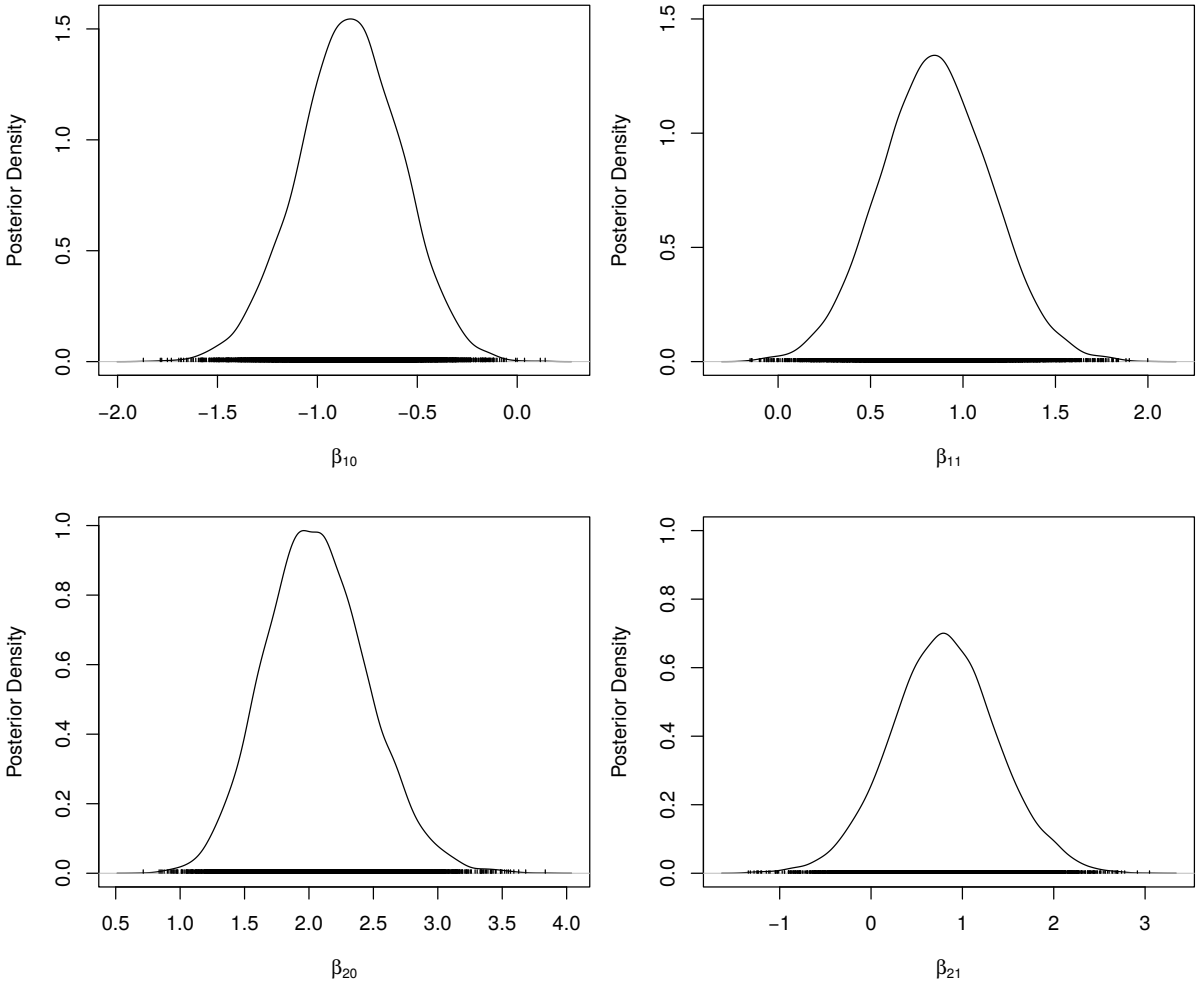
Parameter	Mean	Median	Std. Dev.	ESS	95% BCI		95% HPDI	
					Lower	Upper	Lower	Upper
β_{10}	−0.8395 (−0.33%)	−0.8367 (−0.42%)	0.2549 (−0.55%)	2597.71 (9.66%)	−1.3524	−0.3481	−1.3524	−0.3479
β_{11}	0.8516 (−8.96%)	0.8493 (−8.86%)	0.2979 (0.24%)	2394.79 (4.43%)	0.2723	1.4447	0.2615	1.4326
β_{20}	2.0622 (0.07%)	2.0477 (0.09%)	0.4029 (−0.64%)	2420.74 (0.58%)	1.3267	2.8911	1.2674	2.8193
β_{21}	0.8002 (−1.26%)	0.7951 (−2.32%)	0.5903 (−1.04%)	2848.08 (18.67%)	−0.3396	1.9810	−0.2880	2.0199

Source: Elaborated by the author.

The analysis to verify the existence of influential points is presented in Figure 15. Figure 15a depicts the Kullback-Leibler (KL) divergence (see Appendix B, Section B.1), that used to evaluate the effect of each observation on the parameter estimates. Conservatively, we consider an observation whose distance has a calibration exceeding 0.75 as an influential point. Based on Figure 15b, we have observed the existence of one influential point (36), corresponding to the firm with ten additional bids. As a way to access the influence of this observation, the estimation process was repeated considering the removal of such a firm of the sample. The *posterior* summary for this case and the variation percentage regarding the *posterior* summary

obtained from the full dataset is presented on Table 29. When analyzing the parameter estimates, it can be observed that the removal of observation 36 impacts reasonably the model fit, and the main variation is observed when estimating parameter β_{11}.

Figure 16 – Estimated *posterior* densities of vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.



Source: Elaborated by the author.

For comparison purposes, identical Bayesian procedures were adopted to fit the $\mathcal{P}$, the $\mathcal{PL}$, the Negative Binomial ($\mathcal{NB}$), and the $\mathcal{ZMP}$ regression models. To estimate the fixed dispersion parameter (ϕ) of the $\mathcal{NB}$ model, we have considered a noninformative Inverse-Gamma *prior* distribution with hyperparameters $a = b = 1.0$. For each fitted model, we have estimated the measures presented in Appendix B (Section B.2). The model comparison procedure is summarized in Table 28. One can notice that the zero-modified models have performed considerably better with $\mathcal{ZMPL}$ outperforming all. These results are highlighting that the proposed model is highly competitive with well-established models in the literature.

Figure 16 presents the marginal *posterior* densities of parameters of the $\mathcal{ZMPL}$ regression model. These densities provided the summary displayed on Table 29. The assumption of normality for the generated chains is quite reasonable even in the presence of slightly heavy tails

on some of the estimated densities. Besides, there exists evidence of symmetry since *posterior* mean and median are very close to each other. For each parameter, the effective sample size was estimated greater than $M/2$, which can be considered an indication of good mixing of the generated chains, without any computational waste.

From the results displayed in Table 29, one can make some inferences and take some conclusions. Firstly, we have observed that the BCI/HPDI of the parameter β_{11} does not contain the value zero, which constitute the *white knight* as a relevant covariate to describe the average number of bids. On the other hand, the probability of not receive at least one additional bid is $1-[1+\exp\{-2.0622\}]^{-1} \approx 0.113$ if no additional bidder is invited and $1-[1+\exp\{-(2.0622+0.8002)\}]^{-1} \approx 0.054$ otherwise. However, as parameter ω is not affected by individual *white knights*, one can re-estimate it, not depending on individuals' covariate. Hence, the fitted $\mathcal{ZMPL}$ model can be represented by

$$\hat{\mu}_{i*} = \frac{\hat{\omega}\,\hat{\mu}_i}{\mathrm{P}(Y>0;\hat{\mu}_i)},$$

where $\hat{\mu}_i = \exp\{-0.8395+0.8516x_i\}$ and $\hat{\omega} = 0.9252$. Also, since parameter μ_i was estimated using only positive observations, if at least one additional bid is being offered, its expected value will be 1.312 provided that no additional bidder is invited and 1.808 otherwise.

Table 30 – *Posterior* estimates of extra parameters and goodness-of-fit evaluation.

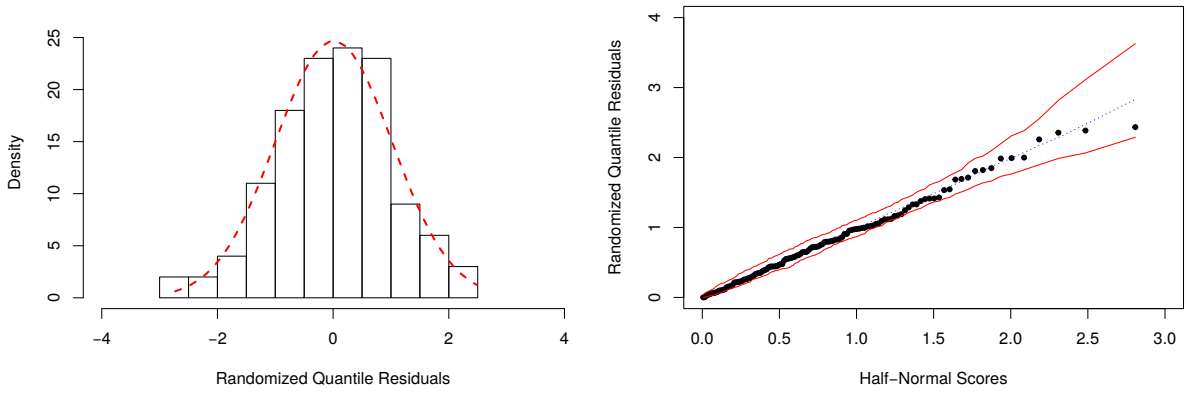
Model	Full Dataset				Without Observation 36			
	$\hat{\mu}_*$	$\hat{p}$	$\hat{n}_0$	χ^2	$\hat{\mu}_*$	$\hat{p}$	$\hat{n}_0$	χ^2
$\mathcal{P}$	1.65	1.00	24	26.19 (<0.001)	1.58	1.00	26	24.65 (<0.001)
$\mathcal{NB}$	1.67	1.00	25	27.50 (<0.001)	1.62	1.00	25	25.75 (<0.001)
$\mathcal{PL}$	1.60	1.00	45	69.67 (<0.001)	1.56	1.00	46	70.59 (<0.001)
$\mathcal{ZMP}$	1.88	1.41	10	8.75 (0.119)	1.78	1.43	10	4.31 (0.366)
$\mathcal{ZMPL}$	1.84	2.22	10	0.61 (0.986)	1.77	2.27	9	0.65 (0.957)

Source: Elaborated by the author.

Table 30 presents the final *posterior* summary of the fitted models. One can notice that the estimates for n_0, obtained from the $\mathcal{P}$, $\mathcal{NB}$, and $\mathcal{PL}$ models are much larger than the real one while those provided by zero-modified models are very close (or exactly equal) to 9. Through these measures, one can better understand how the fitted models are adhering to the data since the nature of the observed counts should be well described regarding its frequency and the average number of nonzero observations. The goodness-of-fit can be evaluated by the χ^2 measure obtained from the observed and expected frequencies. To compute such statistics, we

have grouped cells with frequencies lower or equal than 5, resulting in 5 df (full dataset) and 4 df (when removing observation 36). One can notice that $\mathcal{ZMP}$ model provides reliable fits, but it is quite clear that the proposed model adheres much better on the considered datasets (p-values greater than 0.95).

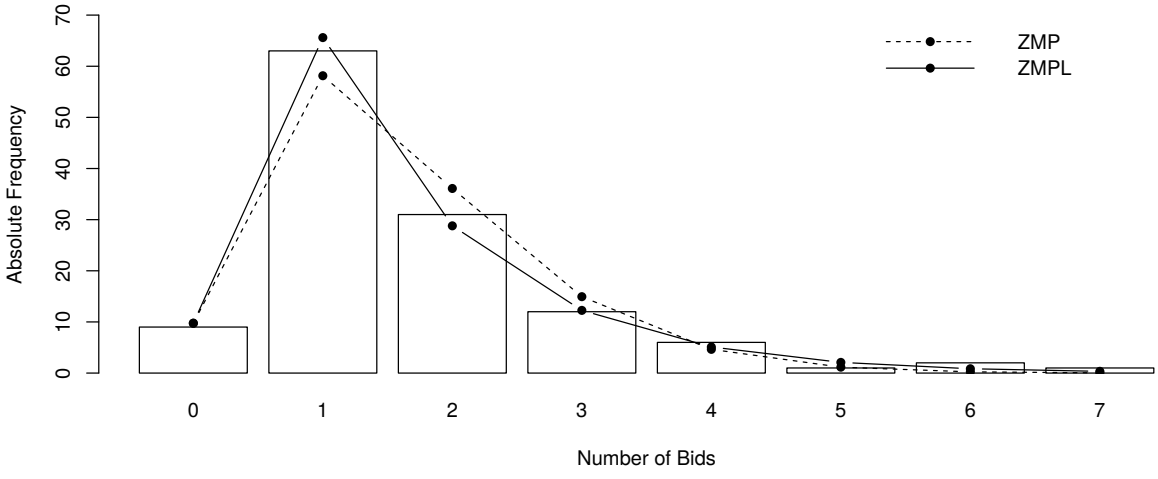
Figure 17 – Frequency distribution and Half-Normal plot with simulated envelope for the RQRs.



Source: Elaborated by the author.

Figure 17 depicts additional evidence based on the randomized quantile residuals (RQRs) for validating the fitted $\mathcal{ZMPL}$ regression model (without observation 36). This residual metric was computed as discussed in Appendix B (Section B.3) using Equation (4.11). One can notice that the normality assumption of the residuals is easily verified by the behavior of its frequency distribution (left-panel). Besides, the Half-Normal probability plot indicates that the fit of the $\mathcal{ZMPL}$ model was very satisfactory since all estimated residuals are lying within the simulated envelope (right-panel).

Figure 18 – *Posterior* expected frequencies under zero-modified models.



Source: Elaborated by the author.

Figure 18 depicts the expected frequencies estimated through predictive $\mathcal{ZMP}$ and $\mathcal{ZMPL}$ models, considering the removal of the influential point. The results highlight the better adherence of the fitted models when observation 36 is discarded in the estimation procedure. Besides, using the χ^2 statistic and the comparison criteria, one can notice that the proposed model provides a more realistic fit for the considered dataset, and inferences about parameter p allow us to classify the observed sample as being zero-deflated ($\hat{p} > 1$). In other words, we have that, by the proposed model, it would be expected that more firms would not have received additional bids after the initial bid.

4.8 Concluding remarks

In this chapter, we have introduced the $\mathcal{ZMPL}$ regression model as an alternative for the analysis of overdispersed datasets exhibiting zero inflation/deflation in the presence of covariates. By using the hurdle version of the $\mathcal{PL}$ distribution, it was possible to write separable likelihood functions for the parameter vectors, which led us to less complicated Bayesian procedures based on the *g-prior* method. Also, we have shown that the mean of the $\mathcal{ZMPL}$ model can be estimated using only the positive observations.

An intensive Monte Carlo simulation study was performed in order to evaluate the empirical properties of the Bayesian estimators and MLEs, and the obtained results highlighted the suitability of the adopted methodology. Due to the "vague" nature of the *prior* distributions, similar results were achieved, but the Bayesian approach remains an excellent option since it does not depend on asymptotic results for inference and has the advantage of incorporating specific information about parameters when available.

The proposed model was considered for the analysis of a real dataset obtained from an economic study with legal implications, where the response variable is the number of additional bids after the initial bid received by 126 US firms. The response variable was identified as being overdispersed and zero-deflated, which justifies the use of the $\mathcal{ZMPL}$ model. A sensitivity analysis was conducted using the Kullback-Leibler divergence, and one firm was identified as locally influent. The main inferential conclusion one can take from the fitted model is that the *white knight* is statistically relevant to describe the average number of additional bids. Besides, when looking at the χ^2 statistic and the *posterior* based comparison criteria, we have noticed that the proposed model had presented a better fit when compared with its competitors. Therefore, the $\mathcal{ZMPL}$ regression model can be considered an excellent addition to the set of models that can be used when analyzing overdispersed and zero-modified count data.

CHAPTER

5

THE ZERO-MODIFIED POISSON-SUJATHA REGRESSION MODEL

5.1 Introduction

Count datasets are traditionally analyzed using the ordinary Poisson $(\mathcal{P})$ distribution. However, such a model has its applicability limited as it can be somewhat restrictive to handle specific data structures. In this case, it arises the need for obtaining alternative models that accommodate, for example, overdispersion and zero modification (inflation/deflation at the frequency of zeros). In practical terms, these are the most prevalent structures ruling the nature of discrete phenomena nowadays. Hence, the primary goal of this chapter was to jointly address these issues by deriving a fixed-effects regression model based on the hurdle version of the Poisson-Sujatha $(\mathcal{PS}_u)$ distribution. In this framework, the zero modification is incorporated by considering that a binary probability model determines which outcomes are zero-valued, and a zero-truncated process is responsible for generating positive observations. *Posterior* inferences for the model parameters were obtained from a fully Bayesian approach based on the *g-prior* method. Intensive Monte Carlo simulation studies were performed as a way to assess the empirical properties of the Bayesian estimators, and the obtained results have been discussed. The proposed model was considered for the analysis of a real dataset, and its competitiveness regarding some well-established fixed-effects models for count data was evaluated. A sensitivity analysis to detect observations that may impact parameter estimates was performed based on some standard divergence measures. The Bayesian p-value and the randomized quantile residuals were considered for the task of model validation.

In this chapter, we have extended the works of Shanker (2016d), Bertoli *et al.* (2018), and Bertoli *et al.* (2019a) in the sense of developing a new fixed-effects regression model for zero-modified count data based on the $\mathcal{PS}_u$ distribution. A discrete random variable Y defined on $\mathcal{Y}_0 = \{0, 1, \ldots\}$ is said to follow a zero-modified Poisson-Sujatha $(\mathcal{ZMPS}_u)$ distribution if

its probability mass function (pmf) can be written as

$$\mathrm{P}_*(Y=y;\mu,p)=(1-p)\,\delta_y+p\,\mathrm{P}(Y=y;\mu),\quad y\in\mathcal{Y}_0,$$

where p is the zero modification parameter and δ_y is an indicator function, so that $\delta_y=1$ if $y=0$ and $\delta_y=0$ otherwise. Also, $\mu\in\mathbb{R}_+$ is the expected value of the ordinary $\mathcal{PS}_u$ distribution, whose reparameterized pmf is given by

$$\mathrm{P}(Y=y;\mu)=\frac{h^3(\mu)}{h^2(\mu)+h(\mu)+2}\left[\frac{y^2+y\,[h(\mu)+4]+\left[h^2(\mu)+3h(\mu)+4\right]}{[h(\mu)+1]^{y+3}}\right],\quad y\in\mathcal{Y}_0,$$

where

$$h(\mu)=\frac{1}{3\mu}\left[(s(\mu)-\mu+1)-\frac{(\mu-1)(5\mu+1)}{s(\mu)}\right],\tag{5.1}$$

with

$$s(\mu)=\left[3\mu\sqrt{21\mu^4+84\mu^3+513\mu^2+96\mu+15}+2\mu\left(4\mu^2+33\mu+3\right)+1\right]^{1/3},$$

and $\mu=(\theta^2+2\theta+6)[\theta(\theta^2+\theta+2)]^{-1}$ for $\theta\in\mathbb{R}_+$ (shape parameter). This parameterization is particularly useful since our primary goal is to derive a regression model, in which the influence of fixed-effects can be evaluated directly over the mean of a zero-modified response variable. Unlike zero-inflated models, here parameter p is defined on the interval $[0,\mathrm{P}^{-1}(Y>0;\mu)]$, and so the $\mathcal{ZMPS}_u$ model is not a mixture distribution since p may assume values greater than 1. The expected value and variance of Y are given, respectively, by $\mathbb{E}(Y)=\lambda=\mu p$ and $\mathbb{V}(Y)=\varsigma^2=p[\sigma^2+(1-p)\mu^2]$, where $\sigma^2\in\mathbb{R}_+$ is the variance of the $\mathcal{PS}_u$ distribution (see Bertoli *et al.* (2019a), Table 4).

The hurdle version of the $\mathcal{PS}_u$ distribution can be obtained by taking $\omega=p\,\mathrm{P}(Y>0;\mu)$, that is,

$$\mathrm{P}_*(Y=y;\mu,\omega)=(1-\omega)\,\delta_y+\omega\,\mathrm{P}^*(Y=y;\mu),\quad y\in\mathcal{Y}_0,\tag{5.2}$$

for $\omega\in[0,1]$ and where $\mathrm{P}^*(Y=y;\mu)$ is the pmf of the zero-truncated Poisson-Sujatha ($\mathcal{ZTPS}_u$) distribution (SHANKER; FESSHAYE, 2016). Noticeably, Equation (5.2) is only a reparameterization of the standard $\mathcal{ZMPS}_u$, and so one can conclude that these models are interchangeable. For ease of notation and understanding, the acronym $\mathcal{ZMPS}_u$ will be used when we refer to the hurdle version of the $\mathcal{PS}_u$ distribution.

The corresponding cumulative distribution function (cdf) of Y is given by

$$\begin{aligned}\mathrm{F}_*(y;\mu,\omega) &= 1-\frac{\omega}{\mathrm{P}(Y>0;\mu)}\left\{\frac{yh(\mu)\left[h^2(\mu)+(y+6)\,h(\mu)+2\right]}{\left[h^2(\mu)+h(\mu)+2\right]\left[h(\mu)+1\right]^{y+3}}+\right.\\ &\left.\frac{h^4(\mu)+4h^3(\mu)+10h^2(\mu)+7h(\mu)+2}{\left[h^2(\mu)+h(\mu)+2\right]\left[h(\mu)+1\right]^{y+3}}\right\},\quad y\in\mathcal{Y}_0.\end{aligned}\tag{5.3}$$

Comparatively, the proposed model can be considered more flexible than zero-inflated models as it allows for zero deflation, which is a structure often encountered when handling

count data. Besides, the proposed model incorporates overdispersion that does not come only from inflation/deflation of zeros, as one of its parts is dedicated to describing the behavior of the positive values. In the regression framework that we have developed, discrepant points (*outliers*) can be identified, and, through a careful sensitivity analysis, it is possible to quantify the influence of such observations. However, since the $\mathcal{PS}_u$ distribution accounts for different levels of overdispersion, its zero-modified version is naturally a robust alternative as it may accommodate discrepant points that would significantly impact the parameter estimates of the zero-modified Poisson ($\mathcal{ZMP}$) model.

This chapter is organized as follows. In Section 5.2, we present the fixed-effects regression model based on the hurdle version of the $\mathcal{PS}_u$ distribution. In Section 5.3, we describe all the Bayesian methodologies and associated numerical procedures that were considered for inferential purposes. In Section 5.4, we discuss the results of an intensive simulation study and in Section 5.5, a real data application using the proposed model is exhibited. General comments and concluding remarks are addressed in Section 5.6.

5.2 The $ZMPS_u$ regression model

Suppose that a random experiment (designed or observational) is conducted with n subjects. The primary response for such an experiment is described by a discrete random variable Y_i denoting the outcome for the i-th subject ($i = 1,\ldots,n$). The full response vector is given by $\boldsymbol{Y} = (Y_1,\ldots,Y_n)$, and we assume that the observed vector $\boldsymbol{y}$ is obtained conditionally to fixed-effects, here denoted by $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Assuming that $Y_i|\boldsymbol{\beta} \sim \mathcal{ZMPS}_u(\mu_i, \omega_i)$ holds for all i, a general fixed-effects regression model for count data based on the $\mathcal{ZMPS}_u$ distribution can be derived by rewriting Equation (5.2) as

$$\mathrm{P}_*(Y_i = y_i;\boldsymbol{\beta}) = (1-\omega_i)\,\delta_{y_i} + \omega_i \mathrm{P}^*(Y_i = y_i;\mu_i)\,, \qquad y_i \in \mathcal{Y}_0, \tag{5.4}$$

where $\mu_i \equiv \mu(\boldsymbol{x}_{1i},\boldsymbol{\beta}_1)$ and $\omega_i \equiv \omega(\boldsymbol{x}_{2i},\boldsymbol{\beta}_2)$ are parameterized nonlinear functions. In this framework, we have $\boldsymbol{\beta}_k^\top = (\beta_{k0},\ldots,\beta_{kq_k})$ $(k = 1,2)$ related to $\boldsymbol{x}_{ki}^\top = (1, x_{ki}^1,\ldots, x_{ki}^{q_k})$, where $\boldsymbol{x}_{ki}$ is a vector of covariates that may include, for example, *dummy* variables, cross-level interactions and polynomials. The quantity q_1 (q_2) denotes the number of covariates considered on the systematic component of a linear predictor for parameter μ_i (ω_i). The full regression matrices of model (5.4) can be written as $\boldsymbol{X}_k = (\boldsymbol{1}_n, \boldsymbol{X}_{k,n\times q_k})$, where $\boldsymbol{1}_n$ is the intercept column and the submatrix $\boldsymbol{X}_{k,n\times q_k}$ is defined in such a way that its i-th row contains the vector $(x_{ki}^1,\ldots,x_{ki}^{q_k})$. The overall dimension of $\boldsymbol{X}_k$ is $n\times(q_k+1)$.

Now, we have to specify two monotonic, invertible and twice differentiable link functions, say g_1 and g_2, in which $\mu_i = g_1^{-1}(\boldsymbol{x}_{1i}^\top\boldsymbol{\beta}_1)$ and $\omega_i = g_2^{-1}(\boldsymbol{x}_{2i}^\top\boldsymbol{\beta}_2)$ are well defined on $\mathbb{R}_+$ and $(0,1)$, respectively. For this purpose, one may choose any suitable mappings g_1 and g_2 such that $g_1^{-1}\colon \mathbb{R}\to\mathbb{R}_+$ and $g_2^{-1}\colon \mathbb{R}\to(0,1)$. The logarithm link function, $\log(\mu_i) = \boldsymbol{x}_{1i}^\top\boldsymbol{\beta}_1$, is the natural

choice for g_1. For g_2, the popular choice is the *logit* link function,

$$\text{logit}(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = \boldsymbol{x}_{2i}^{\top}\boldsymbol{\beta}_2. \tag{5.5}$$

The *probit* link function,

$$\Phi^{-1}(\omega_i) = \boldsymbol{x}_{2i}^{\top}\boldsymbol{\beta}_2, \tag{5.6}$$

is also appropriate for the requested purpose. Another possible choice for g_2 is

$$\log[-\log(1-\omega_i)] = \boldsymbol{x}_{2i}^{\top}\boldsymbol{\beta}_2, \tag{5.7}$$

which corresponds to the *complementary log-log* link function. One can notice that these link functions exclude the limit cases $p_i = 0$ and $p_i = \mathrm{P}^{-1}(Y > 0; \mu_i)$, since it is not possible to obtain either $\hat{\omega}_i = 0$ or $\hat{\omega}_i = 1$. The link function (5.7) is usually preferable when the occurrence probability of a specific outcome is considerably high/low as it accommodates asymmetric behaviors on the unit interval, which is not the case for link functions (5.5) and (5.6). Further, a more sophisticated approach considering power and reversal power link functions was proposed by Bazán *et al.* (2017), and can also be used to add even more flexibility when modeling parameter ω_i.

We may refer to the proposed model as a "semi-compatible" regression model. The term "compatible" alludes for "zero-altered", which defines the class of models proposed by Heilbron (1994). Zero-altered models are similar to zero-modified ones, but the compatibility arises from the fact that the linear predictors of μ_i and ω_i are the same. In our case, specifically, it is worthwhile to mention that identifiability problems may occur if one considers a fixed-effects regression model derived directly from (5.2), with parameters μ and p sharing covariates, even if $\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$. Therefore, the adopted structure allows for more flexibility and robustness as μ and ω may share covariates not necessarily with $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1$, and so the only requirement for ensuring model identifiability is the linear independence between covariates within linear predictors.

Unlike traditional approaches, the proposed model can be used for the analysis of zero-inflated and zero-deflated datasets. In this case, given a set of covariates, the probability of a zero-valued count being observed for the i-th subject is given by $1 - g_2^{-1}(\boldsymbol{x}_{2i}^{\top}\boldsymbol{\beta}_2)$. Under the logistic regression model (5.5), β_{2l} $(l = 1, \ldots, q_2)$ represents the direct change in the log-odds of Y_i being positive per 1-unit change in x_{2i}^{l}, holding the other covariates at fixed values. On the other hand, the same not apply if one adopts the link function (5.7) since $e^{\beta_{2l}}$ is not the odds ratio for the l-th covariate effect and so β_{2l} do not have a straightforward interpretation in terms of contribution to log-odds. Likewise, it is not possible to interpret the coefficients of the *probit* model (5.6) directly, but one can evaluate the marginal effect of β_{2l} by analyzing how much the conditional probability of Y_i being positive is affected when the value of x_{2i}^{l} is changed. The exact interpretation of β_{1l} $(l = 1, \ldots, q_1)$ is not direct in terms of the mean of the hurdle model since the positive counts are modeled by a zero-truncated distribution $(\mathcal{ZTPS}_u)$, and therefore,

β_{1l} represents the overall effect of x_{1i}^{l} on the expected value μ_i when $y_i > 0$, holding the other covariates at fixed values.

The proposed model has $d = \dim(\boldsymbol{\beta}) = q_1 + q_2 + 2$ unknown quantities to be estimated. A fully Bayesian approach will be considered for parameter estimation and associated inference. The next section is dedicated to present details of such an approach.

5.3 Inference

In this section, we address the problem of estimating and making inferences about the proposed model under a fully Bayesian perspective. Firstly, we derive the model likelihood function and then, a suitable set of *prior* distributions is considered in order to obtain a computationally tractable *posterior* density for the vector $\boldsymbol{\beta}$. Beyond the primary distributional assumption that $Y_i|\boldsymbol{\beta} \sim \mathcal{ZMPS}_u(\mu_i, \omega_i)$ holds for all i, here we also assume that the outcomes for different subjects are unconditionally independent.

Let Y be a discrete random variable assuming values on $\mathcal{Y}_0$. Suppose that a random experiment is carried out n times independently and, subject to $\boldsymbol{x}_{ki}$ for each i, a vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ of observed values from Y is obtained. Considering model formulation (5.4), the likelihood function of $\boldsymbol{\beta}$ can be written as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta};\boldsymbol{y}) &= \prod_{i=1}^{n} \omega_i \left(\frac{1-\omega_i}{\omega_i}\right)^{\delta_{y_i}} \left[\frac{\mathrm{P}(Y_i = y_i;\mu_i)}{\mathrm{P}(Y_i > 0;\mu_i)}\right]^{1-\delta_{y_i}} \\
&= \prod_{i=1}^{n} g_2^{-1}\left(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2\right) \left[\frac{1-g_2^{-1}\left(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2\right)}{g_2^{-1}\left(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2\right)}\right]^{\delta_{y_i}} \left\{\frac{\mathrm{P}\left[Y_i = y_i; g_1^{-1}\left(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1\right)\right]}{\mathrm{P}\left[Y_i > 0; g_1^{-1}\left(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1\right)\right]}\right\}^{1-\delta_{y_i}},
\end{aligned}$$

and so the corresponding log-likelihood function is given by

$$\begin{aligned}
\ell(\boldsymbol{\beta};\boldsymbol{y}) &= \sum_{i=1}^{n} (1-\delta_{y_i}) \log\left\{\frac{\mathrm{P}\left[Y_i = y_i; g_1^{-1}\left(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1\right)\right]}{\mathrm{P}\left[Y_i > 0; g_1^{-1}\left(\boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1\right)\right]}\right\} + \\
&\quad \sum_{i=1}^{n} \left\{\log\left[g_2^{-1}\left(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2\right)\right] - \delta_{y_i} \log\left[\frac{g_2^{-1}\left(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2\right)}{1-g_2^{-1}\left(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2\right)}\right]\right\} \\
&= \ell_1(\boldsymbol{\beta}_1;\boldsymbol{y}) + \ell_2(\boldsymbol{\beta}_2;\boldsymbol{y}). \qquad (5.8)
\end{aligned}$$

In this work, we will consider a log-linear model for parameter μ_i, that is, $g_1(\mu_i) = \log(\mu_i) = \boldsymbol{x}_{1i}^{\intercal}\boldsymbol{\beta}_1$. The choice of g_2 is left open and the notation $\omega_i = g_2^{-1}(\boldsymbol{x}_{2i}^{\intercal}\boldsymbol{\beta}_2)$ will be used when necessary. From Equation (5.8), one can easily notice that the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal and that ℓ_1 depends only on the positive values of $\boldsymbol{y}$. In this way, the log-likelihood function of $\boldsymbol{\beta}_1$ takes the form

$$\ell_1(\boldsymbol{\beta}_1;\boldsymbol{y}) = \sum_{j\in\mathcal{J}_1} \log\left\{y_j^2 + y_j\left[h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+4\right] + \left[h^2\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right) + 3h\left(e^{\boldsymbol{x}_{1j}^{\intercal}\boldsymbol{\beta}_1}\right)+4\right]\right\} -$$

$$\sum_{j\in\mathcal{J}_1}\log\left[h^4\left(e^{\boldsymbol{x}_{1j}^\top\boldsymbol{\beta}_1}\right)+4h^3\left(e^{\boldsymbol{x}_{1j}^\top\boldsymbol{\beta}_1}\right)+10h^2\left(e^{\boldsymbol{x}_{1j}^\top\boldsymbol{\beta}_1}\right)+7h\left(e^{\boldsymbol{x}_{1j}^\top\boldsymbol{\beta}_1}\right)+2\right]+$$
$$3\sum_{j\in\mathcal{J}_1}\log\left[h\left(e^{\boldsymbol{x}_{1j}^\top\boldsymbol{\beta}_1}\right)\right]-\sum_{j\in\mathcal{J}_1}y_j\log\left[h\left(e^{\boldsymbol{x}_{1j}^\top\boldsymbol{\beta}_1}\right)+1\right], \tag{5.9}$$

where $\mathcal{J}_1=\{j:y_j>0,y_j\in\boldsymbol{y}\}$ is the finite set of indexes regarding the positive observations of $\boldsymbol{y}$. Adopting this setup is equivalent to assuming that each positive element of $\boldsymbol{y}$ comes from a $\mathcal{ZTPS}_u$ distribution. Here, we are extending the fact that estimating the $\mathcal{P}$ parameter θ using the zero-truncated Poisson $(\mathcal{ZTP})$ distribution results in a loss of efficiency in the inference if there is no zero modification (DIETZ; BÖHNING, 2000; CONCEIÇÃO; ANDRADE; LOUZADA, 2014). Now, the log-likelihood function of $\boldsymbol{\beta}_2$ can be written as

$$\ell_2(\boldsymbol{\beta}_2;\boldsymbol{y})=\sum_{i=1}^{n}\log\left[g_2^{-1}\left(\boldsymbol{x}_{2i}^\top\boldsymbol{\beta}_2\right)\right]-\sum_{j\in\mathcal{J}_2}\log\left[\frac{g_2^{-1}\left(\boldsymbol{x}_{2j}^\top\boldsymbol{\beta}_2\right)}{1-g_2^{-1}\left(\boldsymbol{x}_{2j}^\top\boldsymbol{\beta}_2\right)}\right], \tag{5.10}$$

where $\mathcal{J}_2=\{j:y_j=0,y_j\in\boldsymbol{y}\}$ is the finite set of indexes regarding the zero-valued observations of $\boldsymbol{y}$.

5.3.1 Prior distributions

The *g-prior* (ZELLNER, 1986) is a popular choice among Bayesian users of the multiple linear regression model, mainly due to the fact of providing a closed-form *posterior* distribution for the regression coefficients. The *g-prior* is classified as an objective *prior* method which uses the inverse of the Fisher information matrix up to a scalar variance factor $(\tau\in\mathbb{R}_+)$ to obtain the *prior* correlation structure of the multivariate Normal distribution. Such specification is indeed quite attractive since the Fisher information plays a major role in the determination of large-sample covariance in both Bayesian and classical inference.

The problem of eliciting conjugate *priors* for GLMs was addressed by Chen and Ibrahim (2003). Their approach can be considered as a generalization of the original *g-prior* method, but its application is restricted for the class of GLMs since the proposed *prior* does not have closed-form for non-normal exponential families. As an alternative, Gupta and Ibrahim (2009) have proposed the information matrix *prior* as a way to assess the *prior* correlation structure between the coefficients, not including the intercept since the regression matrix is centered as to ensure that β_0 is orthogonal to the other coefficients. This method uses the Fisher information similarly to a precision matrix whose elements are shrunken by the factor τ, which is considered fixed $(\tau\geqslant 1)$. However, the authors have pointed out that such class of *priors* can only be considered Gaussian *priors* if the Fisher information matrix does not depend on the vector $\boldsymbol{\beta}'=(\beta_1,\ldots,\beta_q)$. In this way, Bové and Held (2011) had considered a similar approach when they proposed a class of *hyper-g priors* for GLMs, where the precision matrix is evaluated at the *prior* mode, hence obtaining an information matrix that is $\boldsymbol{\beta}'$ free.

The formal concept behind the information matrix *prior* is closely related to the unit information *prior* (KASS; WASSERMAN, 1995), whose main idea is that the amount of information provided by a *prior* distribution must be the same as the amount of information contained in a single observation. Such an idea can be applied in the previously mentioned approaches by simply considering that τ is equal the total sample size. Gupta and Ibrahim (2009) have also considered fixed values for τ ($\tau \geqslant 1$). On the other hand, some works, including Hansen and Yu (2003), Wang and George (2007) and Bové and Held (2011) do consider *prior* elicitation and inference procedures for the variance scale factor. Here, we will adopt a methodology based on the unit information *prior* idea combined with the "noninformative *g-prior*" proposed by Marin and Robert (2007) for binary regression models. Based on such an approach, it is possible to obtain a quite simple *prior* distribution for the fixed-effects of the proposed model as $\boldsymbol{\beta}_k \sim \mathcal{N}_{\overline{q}_k}(\mathbf{0}, n(\boldsymbol{X}_k^\top \boldsymbol{X}_k)^{-1})$, where $\overline{q}_k = q_k + 1$.

It is worthwhile to mention that, in cases where $\boldsymbol{X}_k$ is rank deficient ($n < q_k + 1$) or contains collinear covariates, it is highly advisable to compute the generalized inverse of $\boldsymbol{X}_k^\top \boldsymbol{X}_k$ otherwise the *prior* covariance matrix of $\boldsymbol{\beta}_k$ may not be defined.

Analogously to Marin and Robert's approach, we do not consider centered regression matrices in the *prior* specification. Hence, we are able to include β_{10} in the proposed *g-prior* but, in this case, the intercept is *a priori* correlated with the other coefficients $(\beta_{11}, \ldots, \beta_{1q_1})$. The same applies for β_{20} and the vector $(\beta_{21}, \ldots, \beta_{2q_2})$.

5.3.2 Posterior distributions and estimation

Considering the outlined structure for the $\mathcal{ZMPS}_u$ regression model, the unnormalized joint *posterior* distribution of the unknown vector $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}; \boldsymbol{y}) \propto \exp\{\ell(\boldsymbol{\beta}; \boldsymbol{y})\}\, \pi(\boldsymbol{\beta}). \tag{5.11}$$

However, since $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are orthogonal, we have that

$$\pi_1(\boldsymbol{\beta}_1; \boldsymbol{y}) \propto \exp\{\ell_1(\boldsymbol{\beta}_1; \boldsymbol{y})\}\, \pi_1(\boldsymbol{\beta}_1) \quad \text{and} \quad \pi_2(\boldsymbol{\beta}_2; \boldsymbol{y}) \propto \exp\{\ell_2(\boldsymbol{\beta}_2; \boldsymbol{y})\}\, \pi_2(\boldsymbol{\beta}_2), \tag{5.12}$$

where ℓ_1 and ℓ_2 are given by (5.9) and (5.10), respectively. It can be proved that, if the data is discrete, then the use of a proper *prior* distribution (multivariate Normal in our case) avoids the *posterior* density to be improper.

From the Bayesian point of view, inferences for the elements of $\boldsymbol{\beta}_k$ can be obtained from their marginal *posterior* distribution. However, deriving analytical expressions for these densities is infeasible, mainly due to the complexity of the associated log-likelihood function. In this case, to make inferences for $\boldsymbol{\beta}_k$, we must resort to a suitable iterative procedure to drawn pseudo-random samples from their *posterior* densities. Hence, aiming to generate N values for $\boldsymbol{\beta}_k$, we will adopt the well-known Random-walk Metropolis (RwM) algorithm (METROPOLIS

et al., 1953; ROBERTS; GELMAN; GILKS, 1997). For the *posterior* densities in (5.12), we consider a multivariate Normal distributions for the proposal (candidate-generating) densities in the algorithm. These distributions will be used as the main terms in the transition *kernels* when computing the acceptance probabilities. Hence, at any state $t > 0$, the MCMC (Markov Chain Monte Carlo) simulation are performed by proposing a candidate $\boldsymbol{\psi}_k$ for $\boldsymbol{\beta}_k$ as

$$\boldsymbol{\psi}_k|\boldsymbol{\beta}_k^{(t-1)} \sim \mathcal{N}_{\bar{q}_k}\left[\nu\boldsymbol{\beta}_k^{(t-1)}, \nu\boldsymbol{\mathcal{S}}_k^{(t-1)}\right],$$

where $\nu = \tau(\tau+1)^{-1} = n(n+1)^{-1}$. One can notice that transitions depend on the acceptance of pseudo-random vectors generated with mean given by the actual state of the chain, which is shrunken by the factor ν. Besides, at any state $t > 0$, the covariance matrix of the candidate vector $\boldsymbol{\psi}_k$ can be approximated numerically by evaluating $\boldsymbol{\mathcal{S}}_k = \boldsymbol{\mathcal{H}}_k^{-1}$ at $\boldsymbol{\beta}_k = \nu\boldsymbol{\beta}_k^{(t-1)}$, where

$$\boldsymbol{\mathcal{H}}_k = -\frac{\partial^2 \log\left[\pi_k\left(\boldsymbol{\beta}_k; \boldsymbol{y}\right)\right]}{\partial\boldsymbol{\beta}_k \partial\boldsymbol{\beta}_k^{\top}}.$$

The procedure to generate pseudo-random samples from the approximate *posterior* distribution of $\boldsymbol{\beta}$ is summarized in Algorithm 7 (see Appendix C, Section C.1). To run it, one has to specify the size of the chains to be generated (N) and the initial state vectors $\boldsymbol{\beta}_1^{(0)}$ and $\boldsymbol{\beta}_2^{(0)}$ beforehand. For a specific asymptotic Gaussian environment, Roberts, Gelman and Gilks (1997) have shown that the optimal acceptance rate should be around 45% for 1-dimensional problems and asymptotically approaches to 23.40% in higher-dimensional problems. Here we are considering acceptance rates varying between 23.40% and 32% as quite reasonable since the proposed model will generally have at least four parameters to be estimated. Indeed, the higher the value of n, the smaller the acceptance rate in the RwM algorithm, which results in smaller variability of the estimates.

The convergence of the simulated sequences can be monitored by using trace and autocorrelation plots, as well as the run-length control method with half-width test (HEIDELBERGER; WELCH, 1983), the Geweke z-score diagnostic (GEWEKE, 1992), and the Brooks-Gelman-Rubin scale-reduction statistic (BROOKS; GELMAN, 1998). After diagnosing convergence, some samples can be discarded as burn-in. The strategy to decrease the correlation between and within generated chains is based on getting thinned steps, and so the final sample is supposed to have size $M \ll N$ for each parameter. A full descriptive summary of the *posterior* distribution (5.11) can be obtained through Monte Carlo (MC) estimators using the sequence $\{\boldsymbol{\beta}^t\}_{t=1}^M$. We choose the *posterior* expected value as the Bayesian point estimator for $\boldsymbol{\theta}$, that is,

$$\hat{\boldsymbol{\beta}} = \frac{1}{M}\sum_{t=1}^{M}\boldsymbol{\beta}^{(t)}, \tag{5.13}$$

which is also known as the minimum mean square error estimator.

In the next section, we discuss the results of the Monte Carlo simulation studies that were performed as a way to assess the performance of the proposed Bayesian methodology. In

Section 5.5, the usefulness, and the competitiveness of the proposed model are illustrated by using a real dataset. All computations were performed using the R environment (R Development Core Team, 2017).

5.3.3 Posterior predictive distribution

In a Bayesian context, the *posterior* predictive distribution (ppd) is defined as the distribution of possible future (unobserved) values conditioned on the observed ones. Under the $\mathcal{ZMPS}_u$ distribution, the pmf of any observation $w \in \mathcal{Y}_0$ (subject to the vectors $\boldsymbol{x}_{1w}^\top$ and $\boldsymbol{x}_{2w}^\top$ of covariates) is given by

$$\begin{aligned}
\mathrm{P}_\pi(Y=w) &= \int_{\mathbb{R}^d} \mathrm{P}^*(Y=w;\mu_w,\omega_w)\,\pi(\boldsymbol{\beta};\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta} \\
&= \int_{\mathbb{R}^{\bar{q}_1}} \left[\frac{\mathrm{P}\left(Y=w;e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1}\right)}{\mathrm{P}\left(Y>0;e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1}\right)}\right]^{1-\delta_w} \pi_1(\boldsymbol{\beta}_1;\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta}_1 \times \\
&\quad \int_{\mathbb{R}^{\bar{q}_2}} g_2^{-1}\left(\boldsymbol{x}_{2w}^\top\boldsymbol{\beta}_2\right)\left[\frac{1-g_2^{-1}\left(\boldsymbol{x}_{2w}^\top\boldsymbol{\beta}_2\right)}{g_2^{-1}\left(\boldsymbol{x}_{2w}^\top\boldsymbol{\beta}_2\right)}\right]^{\delta_w} \pi_2(\boldsymbol{\beta}_2;\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta}_2,
\end{aligned}$$

where $\delta_w = 1$ if $w=0$ and $\delta_w = 0$ otherwise. Noticeably, the ppd has no closed-form available, and therefore, an MC estimator for this quantity is given by

$$\hat{\mathrm{P}}_\pi(Y=w) = \frac{1}{M^2}\sum_{t=1}^{M} g_2^{-1}\left(\boldsymbol{x}_{2w}^\top\boldsymbol{\beta}_2^{(t)}\right)\left[\frac{1-g_2^{-1}\left(\boldsymbol{x}_{2w}^\top\boldsymbol{\beta}_2^{(t)}\right)}{g_2^{-1}\left(\boldsymbol{x}_{2w}^\top\boldsymbol{\beta}_2^{(t)}\right)}\right]^{\delta_w}\sum_{t=1}^{M} b_t(w), \tag{5.14}$$

where

$$b_t(w) = \left\{\frac{h^3\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)}{h^4\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+4h^3\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+10h^2\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+7h\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+2}\times \right.$$
$$\left.\frac{w^2+w\left[h\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+4\right]+\left[h^2\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+3h\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+4\right]}{\left[h\left(e^{\boldsymbol{x}_{1w}^\top\boldsymbol{\beta}_1^{(t)}}\right)+1\right]^w}\right\}^{1-\delta_w}.$$

From Equation (5.14), one can easily estimate, for example, the *posterior* probability of $Y=0$ (subject to $\boldsymbol{x}_{10}^\top$ and $\boldsymbol{x}_{20}^\top$) as

$$\hat{\mathrm{P}}_\pi\left(Y=0;\boldsymbol{x}_{10}^\top,\boldsymbol{x}_{20}^\top\right) = \frac{1}{M}\sum_{t=1}^{M} g_2^{-1}\left(\boldsymbol{x}_{20}^\top\boldsymbol{\beta}_2^{(t)}\right)\left[\frac{1-g_2^{-1}\left(\boldsymbol{x}_{20}^\top\boldsymbol{\beta}_2^{(t)}\right)}{g_2^{-1}\left(\boldsymbol{x}_{20}^\top\boldsymbol{\beta}_2^{(t)}\right)}\right].$$

5.3.4 Influential points

Identifying influential observations is a crucial step in any statistical analysis. Usually, the presence of influential points impacts the inferential procedures and the subsequent conclusions

considerably. In this way, this subsection is dedicated to present some case deletion Bayesian diagnostic measures that can be used to quantify the influence of observations from each subject in a given dataset.

The computation of divergence measures between *posterior* distributions is a useful way to quantify influence. According to Csiszár (1967), the φ-divergence measure between two densities f and g for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is defined by

$$d_{\varphi} = \int_{\boldsymbol{\Theta}} g(\boldsymbol{\theta})\, \varphi\left[\frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\right] \mathrm{d}\boldsymbol{\theta},$$

where φ is a smooth convex, lower semicontinuous function such that $\varphi(1) = 0$. Some popular divergence measures can be obtained by choosing specific functions for φ. The well-known Kullback-Leibler (KL) divergence is obtained by considering $\varphi(z) = -\log(z)$. A symmetric version of the KL divergence, the Jeffrey (J) divergence, can be obtained by specifying $\varphi(z) = (z-1)\log(z)$ and the variational divergence (L^1 norm) is obtained when $\varphi(z) = 0.50|z-1|$. In addition, the Chi-Square (CS) divergence is obtained by considering $\varphi(z) = (z-1)^2$ and the Hellinger (H) distance arises when $\varphi(z) = 0.50(\sqrt{z}-1)^2$. We refer to Peng and Dey (1995) for a detailed study on several types of φ-divergence.

Let $g(\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}; \boldsymbol{y}_i)$ be the joint *posterior* distribution of $\boldsymbol{\beta}$ based only on the i-th observation and let $f(\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}; \boldsymbol{y}_i^j)$, where $\boldsymbol{y}_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ is the response vector without the i-th observation. After some algebra (see Cho *et al.* (2009) for the KL divergence case), one can verify that the φ-divergence corresponds to

$$d_{\varphi} = \mathbb{E}_{\boldsymbol{\beta}}\left\{ \varphi\left[\frac{\mathbb{E}_{\boldsymbol{\beta}}^{-1}\left[\mathrm{P}_*^{-1}(Y_i = y_i; \boldsymbol{\beta}); \boldsymbol{y}\right]}{\mathrm{P}_*(Y_i = y_i; \boldsymbol{\beta})}\right]; \boldsymbol{y}\right\},$$

where $\mathbb{E}_{\boldsymbol{\beta}}^{-1}[\mathrm{P}_*^{-1}(Y_i = y_i; \boldsymbol{\beta}); \boldsymbol{y}]$ is the conditional predictive ordinate (CPO) statistic (GEISSER; EDDY, 1979) for the i-th observation. Here, we are also not able to compute the inner expectation over $\boldsymbol{\beta}$ analytically and so, an MC estimator for the CPO_i is given by

$$\widehat{\mathrm{CPO}}_i = \left[\frac{1}{M}\sum_{t=1}^{M} \mathrm{P}_*^{-1}\left(Y_i = y_i; \boldsymbol{\beta}^{(t)}\right)\right]^{-1}. \tag{5.15}$$

According to Congdon (2005), the harmonic mean estimator (5.15) is stable when most of the individual log-likelihood values exceed -10. Using the estimated CPO, one can approximate the local influence of a particular y_i on the joint *posterior* distribution (5.11) as

$$\hat{d}_{\varphi} = \frac{1}{M}\sum_{t=1}^{M} \varphi\left[\frac{\widehat{\mathrm{CPO}}_i}{\mathrm{P}_*\left(Y_i = y_i; \boldsymbol{\beta}^{(t)}\right)}\right].$$

One can notice that, if $\pi(\boldsymbol{\beta}; \boldsymbol{y}_{-i}) = \pi(\boldsymbol{\beta}; \boldsymbol{y})$, then there is no divergence caused by observation y_i. In practice, however, it may not be elementary to define a threshold value for the

divergence in order to decide about the magnitude of the influence (WEISS, 1996). A measure of calibration for the KL divergence was proposed by McCulloch (1989). The idea is based on the typical toy binary example of tossing a coin once and observing its upper face. This experiment can be described by $\mathrm{P}(Y=y;\rho)=\rho^{y}(1-\rho)^{1-y}$, $y\in\{0,1\}$, where $\rho\in[0,1]$ is the probability of success. Regardless of what success means, if the coin is unbiased, then $\mathrm{P}(Y=y;\rho)=0.50$. Thus, the φ-divergence between a (possibly) biased and an unbiased coin is given by

$$d_{\varphi}(\rho)=\frac{\varphi(2\rho)+\varphi[2(1-\rho)]}{2},$$

from which one can conclude that the divergence between two *posteriors* distributions can be associated with the biasedness of a coin (PENG; DEY, 1995). By analogy, this implies that predict unobserved responses using $\pi(\boldsymbol{\beta};\boldsymbol{y}_{-i})$ instead of $\pi(\boldsymbol{\beta};\boldsymbol{y})$ is equivalent to describe an unobserved event as having probability ρ_i, when the correct probability is 0.50. Considering some specific choices for φ, in Table 90 (see Appendix C, Section C.2) we present MC estimators that can be used to compute the local influence of each y_i. Besides, we also present the expression of $d_{\varphi}(\rho)$ for each φ. For ease of notation, we assume $f_i^t=\mathrm{P}(Y_i=y_i;\boldsymbol{\beta}^{(t)})$.

The function $d_{\varphi}(\rho)$ is symmetric about 0.50 and increases as ρ moves away from 0.50. In addition, $\inf_{\rho\in(0,1)}d_{\varphi}(\rho)=0$, which is attained at $\rho=0.50$ since $d_{\varphi}(0.50)=\varphi(1)=0$. Therefore, a general measure of calibration based on the φ-divergence can be obtained by solving

$$2d_{\varphi}(\rho)-\varphi(2\rho)-\varphi[2(1-\rho)]=0.$$

An estimator for the calibration measure (ρ_{φ}) associated with each φ-divergence type is also presented in Table 90. Clearly, depending on the form of φ, such an equation may not have a closed-form, which is the case of the J divergence. Besides, one can notice that $\rho_i\in[0.50,1]$ and so, for $\rho_i\gg 0.50$, the i-th observation may be considered an influential point. For example, if $\rho_i>0.80$ is considered a significative bias, then y_i will be classified as influential if $\hat{d}_i>0.223$ $(d_{\varphi}(0.80)\approx 0.223)$ under the KL divergence or yet if $\hat{d}_i>0.051$ $(d_{\varphi}(0.80)\approx 0.051)$ under the H divergence.

5.3.5 Model comparison and adequacy

There are several methods for Bayesian model selection that are useful to compare competing models. The most popular method is the Deviance Information Criterion (DIC), which was proposed to work simultaneously as a measure of fit and complexity of the model. The DIC criterion is defined as

$$\mathrm{DIC}=\mathbb{E}_{\boldsymbol{\beta}}[\mathrm{D}(\boldsymbol{\beta})]+\rho_{\mathrm{D}}=\underline{\mathrm{D}}(\boldsymbol{\beta})+\rho_{\mathrm{D}},$$

where $\mathrm{D}(\boldsymbol{\beta})=-2\ell(\boldsymbol{\beta};\boldsymbol{y})$ is the *deviance* function and $\rho_{\mathrm{D}}=\underline{\mathrm{D}}(\boldsymbol{\beta})-\mathrm{D}(\hat{\boldsymbol{\beta}})$ is the effective number of model parameters, with $\hat{\boldsymbol{\beta}}$ given by (5.13). A negative value for ρ_{D} may suggest that the

log-likelihood function is non-concave, the *prior* distribution is misspecified, or the *posterior* expected value is not a good estimator for $\boldsymbol{\beta}$. On the other hand, when $\rho_{\mathrm{D}} \gg d$, then there is an indication of overfitting with estimate $\hat{\boldsymbol{\beta}}$.

Noticeably, we are not able to compute the expectation of $\mathrm{D}(\boldsymbol{\beta})$ over $\boldsymbol{\beta}$ analytically. In this case, an MC estimator for such a measure is given by

$$\underline{\hat{\mathrm{D}}}(\boldsymbol{\beta}) = -\frac{2}{M}\sum_{t=1}^{M} \ell\left(\boldsymbol{\beta}^{(t)}; \boldsymbol{y}\right),$$

and so the DIC can be estimated by $\widehat{\mathrm{DIC}} = 2\underline{\hat{\mathrm{D}}}(\boldsymbol{\beta}) - \mathrm{D}(\hat{\boldsymbol{\beta}})$.

The Expected Akaike (EAIC) and the Expected Bayesian (EBIC) information criteria can also be used when comparing Bayesian models (CARLIN; LOUIS, 2001; BROOKS, 2002). Using the approximation for the expected value of $\mathrm{D}(\boldsymbol{\beta})$, these measures can be estimated by

$$\widehat{\mathrm{EAIC}} = \underline{\hat{\mathrm{D}}}(\boldsymbol{\beta}) + 2d \quad \text{and} \quad \widehat{\mathrm{EBIC}} = \underline{\hat{\mathrm{D}}}(\boldsymbol{\beta}) + d\log(n).$$

Another widely used criterion is derived from the CPO statistic, which is based on the cross-validation criterion to compare models. For the i-th observation, the CPO can be estimated through Equation (5.15). A summary statistic of the estimated CPO's is the log-marginal pseudo-likelihood (LMPL) given by the sum of the logarithms of $\widehat{\mathrm{CPO}}_i$'s. Regarding model comparison, we have that the lower the values of DIC, EAIC, EBIC, and NLMPL (negative LMPL), the better the fit.

In addition to comparing, researchers are often interested in verifying the adequacy of the fitted models. An effective way to evaluate model adequacy is based on the use of measures derived from the ppd. For instance, if any observation is extremely unlikely relative to the ppd, the adequacy of the obtained fit might be questionable. A widespread discrepancy measure between model and data was proposed by Gelman *et al.* (2004). In our case, we need a slightly adapted version of such a measure, which is given by

$$\mathrm{T}(\boldsymbol{y}, \boldsymbol{\beta}) = -2\sum_{i=1}^{n} \log\left[\mathrm{P}_*(Y_i = y_i; \boldsymbol{\beta})\right].$$

The Bayesian p-value (*posterior* predictive p-value), proposed by Rubin (1984), is defined as

$$p_{\mathrm{B}} = \mathrm{P}\left[\mathrm{T}(\boldsymbol{y}_r, \boldsymbol{\beta}) \geqslant \mathrm{T}(\boldsymbol{y}, \boldsymbol{\beta}); \boldsymbol{y}\right],$$

where $\boldsymbol{y}_r$ denotes the response vector that might have been observed if the conditions generating $\boldsymbol{y}$ were reproduced. This predictive measure can be empirically estimated as the relative number of times that $\mathrm{T}(\boldsymbol{y}_r, \hat{\boldsymbol{\beta}})$ exceeds $\mathrm{T}(\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ out of B simulations. In general, the model fit becomes suspect if the discrepancy is of practical relevance, and the associated Bayesian p-value is close either to 0 or 1 (GELMAN *et al.*, 2004). A large (small) value of p_{B}, say greater than 0.95 (lower than 0.05), indicates model misspecification (lack-of-fit), that is, the observed behavior would be unlikely to be seen if we replicate the response vector using the fitted model.

5.3.6 Residual analysis

The residual analysis plays an essential role in the task of validating the results obtained from a regression model. In general, residual metrics are responsible for indicating departures from the underlying model assumptions by quantifying the portion of data variability that was not explained by the fitted model. Assessing the adequacy of a regression model using residual metrics is a common practice nowadays due to the availability of statistical packages providing diagnostic tools for well-established models. However, deriving appropriate residuals is not always an easy task for non-normal models that accommodate overdispersion and mixed-effects. In this way, we will consider here a popular residual metric proposed by Dunn and Smyth (1996), the randomized quantile residuals (RQRs), which can be straightforwardly used in our context to assess the appropriateness of the proposed model when fitted to real data.

For obvious reasons, we focus on the definition of RQRs for discrete random variables. In this case, the RQR associated to the i-th observation is defined as $r_i = \Phi^{-1}(u_i)$, where Φ denotes the cdf of the standard Normal distribution and u_i is a Uniform random variable defined on $(a_i, b_i]$, with $a_i = \lim_{y \uparrow y_i} \mathrm{F}(y_i)$ and $b_i = \mathrm{F}(y_i)$, where $\mathrm{F}(y_i)$ is the cdf of the current model. In our case, we may obtain an MC estimator for the RQR as $\hat{r}_i = \Phi^{-1}(u_i)$, with $u_i \sim \mathcal{U}(\lim_{y \uparrow y_i} \hat{\mathrm{F}}_*(y_i), \hat{\mathrm{F}}_*(y_i)]$. Here, $\hat{\mathrm{F}}_*(y_i) \equiv \mathrm{F}_*(y_i; \hat{\mu}_i, \hat{\omega}_i)$ is an estimate for the probability of $Y_i \leqslant y_i$ using cdf (5.3), where $\hat{\mu}_i$ and $\hat{\omega}_i$ depend on the fitted model as $\hat{\mu}_i = \log(\boldsymbol{x}_{1i}^\top \hat{\boldsymbol{\beta}}_1)$ and $\hat{\omega}_i = g_2^{-1}(\boldsymbol{x}_{2i}^\top \hat{\boldsymbol{\beta}}_2)$.

The main assumption for this metric is that $\hat{r}_i \sim \mathcal{N}(0,1)$ must hold, whichever the variability degree of $\hat{\mu}_i$ and $\hat{\omega}_i$. In this case, after model fitting, one has to evaluate if these residuals are normally distributed around zero, which can be made through adherence tests and by using graphical techniques as histograms and Half-Normal probability plots. An excellent alternative for checking whether RQRs are consistent with the fitted model is the inclusion of simulated envelopes in their Half-Normal plot. Thus, if a significative subset of estimated residuals falls outside the envelope bands, then the adequacy of the fitted model must be questioned, and further investigation on the corresponding observations are necessary. An algorithm for obtaining simulated envelopes for a Half-Normal plot is provided by Moral, Hinde and Demétrio (2017).

5.4 Simulation study

The empirical properties of an estimator can be accessed through Monte Carlo simulations. In this way, we have performed an intensive simulation study aiming to validade the Bayesian approach in some specific situations. The simulation process was carry out by generating 500 pseudo-random samples of sizes $n = 50, 100, 200,$ and 500 of a variable Y following a $\mathcal{ZMPS}_u$ distribution under the regression framework presented in Section 5.2. For the whole process, it was considered a $n \times 2$ regression matrix $\boldsymbol{X}_1 = (\mathbf{1}_n, \boldsymbol{X}_{1,n\times 1})$ in which $\boldsymbol{X}_{n\times 1}$ is a vector containing n generated values from a Uniform distribution on the unit interval. Here, we have

fixed $\boldsymbol{X}_2 = \boldsymbol{X}_1$. Moreover, we have assigned different values for the vectors $\boldsymbol{\beta}_1^\top = (\beta_{10}, \beta_{11})$ and $\boldsymbol{\beta}_2^\top = (\beta_{20}, \beta_{21})$ in order to generate both zero-inflated and zero-deflated artificial samples. The logarithm link function was considered for g_1. For g_2, we have considered the link functions (5.5)-(5.7) as a way to evaluate how these different specifications affect the estimation of $\boldsymbol{\beta}$. Due to the massive amount of obtained results, here we will only present some of the outputs using the *probit* link function. The remaining results were made available in Appendix C (Section C.3).

Table 31 – Actual parameter values for simulation of zero-modified artificial datasets.

Case	Scenario	Link	β_{10}	β_{11}	β_{20}	β_{21}	Range μ_i	Range p_i
Inflation	1	Logit	1.50	3.00	−1.00	−1.00	(4.48;90.02)	(0.12;0.30)
		Probit						(0.02;0.18)
		CLL						(0.13;0.34)
	2	Logit	1.50	3.00	−1.00	0.50	(4.48;90.02)	(0.30;0.38)
		Probit						(0.18;0.31)
		CLL						(0.34;0.45)
	3	Logit	1.50	−1.50	−1.00	−1.00	(1.00;4.48)	(0.23;0.30)
		Probit						(0.04;0.18)
		CLL						(0.24;0.34)
	4	Logit	1.50	−1.50	−1.00	0.50	(1.00;4.48)	(0.30;0.73)
		Probit						(0.18;0.59)
		CLL						(0.34;0.88)
Deflation	1	Logit	−1.00	1.00	0.50	0.50	(0.37;1.00)	(1.41;2.30)
		Probit						(1.62;2.56)
		CLL						(1.80;2.99)
	2	Logit	−1.00	1.00	1.50	−1.00	(0.37;1.00)	(1.20;3.02)
		Probit						(1.33;3.45)
		CLL						(1.56;3.66)
	3	Logit	−1.00	−1.50	0.50	0.50	(0.08;0.37)	(2.30;9.64)
		Probit						(2.56;11.09)
		CLL						(2.99;12.31)
	4	Logit	−1.00	−1.50	1.50	−1.00	(0.08;0.37)	(3.02;8.21)
		Probit						(3.45;9.12)
		CLL						(3.66;10.65)

CLL: Complementary log-log.

Source: Elaborated by the author.

Algorithm 8 (see Appendix C, Section C.1) can be used to generate a single pseudo-random realization from the $\mathcal{ZMPS}_u$ distribution in the regression framework with covariate $\mathcal{U}(0,1)$ for μ and ω. The extension for the use of more covariates is straightforward. The process to generate a pseudo-random sample of size n consists of running the algorithm as often as necessary, say n^* times ($n^* \geqslant n$). The sequential-search is a black-box algorithm and works with any computable probability vector. The main advantage of such a procedure is its simplicity.

On the other hand, sequential-search algorithms may be slow as the *while-loop* may have to be repeated very often. More details about this algorithm can be found at Hörmann, Leydold and Derflinger (2013).

Under the $\mathcal{ZMPS}_u$ distribution, the expected number of iterations (NI), that is, the expected number of comparisons in the *while* condition, is given by

$$\mathbb{E}(\mathrm{NI}) = \lambda + 1 = \frac{\omega\mu\left[h^2(\mu)+h(\mu)+2\right]\left[h(\mu)+1\right]^3}{h^4(\mu)+4h^3(\mu)+10h^2(\mu)+7h(\mu)+2} + 1,$$

where $h(\mu)$ is given by Equation (5.1).

We have considered four scenarios for each kind of zero modification. Table 31 presents the true parameter values that were considered in our study. For the zero-inflated (zero-deflated) case, the samples were generated from the $\mathcal{ZMPS}_u$ distribution by considering that $p_i \in (0,1)$ ($p_i \in [1, \mathrm{P}^{-1}(Y > 0; \mu_i)]$) for all i. Here, the regression coefficients were chosen by taking into account that zero-inflated (zero-deflated) samples have, naturally, proportion of zeros greater (lower) than expected under an ordinary count distribution and therefore, the variable Y_i $(i = 1, \ldots, n)$ was generated with mean far from zero (close to zero). Table 31 also presents the range of parameters μ_i and p_i in each scenario. The bounds were obtained by evaluating the linear predictors $\beta_{10} + \beta_{11}x$ and $\beta_{20} + \beta_{21}x$ at $x = 0$ and $x = 1$ (limit values of the adopted covariate). Scenarios 1 and 2 of the zero-inflated case were considered to illustrate the behavior of the Bayesian estimators when (right) long-tailed count data are available.

To apply the proposed Bayesian approach to each scenario, we have considered the RwM algorithm for MCMC sampling. A total of $N = 50,000$ values were generated for each parameter, considering a burn-in period of 20% of the size of the chain. To obtain pseudo-independent samples from the *posterior* distributions given in (5.12), one out every 10 generated values were kept, resulting in chains of size $M = 4,000$ for each parameter. Using trace plots and Geweke's z-score diagnostic, the stationarity of the remaining chains was revealed. When running the simulations, the acceptance rates were ranging between 23.40 and 32%. The *posterior* mean (5.13) was considered as the Bayesian point estimator, and its performance was studied by assessing its bias (B), its mean squared error (MSE) and its mean absolute percentage error (MAPE). Besides, the coverage probability (CP) of the 95% highest *posterior* density intervals (HPDIs) was also estimated.

Using the generated samples and letting $\gamma = \beta_{10}, \beta_{11}, \beta_{20}$ or β_{21}, the MC estimators for these measures are given by

$$\widehat{\mathrm{B}}_{\hat{\gamma}} = \frac{1}{500}\sum_{j=1}^{500}\left(\hat{\gamma}_j - \gamma\right), \quad \widehat{\mathrm{MSE}}_{\hat{\gamma}} = \frac{1}{500}\sum_{j=1}^{500}\left(\hat{\gamma}_j - \gamma\right)^2, \quad \text{and} \quad \widehat{\mathrm{MAPE}}_{\hat{\gamma}} = \frac{1}{500}\sum_{j=1}^{500}\left|\frac{\hat{\gamma}_j - \gamma}{\gamma}\right|.$$

Table 32 – Empirical properties of the Bayesian estimators using zero-inflated samples.

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 1		
50	β_{10}	−0.013	0.992	1.000	49.716
	β_{11}	−1.969	9.844	1.284	81.167
	β_{20}	0.020	0.214	1.001	36.875
	β_{21}	0.191	0.754	1.025	69.121
100	β_{10}	−0.007	0.415	1.000	32.842
	β_{11}	−0.847	3.051	1.144	42.982
	β_{20}	0.003	0.148	1.000	30.516
	β_{21}	0.063	0.564	1.004	59.161
200	β_{10}	−0.010	0.134	1.000	18.566
	β_{11}	−0.405	1.003	1.093	24.666
	β_{20}	0.009	0.054	1.001	18.170
	β_{21}	0.010	0.266	1.000	40.134
500	β_{10}	−0.005	0.052	1.000	12.110
	β_{11}	−0.138	0.236	1.043	12.837
	β_{20}	0.001	0.025	1.000	12.704
	β_{21}	0.004	0.096	1.000	24.369
			Scenario 2		
50	β_{10}	−0.213	0.299	1.086	27.133
	β_{11}	−0.117	0.586	1.012	19.488
	β_{20}	0.052	0.180	1.008	33.672
	β_{21}	−0.026	0.497	1.001	108.882
100	β_{10}	−0.131	0.143	1.066	19.972
	β_{11}	−0.026	0.292	1.001	14.063
	β_{20}	0.003	0.108	1.000	26.370
	β_{21}	−0.002	0.311	1.000	89.775
200	β_{10}	−0.056	0.055	1.030	12.321
	β_{11}	−0.018	0.121	1.001	9.102
	β_{20}	−0.002	0.035	1.000	15.016
	β_{21}	0.006	0.114	1.000	54.306
500	β_{10}	−0.030	0.024	1.020	8.184
	β_{11}	0.004	0.052	1.000	6.078
	β_{20}	0.003	0.016	1.000	10.023
	β_{21}	−0.004	0.043	1.000	32.685

Source: Elaborated by the author.

Table 33 – Empirical properties of the Bayesian estimators using zero-deflated samples.

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 1					
50	β_{10}	0.081	0.324	1.010	45.581
	β_{11}	−0.164	0.839	1.016	74.955
	β_{20}	−0.001	0.156	1.000	62.082
	β_{21}	−0.050	0.471	1.003	109.956
100	β_{10}	0.053	0.197	1.007	35.633
	β_{11}	−0.075	0.485	1.006	55.800
	β_{20}	−0.034	0.092	1.006	47.660
	β_{21}	0.030	0.301	1.002	86.294
200	β_{10}	0.012	0.078	1.001	22.714
	β_{11}	−0.017	0.231	1.001	38.300
	β_{20}	−0.022	0.035	1.007	29.615
	β_{21}	0.020	0.123	1.002	55.809
500	β_{10}	0.016	0.040	1.003	15.990
	β_{11}	−0.028	0.098	1.004	25.124
	β_{20}	0.006	0.016	1.001	19.987
	β_{21}	−0.011	0.047	1.001	35.119
Scenario 2					
50	β_{10}	0.050	0.266	1.005	41.381
	β_{11}	−0.084	0.744	1.005	67.351
	β_{20}	−0.072	0.205	1.013	24.018
	β_{21}	0.057	0.512	1.003	56.514
100	β_{10}	0.012	0.171	1.000	32.812
	β_{11}	−0.022	0.445	1.001	52.762
	β_{20}	−0.014	0.138	1.001	19.169
	β_{21}	0.002	0.380	1.000	48.545
200	β_{10}	0.010	0.072	1.001	21.244
	β_{11}	−0.020	0.214	1.001	37.192
	β_{20}	−0.031	0.043	1.012	11.013
	β_{21}	0.042	0.136	1.007	29.241
500	β_{10}	0.019	0.028	1.007	13.449
	β_{11}	−0.028	0.074	1.005	21.794
	β_{20}	0.002	0.026	1.000	8.697
	β_{21}	−0.004	0.061	1.000	19.548

Source: Elaborated by the author.

Table 34 – Coverage probabilities (%) of the HPDIs using zero-inflated samples.

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	2.60	94.60	2.80	1.20	93.20	5.60
	β_{11}	0.80	73.80	25.40	1.60	93.60	4.80
	β_{20}	2.40	97.40	0.20	5.20	93.80	1.00
	β_{21}	2.40	97.60	0.00	1.60	95.20	3.20
100	β_{10}	2.00	96.20	1.80	1.20	95.40	3.40
	β_{11}	0.40	87.00	12.60	1.20	95.20	3.60
	β_{20}	4.20	94.60	1.20	3.40	95.00	1.60
	β_{21}	3.00	95.80	1.20	3.00	93.80	3.20
200	β_{10}	4.20	92.40	3.40	1.00	95.20	3.80
	β_{11}	0.60	90.00	9.40	1.60	95.60	2.80
	β_{20}	3.20	94.80	2.00	2.00	96.40	1.60
	β_{21}	3.40	95.40	1.20	1.80	96.60	1.60
500	β_{10}	3.40	95.20	1.40	2.80	93.60	3.60
	β_{11}	1.00	92.60	6.40	1.60	95.00	3.40
	β_{20}	2.80	95.20	2.00	2.20	96.40	1.40
	β_{21}	2.60	95.40	2.00	2.20	95.60	2.20

Source: Elaborated by the author.

Table 35 – Coverage probabilities (%) of the HPDIs using zero-deflated samples.

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	4.60	94.80	0.60	3.40	96.20	0.40
	β_{11}	0.80	96.80	2.40	1.60	95.80	2.60
	β_{20}	1.60	95.80	2.60	0.00	97.00	3.00
	β_{21}	1.40	96.00	2.60	1.60	97.80	0.60
100	β_{10}	3.60	95.20	1.20	2.60	96.40	1.00
	β_{11}	0.80	96.00	3.20	1.60	96.60	1.80
	β_{20}	1.60	95.40	3.00	1.60	94.20	4.20
	β_{21}	2.80	94.40	2.80	3.40	94.60	2.00
200	β_{10}	2.60	96.20	1.20	2.80	95.40	1.80
	β_{11}	2.20	96.00	1.80	1.60	96.20	2.20
	β_{20}	1.80	94.80	3.40	0.80	95.60	3.60
	β_{21}	1.80	95.40	2.80	3.40	95.40	1.20
500	β_{10}	4.60	93.60	1.80	4.00	94.80	1.20
	β_{11}	2.00	94.40	3.60	1.00	96.60	2.40
	β_{20}	1.80	95.40	2.80	3.00	94.00	3.00
	β_{21}	3.20	94.60	2.20	2.60	94.20	3.20

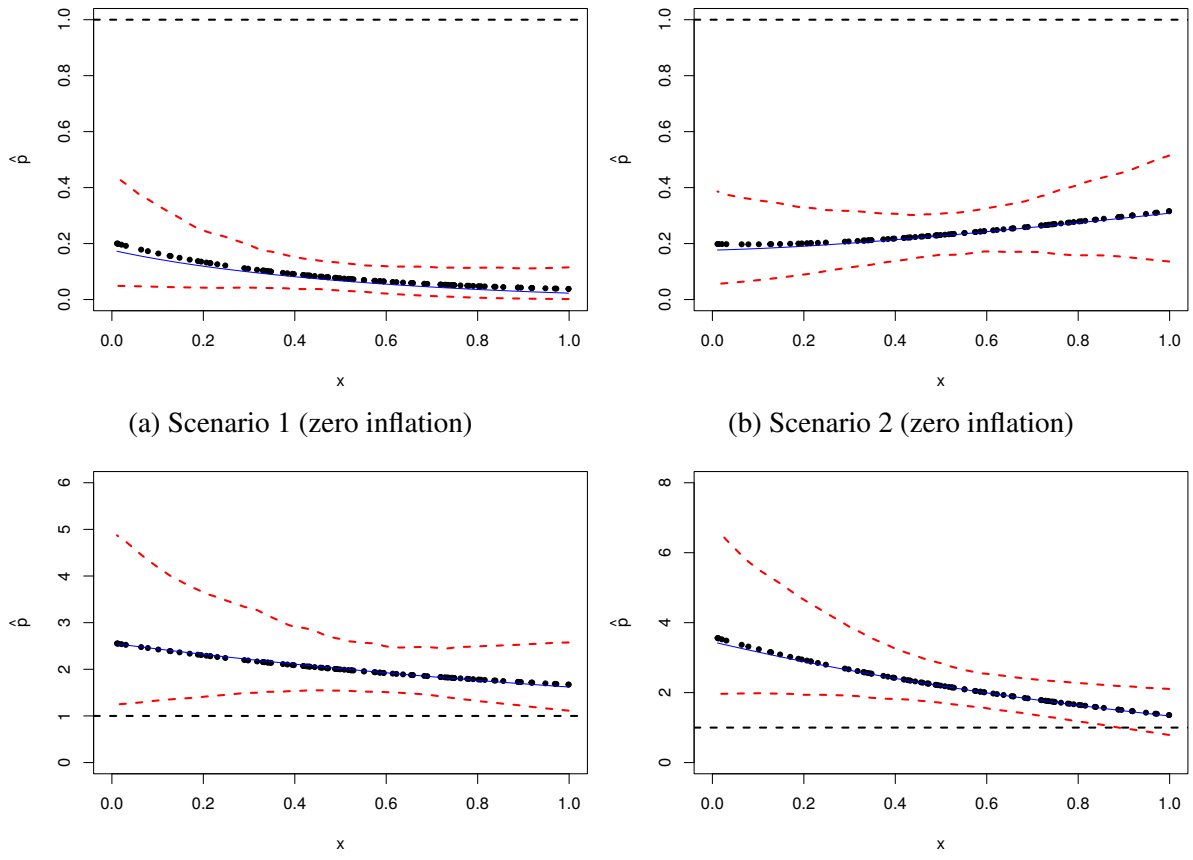
Source: Elaborated by the author.

The variance of $\hat{\gamma}$ was estimated as the difference between the MSE and the square of the bias. Moreover, the CP of the HPDIs was estimated by

$$\widehat{\mathrm{CP}}_{\gamma} = \frac{1}{500} \sum_{j=1}^{500} \delta_j(\gamma),$$

where $\delta_j(\gamma)$ assumes 1 if the j-th HPDI contains the true value γ and 0 otherwise. Also, we have estimated the below noncoverage probability (BNCP) and the above noncoverage probability (ANCP) of the HPDIs. These measures are computed analogously to CP. The BNCP and ANCP may be useful measures to determine asymmetrical behaviors as they provide the probabilities of finding the actual value of γ on the tails of its *posterior* distribution.

Figure 19 – *Posterior* estimates for parameter p using zero-modified samples ($n = 100$).



(a) Scenario 1 (zero inflation)

(b) Scenario 2 (zero inflation)

(c) Scenario 1 (zero deflation)

(d) Scenario 2 (zero deflation)

Source: Elaborated by the author.

The computed measures are presented in Tables 32 and 33 (*posterior* estimates) and Tables 34 and 35 (coverage probabilities). We have noticed in the simulation study that, as expected, the parameter estimates became more accurate with increasing sample sizes since the estimated biases and mean squared errors have decreased considerably as n increased. Also, the squared ratio between the mean squared error and the estimated variance approaches 1 as n increases. Although high MAPE values were obtained for some parameters (when using small

sample sizes), this does not compromise the overall accuracy in estimation. For example, on Table 33 (Scenario 1), for $n = 100$, we have obtained a estimated MAPE value of approximately 56% for β_{11}. Taking into account that the true value of this parameter is 1.00, we have that the estimates for β_{11} were ranging mostly between 0.44 and 1.56, which do not represent such a significant impact on the estimated mean (μ). When (right) long-tailed count data are available, the CP of the HPDI for β_{11} is considerably lower than the adopted nominal level (for small sample sizes) as its *posterior* distribution tends to be more asymmetric towards higher values on the parametric space. However, we have observed that the estimated CP of the HPDIs is converging to 95% in both zero-modified cases, and the *posterior* distributions became more symmetric with increasing sample sizes.

Figure 19 depicts the Bayesian estimates obtained for parameter p using zero-inflated and zero-deflated artificial samples with $n = 100$. The true values are represented by the straight blue lines, and the 95% HPDIs are represented by the red dashed lines. The filled black dots represent the estimated values for each generated observation. The uncertainty about parameter p is higher under zero-deflated data when the value of x is close to zero.

Considering the predefined scenarios, we conclude that our simulation study provides favorable indications about the suitability of the adopted Bayesian methodology to estimate the parameters of the proposed model. We believe that in a similar procedure with a different set of actual values, the overall behavior of the estimators should resemble the results that we have described here. Besides, the adopted methodology would also be reliable if one or more than one covariates (possibly of other nature) were included in the linear predictors of μ_i and ω_i.

5.5 Chromosomal aberration data analysis

In this section, the $\mathcal{ZMPS}_u$ regression model is considered for the analysis of a real dataset obtained from a cytogenetic dosimetry experiment that was first presented by Heimers *et al.* (2006). In this study, the response variable is the number of cytogenetic chromosomal aberrations after the DNA molecule is treated with induced radiation. The dataset was obtained by irradiating five blood samples from healthy donors with different doses x_i $(i = 1,\ldots,5)$ ranging between 0.1 and 1.0 Gy (gray) with 2.1 MeV (million electron volts) neutrons in a culture time of 72h, considering partial-body exposure-densely ionizing radiation. In the following, n_i cells were examined in each irradiated sample and the number of dicentrics and centric rings aberrations y_{ij} $(j = 1,\ldots,n_i)$ was recorded.

The frequency distribution of the collected data is available in Table 36, along with some descriptive statistics. From the observed dataset, there exist evidences that the response variable is slightly overdispersed since $\bar{y}_{\cdot} = 0.131 < s^2_{\cdot} = 0.210$ and $s^2_{\cdot}/\bar{y}_{\cdot} = 1.607$. Also, the number of aberrations appears to be heavily zero-inflated, as can be seen in the left-panel of Figure 20. On the other hand, one can notice that, as the dose of ionizing radiation increases, the number of

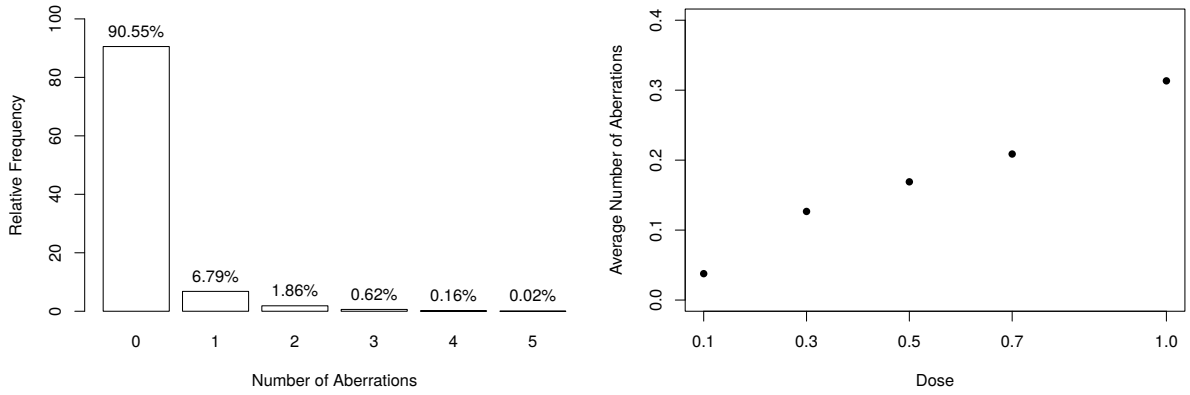
observed zeros decreases, but the distribution becomes more overdispersed since it naturally increases the number of aberrations.

Table 36 – Descriptive summary of the number of dicentrics and centric rings aberrations.

x_i	y_{ij}						n_i	$\overline{y}_i$	s_i	$s_i^2/\overline{y}_i$
	0	1	2	3	4	5				
0.1	2130	59	9	2	0	0	2200	0.038	0.224	1.316
0.3	1088	84	19	6	3	0	1200	0.127	0.449	1.591
0.5	875	88	30	7	0	0	1000	0.169	0.493	1.438
0.7	679	88	23	8	1	1	800	0.209	0.568	1.545
1.0	480	75	27	13	5	0	600	0.313	0.732	1.712

Source: Elaborated by the author.

Figure 20 – Summary of the number of dicentrics and centric rings aberrations.



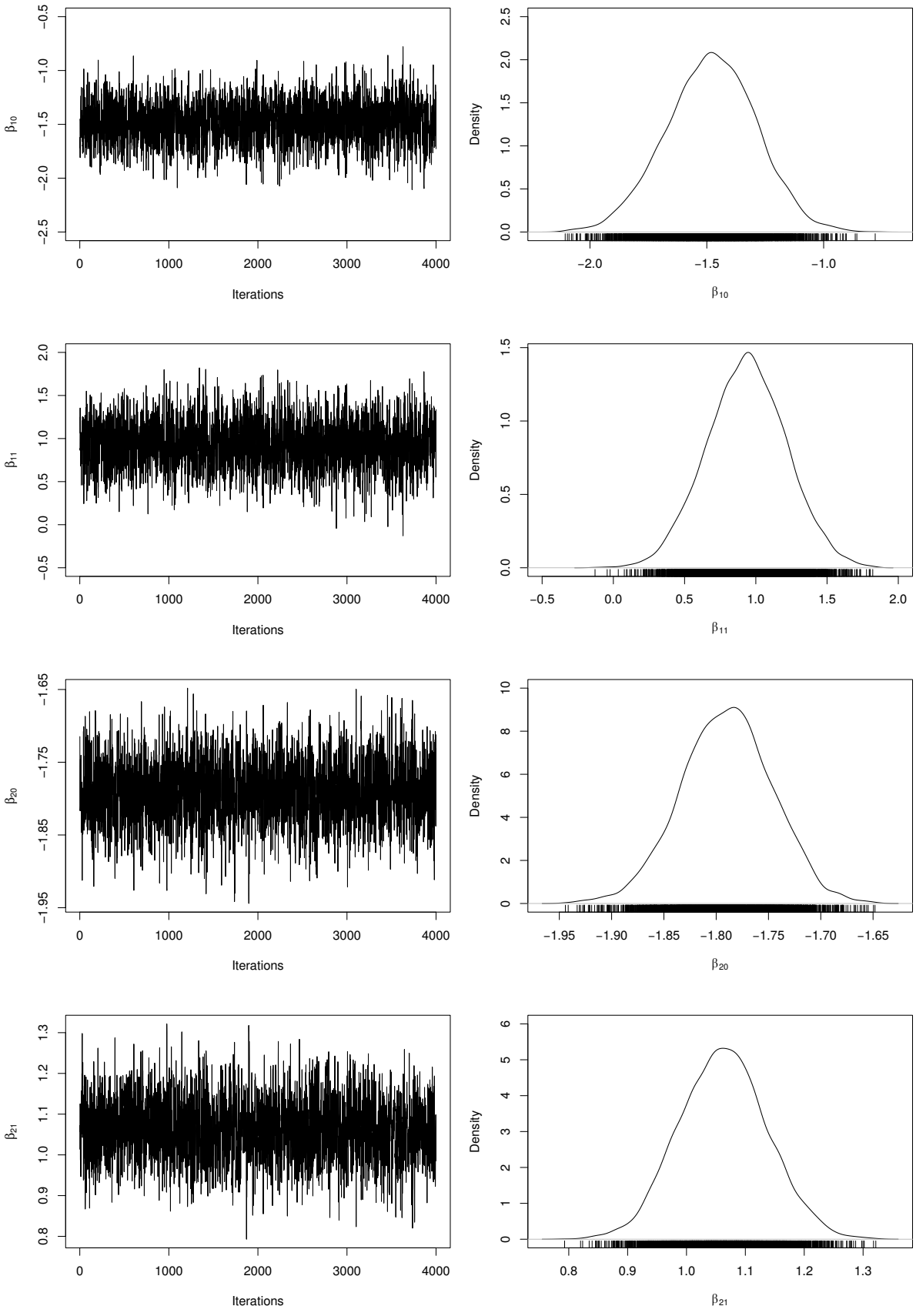
Source: Elaborated by the author.

According to Oliveira *et al.* (2016), when considering higher linear energy transfer radiations, the incidence of chromosomal aberrations becomes a linear function of the dose because the more densely ionizing nature of the radiation leads to an "one-track" distribution of damage. Such an aspect can be seen in the right-panel of Figure 20, which highlights the linear behavior between the average number of aberrations and the doses. In this way, our assumption is that $Y_{ij}|x_i \sim \mathcal{ZMPS}_u(\mu_{ij}, \omega_{ij})$, where parameters μ_{ij} and ω_{ij} are specified as linear dose models, that is,

$$\log\left(\mu_{ij}\right) = \beta_{10} + \beta_{11}x_i \quad \text{and} \quad g_2\left(\omega_{ij}\right) = \beta_{20} + \beta_{21}x_i.$$

To fit the $\mathcal{ZMPS}_u$ regression model with dose as the only covariate, we have adopted the same procedure used in the previous section. The link function (5.6) was chosen to relate ω_{ij} with the linear predictor $\beta_{20} + \beta_{21}x_i$ and so we have the *probit* hurdle regression model. In this

Figure 21 – Trace plots and marginal *posterior* distributions of the $\mathcal{ZMPS}_u$ model parameters.



Source: Elaborated by the author.

framework, the coefficient β_{11} represents the effect of the dose of ionizing radiation on the expected count μ_i when $Y_{ij} > 0$, and β_{21} indicates the effect of the dose on the probability of aberrations to occur. We have considered the RwM for MCMC sampling, generating a chain of size $N = 50,000$ for each parameter whereby the first $10,000$ values were discarded as burn-in. The stationarity of the chains was revealed using the Geweke z-score diagnostic of convergence. To obtain the pseudo-independent samples from the *posterior* distributions given in (5.12), we have considered one value out of every 10 generated ones, resulting in chains of size $M = 4,000$ for each parameter.

Table 37 presents the *posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPS}_u$ fitted model. When obtaining the MCMC samples, the acceptance rate in the RwM algorithm was approximately 32%. Besides, we have computed the number of effectively pseudo-independent draws, that is, the Effective Sample Size (ESS) for each parameter. Figures 21 depicts the history of the chains (trace plots) and the marginal *posterior* distributions of the regression coefficients. The normality assumption of the generated chains is quite reasonable, even in the presence of slight tails on the estimated densities. Also, there exists evidence of symmetry since the *posterior* means and medians are very close to each other. For each parameter, the ESS was estimated at approximately half of M, which can be considered an indication of good mixing of the generated chains, without computational waste.

Table 37 – *Posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPS}_u$ fitted model.

Parameter	Mean	Median	Std. Dev.	ESS	95% HPDI	
					Lower	Upper
β_{10}	−1.481	−1.479	0.192	1874.876	−1.868	−1.119
β_{11}	0.935	0.937	0.279	1912.372	0.411	1.497
β_{20}	−1.790	−1.790	0.044	1834.592	−1.873	−1.706
β_{21}	1.062	1.062	0.074	1910.648	0.924	1.211

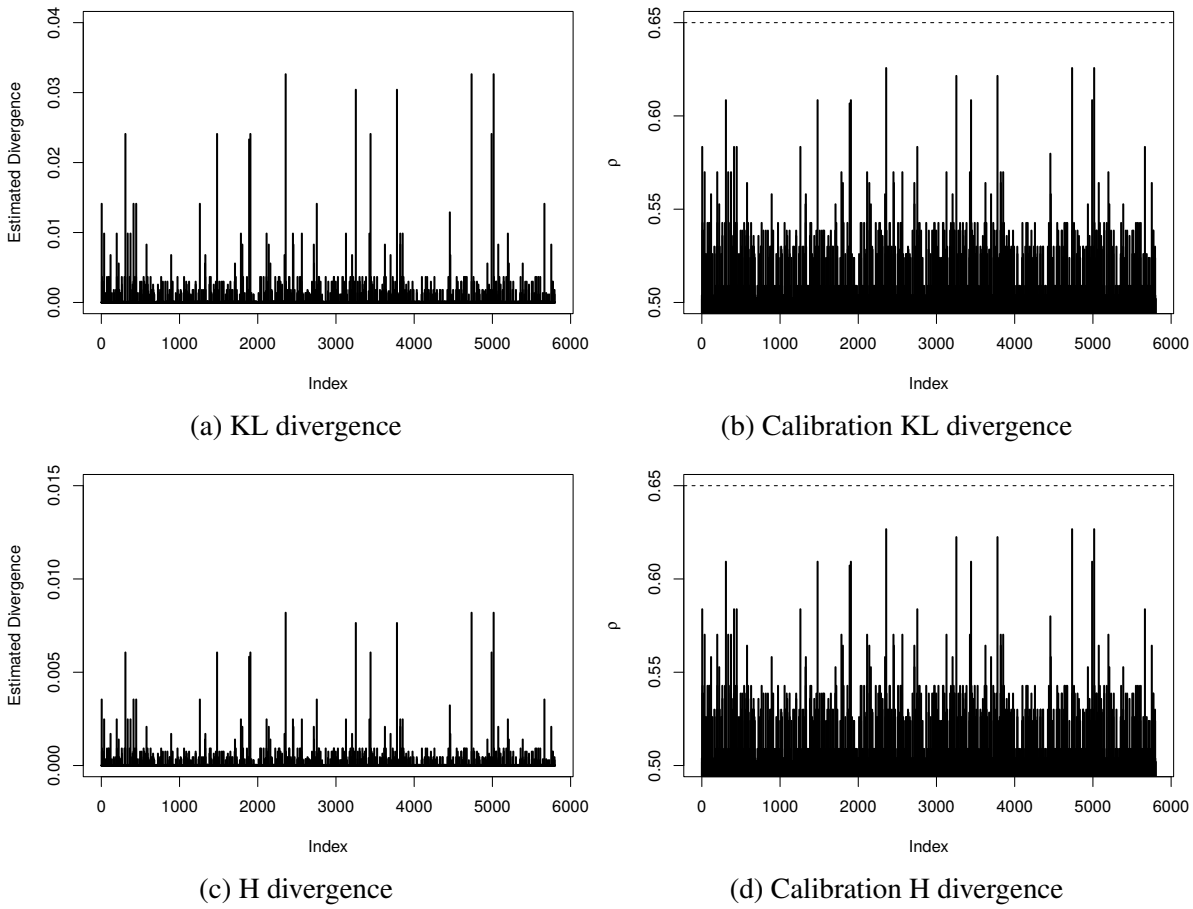
Source: Elaborated by the author.

A sensitivity analysis to verify the existence of influential points is presented in Figure 22. We have estimated all divergence measures presented in Table 90 but, since the obtained results led to the same conclusions, here we are only reporting the KL and H divergences and their calibration for each observation. Even being very conservative by considering an observation whose distance has a calibration exceeding 0.65 as an influential point, we do not have found evidence that any observation has influenced the estimation of any coefficient of the $\mathcal{ZMPS}_u$ regression model significantly.

For comparison purposes, identical Bayesian procedures were adopted to fit the $\mathcal{P}$, the Negative Binomial ($\mathcal{NB}$), the $\mathcal{PS}_u$, the $\mathcal{ZMP}$ and the zero-modified Negative Binomial ($\mathcal{ZMNB}$) regression models. To estimate the fixed dispersion parameter (ϕ) of $\mathcal{NB}$ and

$\mathcal{ZMNB}$ models, we have considered a noninformative Inverse-Gamma *prior* distribution with hyperparameters $a = b = 1.0$. For each fitted model, we have estimated the measures previously discussed in Subsection 5.3.5. The model comparison procedure is summarized in Table 38. One can notice that the zero-modified models have performed considerably better with $\mathcal{ZMPS}_u$ outperforming all. These results are highlighting that the proposed model is highly competitive with well-established models in the literature. This feature can be considered one of the most relevant achievements of the $\mathcal{ZMPS}_u$ model since it has to deal with the positive observations using fewer parameters than, for example, the $\mathcal{ZMNB}$ model.

Figure 22 – Sensitivity analysis for diagnostic of influential points.



(a) KL divergence

(b) Calibration KL divergence

(c) H divergence

(d) Calibration H divergence

Source: Elaborated by the author.

In Table 38, we have also reported the Bayesian p-values as a way to evaluate the adequacy of the fitted models. As expected, the $\mathcal{P}$ model is unsuitable to describe the considered dataset, and the fit provided by the $\mathcal{NB}$ regression model is also highly questionable. For the zero-modified models, there is no indication of overall lack-of-fit, since the *posterior* values of p_{B} were estimated close to 0.50. Figure 23 depicts additional evidence based on the RQRs for validating the fitted $\mathcal{ZMPS}_u$ regression model. This residual metric was computed as discussed in Subsection 5.3.6 using Equation (5.3). One can notice that the normality assumption of the

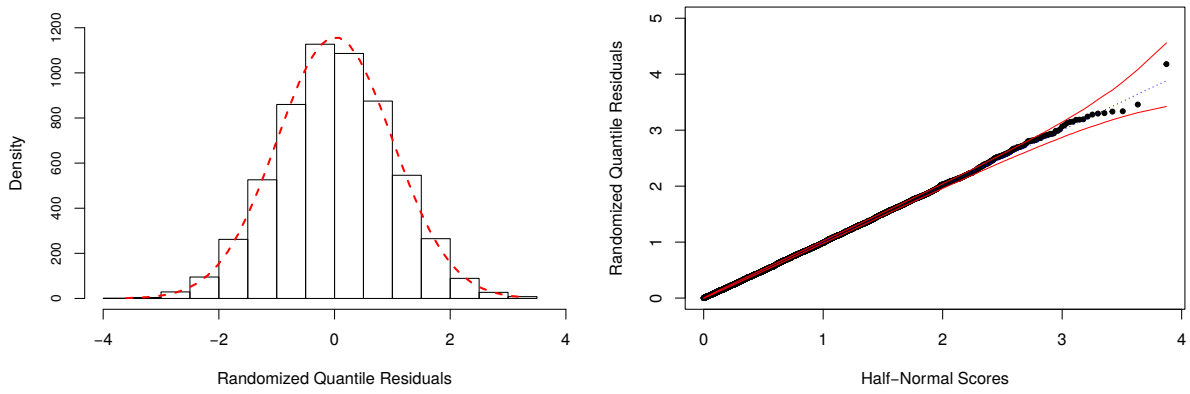
residuals is easily verified by the behavior of its frequency distribution (left-panel). Also, the Half-Normal probability plot indicates that the fit of the $\mathcal{ZMPS}_u$ model was very satisfactory since all estimated residuals are lying within the simulated envelope (right-panel).

Table 38 – Comparison criteria and adequacy measures for the fitted models.

Model	DIC	EAIC	EBIC	NLMPL	p_{B}
$\mathcal{P}$	4650.631	4652.624	4665.955	2325.750	1.000
$\mathcal{NB}$	4340.938	4343.915	4363.912	2170.321	0.936
$\mathcal{PS}_u$	4436.313	4438.312	4451.643	2218.355	0.578
$\mathcal{ZMP}$	4323.300	4327.164	4353.826	2161.734	0.516
$\mathcal{ZMNB}$	4321.668	4326.960	4360.288	2160.530	0.598
$\mathcal{ZMPS}_u$	4320.539	4324.549	4351.212	2160.138	0.554

Source: Elaborated by the author.

Figure 23 – Frequency distribution and Half-Normal plot with simulated envelope for the RQRs.



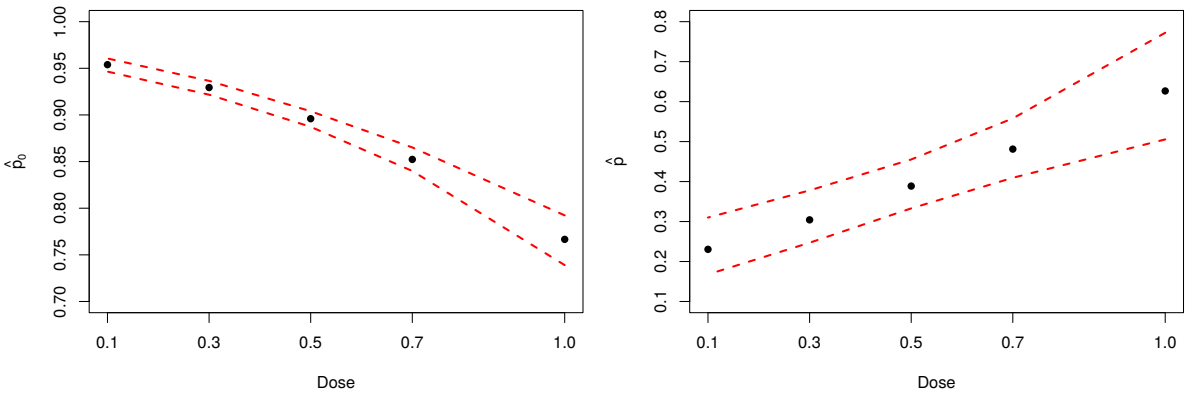
Source: Elaborated by the author.

From the results displayed in Table 37, one can make some conclusions. Firstly, we have observed that the HPDIs of parameters β_{11} and β_{21} do not contain the value zero, which constitutes the dose of ionizing radiation as a relevant covariate to describe the average number of chromosomal aberrations as well the probability of not observing at least one aberration (p_0). For example, the expected number of dicentrics and centric rings in a cell that was exposed to 1.0 Gy is 0.363, and the probability of such aberrations not to occur is $\hat{p}_0 = \Phi(1.790 - 0.319) = 0.929$. Therefore, based on the *posterior* estimates, the components of the fitted $\mathcal{ZMPS}_u$ model can be expressed by

$$\hat{\mu}_{ij} = \exp\{-1.481 + 0.935x_i\} \quad \text{and} \quad \hat{\omega}_{ij} = \Phi(-1.790 + 1.062x_i),$$

where x_i is the dose of ionizing radiation.

Figure 24 – *Posterior* estimates of parameters p_0 and p.

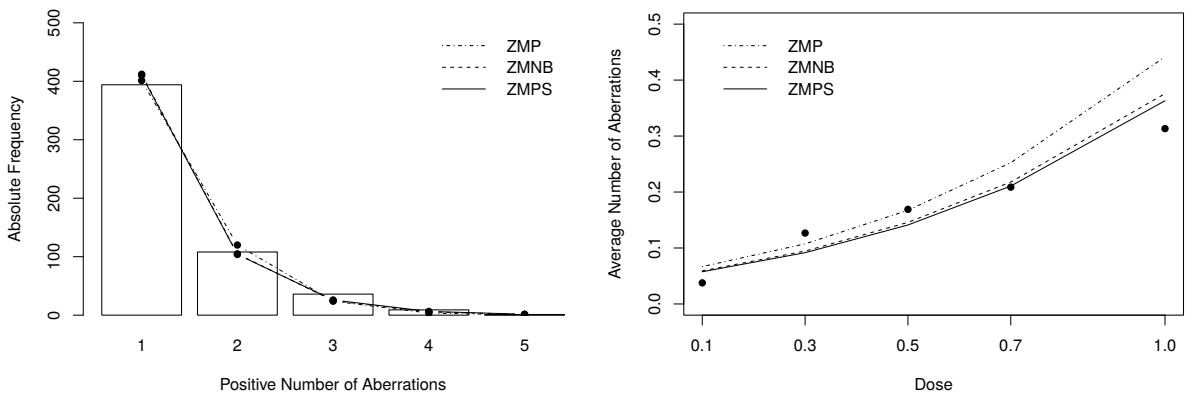


Source: Elaborated by the author.

Figure 24 present the Bayesian estimates, by dose, for the probability of not observing at least one aberration (left-panel) and for parameter p (right-panel). The 95% HPDIs are represented by the red dashed lines. Noticeably, our inferences about parameter p are confirming the initial assumption that the analyzed sample has an excessive number of zero-valued dicentrics and centric rings aberrations.

Table 39 presents a general *posterior* summary of the models that were fitted to the chromosomal aberration data. Here, parameter λ as estimated as $n^{-1}\sum_{i=1}^{5}\sum_{j=1}^{n_i}\hat{\lambda}_{ij}$ and ς^2 was estimated analogously. One can notice that the expected number of zeros $(\hat{n}_0)$ obtained by the $\mathcal{P}$, the $\mathcal{NB}$ and the $\mathcal{PS}_u$ models are slightly lower than the observed n_0 while those provided by the zero-modified models are very close (or exactly equal) to 5252. Through these measures, one can better understand how the fitted models are adhering to the data since the nature of the observed counts should be well described regarding its overdispersion level as well as the frequency and the average number of nonzero observations.

Figure 25 – *Posterior* expected frequencies and dose-response curve fitted by the zero-modified models.



Source: Elaborated by the author.

The goodness-of-fit of the fitted models can be evaluated by the χ^2 statistic, which is obtained from the observed and expected frequencies. To compute such measure, we have grouped cells with frequencies lower or equal than 5, resulting in 4 df. The obtained statistics are also presented in Table 39. Figure 25 depict the positive expected frequencies (left-panel) and the dose-response curves (right-panel) that were estimated using the zero-modified fitted models. From the obtained results, one can conclude that despite the reasonable fit provided by the $\mathcal{ZMNB}$ regression model, the proposed model have adhered better on the chromosomal aberration data and has proved to be an excellent alternative to overlaps all the variability of the response variable in terms of the dose of ionizing radiation.

Table 39 – *Posterior* parameter estimates and goodness-of-fit evaluation.

Model	Parameter	$\hat{\lambda}$	$\hat{\varsigma}^2$	$\hat{n}_0$	χ^2	p-value
$\mathcal{P}$	$\hat{\beta}_{10} = -2.97$ $\hat{\beta}_{11} = 1.95$	0.131	0.131	5086	2343.773	< 0.001
$\mathcal{NB}$	$\hat{\beta}_{10} = -3.02$ $\hat{\beta}_{11} = 2.07$ $\hat{\phi} = 0.28$	0.133	0.232	5202	20.050	< 0.001
$\mathcal{PS}_u$	$\hat{\beta}_{10} = -2.99$ $\hat{\beta}_{11} = 1.98$	0.132	0.157	5126	266.458	< 0.001
$\mathcal{ZMP}$	$\hat{\beta}_{10} = -0.86$ $\hat{\beta}_{11} = 0.82$ $\hat{\beta}_{20} = -1.79$ $\hat{\beta}_{21} = 1.06$	0.132	0.199	5251	16.456	0.002
$\mathcal{ZMNB}$	$\hat{\beta}_{10} = -1.33$ $\hat{\beta}_{11} = 0.88$ $\hat{\beta}_{20} = -1.79$ $\hat{\beta}_{21} = 1.07$ $\hat{\phi} = 1.51$	0.131	0.206	5251	7.255	0.123
$\mathcal{ZMPS}_u$	See Table 37	0.132	0.210	5252	5.298	0.258

Source: Elaborated by the author.

5.6 Concluding remarks

In this chapter, we have introduced the $\mathcal{ZMPS}_u$ regression model as an alternative for the analysis of overdispersed datasets exhibiting zero modification in the presence of covariates. By considering the hurdle version of the $\mathcal{PS}_u$ distribution, it was possible to derive separable likelihood functions for the parameter vectors, which led us to less complicated Bayesian proce-

dures, based on the *g-prior* method. Intensive Monte Carlo simulation studies were performed, and the obtained results have allowed us to assess the empirical properties of the Bayesian estimators and then conclude about the suitability of the adopted methodology to the predefined scenarios.

The proposed model was considered for the analysis of a real dataset on the number of cytogenetic chromosomal aberrations, considering the dose of ionizing radiation as the covariate for both model components. The response variable was identified as overdispersed and heavily zero-inflated, which justified the use of the $\mathcal{ZMPS}_u$ regression model. A sensitivity analysis was conducted by using some standard divergence measures, and no locally influent observations were found. The adequacy of the fitted models was evaluated by using the Bayesian p-value and the RQRs. The main conclusion one can make from the fitted models is that the dose is statistically relevant to describe either the probability of occurrence as well as the average incidence of aberrations. Besides, when looking at the χ^2 statistic and the *posterior*-based comparison criteria, we have noticed that the proposed model has presented a better fit when compared to its competitors and therefore, it can be considered an excellent addition to the set of models that can be used for the analysis of overdispersed and zero-modified count data.

CHAPTER

6

THE ZERO-MODIFIED POISSON-LINDLEY REGRESSION MODEL WITH MIXED-EFFECTS

6.1 Introduction

Count datasets are traditionally analyzed using the ordinary Poisson ($\mathcal{P}$) distribution. However, such a model has its applicability limited as it can be somewhat restrictive to handle specific data structures. In this case, it arises the need for obtaining alternative models that accommodate, for example, (i) zero-modification (inflation or deflation at the frequency of zeros), (ii) overdispersion, and (iii) individual heterogeneity arising from clustering or repeated (correlated) measurements made on the same subject. Cases (i)-(ii) and (ii)-(iii) are often treated together in the statistical literature with several practical applications, but models supporting all at once are less common. Hence, the primary goal of this chapter was to jointly address these issues by deriving a mixed-effects regression model based on the hurdle version of the Poisson-Lindley ($\mathcal{PL}$) distribution. In this framework, the zero-modification is incorporated by considering that a binary probability model determines which outcomes are zero-valued, and a zero-truncated process is responsible for generating positive observations. Approximate *posterior* inferences for the model parameters were obtained from a fully Bayesian approach based on the Adaptive Metropolis algorithm. Intensive Monte Carlo simulation studies were performed as a way to assess the empirical properties of the Bayesian estimators, and the obtained results have been discussed. The proposed model was considered for the analysis of a real dataset, and its competitiveness regarding some well-established mixed-effects models for count data was evaluated. A sensitivity analysis to detect observations that may impact parameter estimates was performed based on some standard divergence measures. The Bayesian p-value and the randomized quantile residuals were considered for the task of model validation.

In this chapter, we have extended the works of Sankaran (1970) and Bertoli *et al.* (2019b) in the sense of developing a new mixed-effects regression model for zero-modified count data based on the $\mathcal{PL}$ distribution. The zero-modified Poisson-Lindley ($\mathcal{ZMPL}$, for short) distribution was firstly introduced by Bertoli *et al.* (2019a). A discrete random variable Y defined on $\mathcal{Y}_0 = \{0, 1, \ldots\}$ is said to follow a $\mathcal{ZMPL}$ distribution if its probability mass function (pmf) can be written as

$$\mathrm{P}_*(Y = y; \mu, p) = (1 - p)\, \delta_y + p\, \mathrm{P}(Y = y; \mu), \quad y \in \mathcal{Y}_0,$$

where p is the zero modification parameter and δ_y is an indicator function, so that $\delta_y = 1$ if $y = 0$ and $\delta_y = 0$ otherwise. Also, $\mu \in \mathbb{R}_+$ is the expected value of the ordinary $\mathcal{PL}$ distribution, whose reparameterized pmf is given by

$$\mathrm{P}(Y = y; \mu) = \frac{(2\mu)^y \left[h(\mu) - \mu + 1\right]^2 \left[h(\mu) + (2y+3)\,\mu + 1\right]}{\left[h(\mu) + \mu + 1\right]^{y+3}}, \quad y \in \mathcal{Y}_0,$$

where $h(\mu) = \sqrt{(\mu-1)^2 + 8\mu}$ and $\mu = (\theta+2)[\theta(\theta+1)]^{-1}$ for $\theta \in \mathbb{R}_+$ (shape parameter). This parameterization is particularly useful since our primary goal is to derive a regression model, in which the influence of fixed and random-effects can be evaluated directly over the mean of a zero-modified response variable. Unlike zero-inflated models, here parameter p is defined on the interval $[0, \mathrm{P}^{-1}(Y > 0; \mu)]$, and so the $\mathcal{ZMPL}$ model is not a mixture distribution since p may assume values greater than 1. The expected value and variance of Y are given, respectively, by $\mathbb{E}(Y) = \lambda = \mu p$ and $\mathbb{V}(Y) = \varsigma^2 = p[\sigma^2 + (1-p)\mu^2]$, where $\sigma^2 \in \mathbb{R}_+$ is the variance of the $\mathcal{PL}$ distribution (see Bertoli *et al.* (2019a), Table 4).

The hurdle version of the $\mathcal{PL}$ distribution can be obtained by taking $\omega = p\,\mathrm{P}(Y > 0; \mu)$, that is,

$$\mathrm{P}_*(Y = y; \mu, \omega) = (1 - \omega)\, \delta_y + \omega\, \mathrm{P}^*(Y = y; \mu), \quad y \in \mathcal{Y}_0, \tag{6.1}$$

for $\omega \in [0, 1]$ and where $\mathrm{P}^*(Y = y; \mu)$ is the pmf of the zero-truncated Poisson-Lindley ($\mathcal{ZTPL}$) distribution (GHITANY; AL-MUTAIRI; NADARAJAH, 2008). Noticeably, Equation (6.1) is only a reparameterization of the standard $\mathcal{ZMPL}$, and so one can conclude that these models are interchangeable. For ease of notation and understanding, the acronym $\mathcal{ZMPL}$ will be used when we refer to the hurdle version of the $\mathcal{PL}$ distribution.

The corresponding cumulative distribution function (cdf) of Y is given by

$$\mathrm{F}_*(y; \mu, \omega) = 1 - \omega \left[\frac{h^2(\mu) + 2\,(\mu y + 2\mu + 1) h(\mu) - (2y+1)\mu^2 + 2\,(y+2)\mu + 1}{(2\mu)^{-y} [h(\mu) + \mu + 1]^y [h^2(\mu) + 2\,(2\mu+1)\,h(\mu) - \mu\,(\mu - 4) + 1]} \right], \quad y \in \mathcal{Y}_0. \tag{6.2}$$

Comparatively, the proposed model can be considered more versatile than the $\mathcal{ZMPL}$ model with fixed-effects, since it is designed to accommodate hierarchical data due to repeated measurements on the same subject. Moreover, it can also be considered more flexible than zero-inflated mixed models as it allows for zero deflation within clusters, which is a structure often

encountered when handling count data. Besides, the proposed model incorporates overdispersion that does not come only from inflation/deflation of zeros, as one of its parts is dedicated to describing the behavior of the positive values. In the regression framework that we have developed, discrepant points (*outliers*) can be identified, and, through a careful sensitivity analysis, it is possible to quantify the influence of such observations. However, since the $\mathcal{PL}$ distribution accounts for different levels of overdispersion, its zero-modified version is naturally a robust alternative as it may accommodate discrepant points that would significantly impact the parameter estimates of the zero-modified Poisson ($\mathcal{ZMP}$) model.

This chapter is organized as follows. In Section 6.2, we present the mixed-effects regression model based on the hurdle version of the $\mathcal{PL}$ distribution. In Section 6.3, we describe all the Bayesian methodologies and associated numerical procedures that were considered for inferential purposes. In Section 6.4, we discuss the results of an intensive simulation study and in Section 6.5, a real data application using the proposed model is exhibited. General comments and concluding remarks are addressed in Section 6.6.

6.2 The ZMPL mixed-effects regression model

Standard mixed-effects models incorporate two vector-valued random variables $(\boldsymbol{Y}, \boldsymbol{b})$ and is fully described by the joint distribution of these vectors. The basic assumption in this framework is that the vector $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ of correlated responses is subject to unobserved random-effects $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, whose elements can be interpreted as latent vectors encompassing characteristics that cannot be measured or pre-specified for each individual, regardless the process of data collection.

Suppose that a random experiment (designed or observational) is conducted with n subjects (clusters) being observed m_i times $(i = 1, \ldots, n)$. The primary response for such an experiment is described by a discrete random variable Y_{ij} denoting the j-th outcome $(j = 1, \ldots, m_i)$ for the i-th subject. The full response vector for each i is given by $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})$, and we assume that the observed vector $\boldsymbol{y}_i$ is obtained conditionally to fixed and random-effects (designed to account for within-subject correlation), here denoted respectively by $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}, \boldsymbol{b}_{2i})$. Assuming that $Y_{ij} | \boldsymbol{\beta}, \boldsymbol{b}_i \sim \mathcal{ZMPL}(\mu_{ij}, \omega_{ij})$ holds for all pairs (i, j), a general mixed-effects regression model for count data based on the $\mathcal{ZMPL}$ distribution can be derived by rewriting Equation (6.1) as

$$\mathrm{P}_*\big(Y_{ij} = y_{ij}; \boldsymbol{\beta}, \boldsymbol{b}_i\big) = \big(1 - \omega_{ij}\big)\, \delta_{y_{ij}} + \omega_{ij} \mathrm{P}^*\big(Y_{ij} = y_{ij}; \mu_{ij}\big), \qquad y_{ij} \in \mathcal{Y}_0, \tag{6.3}$$

where $\mu_{ij} \equiv \mu(\boldsymbol{x}_{1ij}, \boldsymbol{\beta}_1, \boldsymbol{z}_{1ij}, \boldsymbol{b}_{1i})$ and $\omega_{ij} \equiv \omega(\boldsymbol{x}_{2ij}, \boldsymbol{\beta}_2, \boldsymbol{z}_{2ij}, \boldsymbol{b}_{2i})$ are parameterized nonlinear functions. In this framework, we have $\boldsymbol{\beta}_k^\intercal = (\beta_{k0}, \ldots, \beta_{kq_k})$ $(k = 1, 2)$ related to $\boldsymbol{x}_{kij}^\intercal = (1, x_{kij}^1, \ldots, x_{kij}^{q_k})$, where $\boldsymbol{x}_{kij}$ is a vector of covariates that may include, for example, *dummy* variables, cross-level interactions and polynomials. The quantity q_1 (q_2) denotes the number of covariates considered on the systematic component of a linear predictor for parameter μ_{ij} (ω_{ij}). The $\boldsymbol{b}_{ki} = (b_{ki}^1, \ldots, b_{ki}^{r_k})$

is a r_k-dimensional vector containing the coefficients of an individual residual component after the fixed-effects have been accounted for, and $\boldsymbol{z}_{kij}^{\intercal} = (1, z_{kij}^1, \ldots, z_{kij}^{r_k-1})$ is the vector of variables associated to the i-th subject-specific random-effect.

While the order of $\boldsymbol{b}_{ki}$ can be arbitrarily specified, it typically includes only an intercept when the responses are taken from a collection of clustered individuals, and an intercept and slope when the responses are taken sequentially over time for each subject. However, in many practical situations, the most straightforward approach of using a scalar random-effect (random-intercept) as an attempt to overlaps all the individual heterogeneity often provides accurate results combined with model parsimony. In this sense, the forthcoming methodological developments are focused in this particular case, that is, we will consider the univariate setup ($\boldsymbol{b}_{ki} = b_{ki}$ and $\boldsymbol{z}_{kij} = 1$) henceforth.

The full regression matrices of model (6.3) can be written as

$$\boldsymbol{X}_1 = \begin{bmatrix} \mathbf{1}_{m_1} : \boldsymbol{X}_{1,m_1\times q_1} \\ \vdots \\ \mathbf{1}_{m_n} : \boldsymbol{X}_{1,m_n\times q_1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}_2 = \begin{bmatrix} \mathbf{1}_{m_1} : \boldsymbol{X}_{2,m_1\times q_2} \\ \vdots \\ \mathbf{1}_{m_n} : \boldsymbol{X}_{2,m_n\times q_2} \end{bmatrix},$$

where $\mathbf{1}_{m_i}$ is the intercept column and the dense submatrix $\boldsymbol{X}_{k,m_i\times q_k}$ is defined in such a way that its i-th row contains the vector $(x_{kij}^1, \ldots, x_{kij}^{q_k})$. The dimension of $\boldsymbol{X}_k$ is $m \times (q_k + 1)$, where $m = \sum_{i=1}^n m_i$.

Now, we have to specify two monotonic, invertible and twice differentiable link functions, say g_1 and g_2, in which $\mu_{ij} = g_1^{-1}(\boldsymbol{x}_{1ij}^{\intercal}\boldsymbol{\beta}_1 + b_{1i})$ and $\omega_{ij} = g_2^{-1}(\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + b_{2i})$ are well defined on $\mathbb{R}_+$ and $(0,1)$, respectively. For this purpose, one may choose any suitable mappings g_1 and g_2 such that $g_1^{-1}\colon \mathbb{R} \to \mathbb{R}_+$ and $g_2^{-1}\colon \mathbb{R} \to (0,1)$. The logarithm link function, $\log(\mu_{ij}) = \boldsymbol{x}_{1ij}^{\intercal}\boldsymbol{\beta}_1 + b_{1i}$, is the natural choice for g_1. For g_2, the popular choice is the *logit* link function,

$$\operatorname{logit}(\omega_{ij}) = \log\left(\frac{\omega_{ij}}{1-\omega_{ij}}\right) = \boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + b_{2i}. \tag{6.4}$$

The *probit* link function,

$$\Phi^{-1}(\omega_{ij}) = \boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + b_{2i}, \tag{6.5}$$

is also appropriate for the requested purpose. Another possible choice for g_2 is

$$\log\left[-\log(1-\omega_{ij})\right] = \boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + b_{2i}, \tag{6.6}$$

which corresponds to the *complementary log-log* link function. The inclusion of a random-intercept in the linear predictor of ω_i is a way to verify if the heterogeneity between subjects affects the probability of occurrence more than the average number of positive counts.

One can notice that the link functions (6.4)-(6.6) exclude the limit cases $p_{ij} = 0$ and $p_{ij} = \mathrm{P}^{-1}(Y > 0; \mu_{ij})$, since it is not possible to obtain either $\hat{\omega}_i = 0$ or $\hat{\omega}_i = 1$. The link function

(6.6) is usually preferable when the occurrence probability of a specific outcome is considerably high/low as it accommodates asymmetric behaviors on the unit interval, which is not the case for link functions (6.4) and (6.5). Further, a more sophisticated approach considering power and reversal power link functions was proposed by Bazán *et al.* (2017), and can also be used to add even more flexibility when modeling parameter ω_{ij}.

We may refer to the proposed model as a "semi-compatible" regression model. The term "compatible" alludes for "zero-altered", which defines the class of models proposed by Heilbron (1994) and extended by Min and Agresti (2005) as to accommodate repeated measures. Zero-altered models are similar to zero-modified ones, but the compatibility arises from the fact that the linear predictors of μ_{ij} and ω_{ij} are the same. In our case, specifically, it is worthwhile to mention that identifiability problems may occur if one considers a mixed-effects regression model derived directly from (6.1), with parameters μ and p sharing covariates, even if $\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$. Therefore, the adopted structure allows for more flexibility and robustness as μ and ω may share covariates not necessarily with $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1$, and so the only requirement for ensuring model identifiability is the linear independence between covariates within linear predictors.

Unlike traditional approaches, the proposed model can be used for the analysis of zero-inflated and zero-deflated datasets. In this case, given a set of covariates, the probability of a zero-valued count being observed in the j-th outcome for i-th subject is given by $1-g_2^{-1}(\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + b_{2i})$. Under the logistic regression model (6.4), β_{2l} $(l = 1,\ldots,q_2)$ represents the direct change (adjusted by a prediction of b_{2i}) in the log-odds of Y_{ij} being positive per 1-unit change in x_{2ij}^{l}, holding the other covariates at fixed values. On the other hand, the same not apply if one adopts the link function (6.6) since $e^{\beta_{2l}}$ is not the odds ratio for the l-th covariate effect and so β_{2l} do not have a straightforward interpretation in terms of contribution to log-odds. Likewise, it is not possible to interpret the coefficients of the *probit* model (6.5) directly, but one can evaluate the marginal effect of β_{2l} by analyzing how much the conditional probability of Y_{ij} being positive is affected when the value of x_{2ij}^{l} is changed. The exact interpretation of β_{1l} $(l = 1,\ldots,q_1)$ is not direct in terms of the mean of the hurdle model since the positive counts are modeled by a zero-truncated distribution $(\mathcal{ZTPL})$, and therefore, β_{1l} represents the overall effect of x_{1ij}^{l} (adjusted by a prediction of b_{1i}) on the expected value μ_{ij} when $y_{ij} > 0$, holding the other covariates at fixed values.

We complete model specification by choosing an appropriate distribution for the vector of individual random-effects, $\boldsymbol{b}_i = (b_{1i}, b_{2i})^{\intercal}$. The bivariate Normal distribution is the natural choice in this case, but the bivariate t-Student can be considered as well. Other models as the penalized Gaussian mixture (KOMÁREK; LESAFFRE, 2008), the Skew-Normal (HOSSEINI; EIDSVIK; MOHAMMADZADEH, 2011) and the Generalized Log-Gamma (FABIO; PAULA; CASTRO, 2012) can be used to accommodate different behaviors of random-effects. According to McCulloch and Neuhaus (2011), one may obtain good performance in predicting the random-effects by assuming that they are Gaussian, even when comparing this choice with other

distributions. Therefore, here we assume that $\boldsymbol{b}_i \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{b}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ is a 2×2 symmetric unstructured covariance matrix given by

$$\boldsymbol{\Sigma}_{\boldsymbol{b}} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

where $\sigma_k \in \mathbb{R}_+$ is the standard deviation of b_{ki} and $\rho \in [-1, 1]$ is the correlation between b_{1i} and b_{2i}. We may represent the bivariate distribution of $\boldsymbol{b}_i$ using a product-Normal parameterization (COOPER *et al.*, 2007), that is,

$$b_{1i} \sim \mathcal{N}\big(0, \sigma_1^2\big) \quad \text{and} \quad b_{2i}|b_{1i} \sim \mathcal{N}\big(\rho\sigma_2\sigma_1^{-1}b_{1i}, (1-\rho^2)\sigma_2^2\big)\,. \tag{6.7}$$

It can be noticed that, under this formulation, the elements of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ appear exclusively in the (marginal) distribution of $\boldsymbol{b}_i$. However, to increase computational stability and efficiency when estimating mixed-effects models, it is highly advisable to represent the random-effects using a spherical distribution (BATES, 2011). Therefore, assuming $\boldsymbol{u}_i = (u_{1i}, u_{2i})^\top \sim \mathcal{N}_2(\mathbf{0}, \mathcal{I}_2)$, the random-intercepts of model (6.3) will be expressed by $b_{1i} = u_{1i}\sigma_1$ and $b_{2i} = \sigma_2(u_{1i}\rho + u_{2i}\sqrt{1-\rho^2})$ from now on.

By definition, $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ is positive semidefinite, but it becomes natural to restrict the covariance matrix of $\boldsymbol{b}_i$ to be strictly positive-definite since degenerate distributions do not have practical appeal to describe the behavior of random-effects. To ensure that the elements of the main diagonal of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$ are strictly positive, we consider the reparameterization $\sigma_k = e^{\xi_k}$, $\xi_k \in \mathbb{R}$, which also allows for unrestricted estimation. Likewise, for the correlation between the random-intercepts, we adopt the parameterization

$$\rho = \frac{1 - e^{-\upsilon}}{1 + e^{-\upsilon}},$$

for $\upsilon \in \mathbb{R}$.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \xi_1, \xi_2, \upsilon)$ be the full vector of model parameters. Hence, the proposed model has $d = \dim(\boldsymbol{\theta}) = q_1 + q_2 + 5$ unknown quantities to be estimated. A fully Bayesian approach will be considered for parameter estimation and associated inference. The next section is dedicated to present details of such an approach.

6.2.1 Testing for zero modification

When analyzing count data, one may obtain empirical evidence of zero modification by computing the expected frequency at zero under any ordinary discrete distribution, and then checking whether this frequency is notably lower/higher than the observed one. Nonetheless, whenever formal evidence is required, one may resort to zero-altered models. Although somewhat restrictive for practical applications, such a class of models can be used to test for zero modification. In our context, if we decide to evaluate if data provide enough evidence favoring

zero modification regarding the $\mathcal{P}$ distribution, we may consider the zero-altered Poisson ($\mathcal{ZAP}$) model with random-intercept (NEELON; O'MALLEY; NORMAND, 2010), which is defined by the linear predictors

$$\log\left(\mu_{ij}\right) = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + b_i \quad \text{and} \quad \log\left[-\log\left(1-\omega_{ij}\right)\right] = \nu + \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + b_i, \tag{6.8}$$

where $b_i \sim \mathcal{N}(0, \sigma_b^2)$ is a shared random-intercept between the two components. In the $\mathcal{ZAP}$ model, the *complementary log-log* link function for ω_{ij} arises naturally as the induced transformation from $\mathrm{P}(Y_{ij} > 0; \mu_{ij}) = 1 - e^{-\mu_{ij}}$ to the linear predictor $\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + b_i$ (see Heilbron (1994) for the uncorrelated responses case). Therefore, testing for zero modification comes down to testing $\nu = 0$. If $\nu < 0$ ($\nu > 0$), then there is evidence of zero inflation (zero deflation). Conversely, if zero is a plausible value for ν, then there is no evidence supporting zero modification, and so a standard generalized linear model (GLM) suffices for modeling the available data.

Analogously, we may derive a zero-altered version of the $\mathcal{PL}$ ($\mathcal{ZAPL}$) distribution by considering the same structure of model (6.8) but, instead of using the *complementary log-log* link function for ω_{ij}, we should induce a transformation $g(\omega_{ij})$ from $\mathrm{P}(Y_{ij} > 0; \mu_{ij})$ to the linear predictor $\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + b_i$, where

$$\mathrm{P}(Y_{ij} > 0; \mu_{ij}) = 1 - \frac{\left[h(\mu_{ij}) - \mu_{ij} + 1\right]^2 \left[h(\mu_{ij}) + 3\mu_{ij} + 1\right]}{\left[h(\mu_{ij}) + \mu_{ij} + 1\right]^3}.$$

Unfortunately, the desired transformation does not have a closed-form as we are not able to isolate $\log\left(\mu_{ij}\right)$ in the expression of $\mathrm{P}(Y_{ij} > 0; \mu_{ij})$. In this case, it is not possible to obtain a known link function for parameter ω_{ij}. However, despite such an inconvenience, the $\mathcal{ZAPL}$ model is well-defined this way and can be used to testing for zero modification regarding the $\mathcal{PL}$ distribution, likewise described for the $\mathcal{ZAP}$ model.

6.3 Approximate Bayesian inference and prediction

In this section, we address the problem of estimating and making inferences from the proposed model under a fully Bayesian perspective. Firstly, we derive the model likelihood function and then, a suitable set of *prior* distributions is considered in order to obtain a computationally tractable *posterior* density for the vector $\boldsymbol{\theta}$. Beyond the primary distributional assumption that $Y_{ij}|\boldsymbol{\theta}, \boldsymbol{u}_i \equiv Y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i \sim \mathcal{ZMPL}(\mu_{ij}, \omega_{ij})$ holds for all pairs (i, j), here we also assume that (i) the outcomes for the i-th subject are conditionally independent given the vector of random-intercepts, and (ii) the outcomes for different subjects are unconditionally independent.

These assumptions are essential for the hierarchical formulation of the proposed model. Particularly, by (i), one can obtain the conditional distribution of the i-th subject response vector $\boldsymbol{Y}_i$ as

$$\mathrm{P}_*(\boldsymbol{Y}_i = \boldsymbol{y}_i; \boldsymbol{\theta}, \boldsymbol{u}_i) \;=\; \prod_{j=1}^{m_i} \mathrm{P}_*\left(Y_{ij} = y_{ij}; \boldsymbol{\theta}, \boldsymbol{u}_i\right)$$

$$= \prod_{j=1}^{m_i} \omega_{ij} \left(\frac{1-\omega_{ij}}{\omega_{ij}} \right)^{\delta_{y_{ij}}} \left[\frac{\mathrm{P}\left(Y_{ij}=y_{ij};\mu_{ij}\right)}{\mathrm{P}\left(Y_{ij}>0;\mu_{ij}\right)} \right]^{1-\delta_{y_{ij}}},$$

where $\mu_{ij} = g_1^{-1}(\boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1})$ and

$$\omega_{ij} = g_2^{-1}\left[\boldsymbol{x}_{2ij}^{\top}\boldsymbol{\beta}_2 + u_{1i}e^{\xi_2}\left(1-e^{-\upsilon}\right)\left(1+e^{-\upsilon}\right)^{-1} + u_{2i}e^{\xi_2}\sqrt{1-\left(1-e^{-\upsilon}\right)^2\left(1+e^{-\upsilon}\right)^{-2}} \right].$$

For a fixed value of vector $\boldsymbol{u}_i$, the conditional distribution of $\boldsymbol{Y}_i = \boldsymbol{y}_i$ can be recognized as the likelihood function of $\boldsymbol{\theta}$, considering that the available data are provided only by the i-th subject. In this case, the subject-specific log-likelihood function of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} \ell_i(\boldsymbol{\theta};\boldsymbol{y}_i,\boldsymbol{u}_i) &= \sum_{j=1}^{m_i}\left(1-\delta_{y_{ij}}\right)\left\{\log\left[\mathrm{P}\left(Y_{ij}=y_{ij};\mu_{ij}\right)\right] - \log\left[\mathrm{P}\left(Y_{ij}>0;\mu_{ij}\right)\right]\right\} + \\ &\quad \sum_{j=1}^{m_i}\left[\log\left(\omega_{ij}\right) - \delta_{y_{ij}}\log\left(\frac{\omega_{ij}}{1-\omega_{ij}}\right)\right] \\ &= \ell_{1i}(\boldsymbol{\beta}_1,\xi_1;\boldsymbol{y}_i,u_{1i}) + \ell_{2i}(\boldsymbol{\beta}_2,\xi_2,\upsilon;\boldsymbol{y}_i,\boldsymbol{u}_i). \end{aligned}$$

In this work, we will consider a log-linear model for parameter μ_{ij}, that is, $\log(\mu_{ij}) = \boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1}$. From the individual log-likelihood function, one can easily notice that ℓ_{1i} depends only on the positive-valued elements of $\boldsymbol{y}_i$ and therefore, it can be expressed as

$$\begin{aligned} \ell_{1i}(\boldsymbol{\beta}_1,\xi_1;\boldsymbol{y}_i,u_{1i}) &= \sum_{j\in\mathcal{J}_1}\log\left[h\left(\exp\left\{\boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1}\right\}\right) + y_{ij} + 2\right] + \\ &\quad 2\sum_{j\in\mathcal{J}_1}\log\left[h\left(\exp\left\{\boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1}\right\}\right)\right] - \\ &\quad \sum_{j\in\mathcal{J}_1}\log\left[h^2\left(\exp\left\{\boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1}\right\}\right) + \right. \\ &\quad \left. 3h\left(\exp\left\{\boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1}\right\}\right) + 1\right] - \\ &\quad \sum_{j\in\mathcal{J}_1} y_{ij}\log\left[h\left(\exp\left\{\boldsymbol{x}_{1ij}^{\top}\boldsymbol{\beta}_1 + u_{1i}e^{\xi_1}\right\}\right) + 1\right], \end{aligned}$$

where $\mathcal{J}_1 = \{j : y_{ij} > 0, y_{ij} \in \boldsymbol{y}_i\}$ is the finite set of indexes regarding the positive observations of $\boldsymbol{y}_i$. Adopting this setup is equivalent to assuming that each positive element of $\boldsymbol{y}_i$ comes from a $\mathcal{ZTPL}$ distribution. Here, we are extending the fact that estimating the $\mathcal{P}$ parameter θ using the zero-truncated Poisson ($\mathcal{ZTP}$) distribution results in a loss of efficiency in the inference if there is no zero modification (DIETZ; BÖHNING, 2000; CONCEIÇÃO; ANDRADE; LOUZADA, 2014).

The choice of g_2 is left open and so ℓ_{2i} will be generically expressed by

$$\ell_{2i}(\boldsymbol{\beta}_2,\xi_2,\upsilon;\boldsymbol{y}_i,\boldsymbol{u}_i) = \sum_{j=1}^{m_i}\log\left(\omega_{ij}\right) - \sum_{j\in\mathcal{J}_2}\log\left(\omega_{ij}\right) + \sum_{j\in\mathcal{J}_2}\log\left(1-\omega_{ij}\right),$$

where $\mathcal{J}_2 = \{j : y_{ij} = 0, y_{ij} \in \boldsymbol{y}_i\}$ is a finite set of indexes regarding the zero-valued observations of $\boldsymbol{y}_i$. Due to orthogonality between the vectors of fixed-effects, if we consider ξ_1, ξ_2 and υ as *nuisance* parameters, then it is possible to assess the contribution of each subject to the log-likelihood functions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

Now, by integrating out the 2-dimensional spherical random variable $\boldsymbol{u}_i$, one can obtain the marginal distribution of $\boldsymbol{Y}_i$ as

$$\mathrm{P}_*(\boldsymbol{Y}_i = \boldsymbol{y}_i; \boldsymbol{\theta}) = \int_{\mathbb{R}^2} \exp\{\ell_i(\boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{u}_i)\} f(\boldsymbol{u}_i) \, \mathrm{d}\boldsymbol{u}_i,$$

which clearly does not have a closed-form, whichever the distribution adopted for $\boldsymbol{u}_i$. Since our assumption is that $\boldsymbol{u}_i \sim \mathcal{N}_2(\boldsymbol{0}, \mathcal{I}_2)$, it follows from (ii) that the full likelihood function of $\boldsymbol{\theta}$ can be written as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) &= \frac{1}{(2\pi)^n} \prod_{i=1}^{n} \int_{\mathbb{R}^2} \exp\{\ell_i(\boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{u}_i)\} \exp\left\{-\frac{1}{2}\boldsymbol{u}_i^\top \boldsymbol{u}_i\right\} \mathrm{d}\boldsymbol{u}_i \\
&= \frac{1}{(2\pi)^n} \prod_{i=1}^{n} \int_{\mathbb{R}^2} \exp\left\{-\left[\frac{1}{2}\boldsymbol{u}_i^\top \boldsymbol{u}_i - \ell_i(\boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{u}_i)\right]\right\} \mathrm{d}\boldsymbol{u}_i \\
&= \frac{1}{(2\pi)^n} \prod_{i=1}^{n} \int_{\mathbb{R}^2} \exp\{-r_i(\boldsymbol{u}_i; \boldsymbol{y}_i, \boldsymbol{\theta})\} \, \mathrm{d}\boldsymbol{u}_i,
\end{aligned}$$

and so the corresponding log-likelihood function has the form

$$\ell(\boldsymbol{\theta}; \boldsymbol{y}) = -n \log(2\pi) + \sum_{i=1}^{n} \log\left[\int_{\mathbb{R}^2} \exp\{-r_i(\boldsymbol{u}_i; \boldsymbol{y}_i, \boldsymbol{\theta})\} \, \mathrm{d}\boldsymbol{u}_i\right]. \tag{6.9}$$

A particular case arises if one considers that the linear predictors of μ_{ij} and ω_{ij} share the same random-intercept, that is, $b_{1i} = b_{2i} = b_i$. In this framework, if we specify $b_i = u_i \sigma_b$, with $u_i \sim \mathcal{N}(0, 1)$, then the covariance matrix $\boldsymbol{\Sigma_b}$ reduces to the scalar σ_b, and so the log-likelihood function of vector $(\boldsymbol{\beta}, \sigma_b)$ can be expressed as

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \sigma_b; \boldsymbol{y}) &= -\frac{n}{2}\log(2\pi) + \sum_{i=1}^{n} \log\left[\int_{\mathbb{R}} \exp\{\ell_i(\boldsymbol{\beta}, \sigma_b; \boldsymbol{y}_i, u_i)\} \exp\left\{-\frac{1}{2}u_i^2\right\} \mathrm{d}u_i\right] \\
&= -\frac{n}{2}\log(2\pi) + \sum_{i=1}^{n} \log\left[\int_{\mathbb{R}} \exp\{-r_i(u_i; \boldsymbol{y}_i, \boldsymbol{\beta}, \sigma_b)\} \, \mathrm{d}u_i\right],
\end{aligned}$$

where

$$\begin{aligned}
\ell_i(\boldsymbol{\theta}; \boldsymbol{y}_i, \boldsymbol{u}_i) &= \sum_{j \in \mathcal{J}_1} \log\left[\mathrm{P}\left(Y_{ij} = y_{ij}; \exp\left\{\boldsymbol{x}_{1ij}^\top \boldsymbol{\beta}_1 + u_i \sigma_b\right\}\right)\right] - \\
&\quad \sum_{j \in \mathcal{J}_1} \log\left[\mathrm{P}\left(Y_{ij} > 0; \exp\left\{\boldsymbol{x}_{1ij}^\top \boldsymbol{\beta}_1 + u_i \sigma_b\right\}\right)\right] + \\
&\quad \sum_{j=1}^{m_i} \log\left[g_2^{-1}\left(\boldsymbol{x}_{2ij}^\top \boldsymbol{\beta}_2 + u_i \sigma_b\right)\right] - \sum_{j \in \mathcal{J}_2} \log\left[\frac{g_2^{-1}\left(\boldsymbol{x}_{2ij}^\top \boldsymbol{\beta}_2 + u_i \sigma_b\right)}{1 - g_2^{-1}\left(\boldsymbol{x}_{2ij}^\top \boldsymbol{\beta}_2 + u_i \sigma_b\right)}\right].
\end{aligned}$$

Since the integral over the unbounded convex set $\mathbb{R}^2$ in Equation (6.9) is analytically intractable, successive numerical approximations are required to compute the full log-likelihood function of the proposed model. For each subject, let this bidimensional integral be denoted by A_i. Here we will consider the Laplace method (SMALL, 2010) to obtain approximations for A_i, but several other approaches can be found in Evans and Swartz (2000). Thus, let $\tilde{\boldsymbol{u}}_i = (\tilde{u}_{1i}, \tilde{u}_{2i})$ be the unique minimum of the nonlinear function r_i (the gradient of r_i vanishes at $\tilde{\boldsymbol{u}}_i$). A crucial assumption for applying the Laplace method is that the Hessian matrix of r_i,

$$\boldsymbol{\mathcal{V}}_i = \frac{\partial^2 r_i(\boldsymbol{u}_i; \boldsymbol{y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{u}_i \partial \boldsymbol{u}_i^\top},$$

is positive-definite for all $\boldsymbol{u}_i \in \mathbb{R}^2$. In this case, by performing a multivariate Taylor series expansion of r_i and the exponential function, the Laplace approximation for A_i is given by

$$A_i^L \approx \frac{2\pi \exp\{-r_i(\tilde{\boldsymbol{u}}_i; \boldsymbol{y}_i, \boldsymbol{\theta})\}}{\sqrt{\det \tilde{\boldsymbol{\mathcal{V}}}_i}},$$

where $\tilde{\boldsymbol{\mathcal{V}}}_i$ is the Hessian matrix evaluated at $\tilde{\boldsymbol{u}}_i$. Therefore, the full log-likelihood function of $\boldsymbol{\theta}$ can be approximated by

$$\ell^L(\boldsymbol{\theta}; \boldsymbol{y}) \approx -n \log(2\pi) + \sum_{i=1}^{n} \log\left(A_i^L\right) = -\frac{1}{2} \sum_{i=1}^{n} \log(\det \tilde{\boldsymbol{\mathcal{V}}}_i) - \sum_{i=1}^{n} r_i(\tilde{\boldsymbol{u}}_i; \boldsymbol{y}_i, \boldsymbol{\theta}). \tag{6.10}$$

The Laplace approximation is accurate to $\mathcal{O}(n^{-1})$ since it takes into account only the first-order terms of the Taylor series expansion (RULI; SARTORI; VENTURA, 2016). The application of this method in the univariate case is less cumbersome as it only requires the evaluation of a unidimensional minimization problem.

6.3.1 Prior *distributions*

6.3.1.1 *Fixed-effects parameters*

The *g-prior* (ZELLNER, 1986) is a popular choice among Bayesian users of the multiple linear regression model, mainly due to the fact of providing a closed-form *posterior* distribution for the regression coefficients. The *g-prior* is classified as an objective *prior* method which uses the inverse of the Fisher information matrix up to a scalar variance factor ($\tau \in \mathbb{R}_+$) to obtain the *prior* correlation structure of the multivariate Normal distribution. Such specification is indeed quite attractive since the Fisher information plays a major role in the determination of large-sample covariance in both Bayesian and classical inference.

The problem of eliciting conjugate *priors* for GLMs was addressed by Chen and Ibrahim (2003). Their approach can be considered as a generalization of the original *g-prior* method, but its application is restricted for the class of GLMs since the proposed *prior* does not have closed-form for non-normal exponential families. As an alternative, Gupta and Ibrahim (2009)

have proposed the information matrix *prior* as a way to assess the *prior* correlation structure between the coefficients, not including the intercept since the regression matrix is centered as to ensure that β_0 is orthogonal to the other coefficients. This method uses the Fisher information similarly to a precision matrix whose elements are shrunken by the factor τ, which is considered fixed ($\tau \geqslant 1$). However, the authors have pointed out that such class of *priors* can only be considered Gaussian *priors* if the Fisher information matrix does not depend on the vector $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_q)$. In this way, Bové and Held (2011) had considered a similar approach when they proposed a class of *hyper-g priors* for GLMs, where the precision matrix is evaluated at the *prior* mode, hence obtaining an information matrix that is $\boldsymbol{\beta}'$ free.

The formal concept behind the information matrix *prior* is closely related to the unit information *prior* (KASS; WASSERMAN, 1995), whose main idea is that the amount of information provided by a *prior* distribution must be the same as the amount of information contained in a single observation. Such an idea can be applied in the previously mentioned approaches by simply considering $\tau = n$ (total sample size). Gupta and Ibrahim (2009) have also considered fixed values for τ ($\tau \geqslant 1$). On the other hand, some works, including Hansen and Yu (2003), Wang and George (2007) and Bové and Held (2011) do consider *prior* elicitation and inference procedures for the variance scale factor. Here, we will adopt a methodology based on the unit information *prior* idea combined with the "noninformative *g-prior*" proposed by Marin and Robert (2007) for binary regression models. Based on such an approach, it is possible to obtain a quite simple *prior* distribution for the fixed-effects of the proposed model, which is

$$\boldsymbol{\beta} \sim \mathcal{N}_q\left(\mathbf{0}, m\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0\right),$$

where $q = q_1 + q_2 + 2$. Unlike Marin and Robert's approach, we consider standardized regression matrices in the *prior* specification. We denote the full standardized vector of regression coefficients by $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*)$. In this case, the above *prior* distribution becomes $\boldsymbol{\beta}^* \sim \mathcal{N}_q(\mathbf{0}, m(\boldsymbol{G}^\top \boldsymbol{G})^{-1})$, in which

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{X}_1^* & \mathbf{0}_{m\times(q_2+1)} \\ \mathbf{0}_{m\times(q_1+1)} & \boldsymbol{X}_2^* \end{bmatrix},$$

where $\boldsymbol{X}_k^*$ is the standardized version of $\boldsymbol{X}_k$. The first column of $\boldsymbol{X}_k^*$ is the same as for $\boldsymbol{X}_k$. The remaining entries can be obtained by taking the difference between the element and the column mean (c_k^{l+1}) ($l = 1, \ldots, q_k$), divided by the respective standard deviation (s_k^{l+1}). It is worthwhile to mention that, in cases where $\boldsymbol{X}_k^*$ is rank deficient ($m < q_k + 1$) or contains collinear covariates, it is highly advisable to compute the generalized inverse of $\boldsymbol{G}^\top \boldsymbol{G}$ otherwise the *prior* covariance matrix of $\boldsymbol{\beta}^*$ may not be defined.

Such structure for the *prior* covariance matrix ensures that β_{k0}^* and $(\beta_{k1}^*, \ldots, \beta_{kq_k}^*)$ are uncorrelated a *priori*, with β_{k0}^* and β_{kl}^* denoting the (fixed) standardized intercepts and slopes, respectively. In this setup, the original regression coefficients can be expressed as

$$\beta_{kl} = \frac{\beta_{kl}^*}{s_k^{l+1}} \quad \text{and} \quad \beta_{k0} = \beta_{k0}^* - \sum_{l=1}^{q_k} c_k^{l+1} \beta_{kl},$$

and so we have to redefine the full vector of parameters as $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*, \xi_1, \xi_2, \upsilon)$, since here we are also considering reparameterizations for the fixed-effects of the proposed model. Finally, when fitting the $\mathcal{ZAPL}$ regression model, we may consider a noninformative Gaussian *prior* distribution for ν with zero mean, that is, $\nu \sim \mathcal{N}(0, 10^3)$.

6.3.1.2 Random-effects parameters

The product-Normal parameterization (interrelated univariate Normal distributions) considered for the vector of random-intercepts (see Section 6.2) makes it possible to assign any suitable marginal *prior* distribution for the random-effects parameters, instead of using, for example, an Inverse-Wishart or Wishart *prior* distribution for $\boldsymbol{\Sigma_b}$, which usually yield biased *posterior* estimates (particularly if the sample size is small) as it considers common degrees of freedom for all variance components (BARNARD; MCCULLOCH; MENG, 2000).

Our primary assumption for *prior* specification here is that the parameters of (6.7) are mutually independent, and the whole set (of random-effects parameters) is independent from vector $\boldsymbol{\beta}$. In this case, we firstly assume a Half-Cauchy *prior* distribution for σ_k, $\sigma_k \sim \mathcal{HC}(\alpha_k)$, whose probability density function (pdf) is given by

$$\pi(\sigma_k; \alpha_k) = \frac{2\alpha_k}{\pi\left(\sigma_k^2 + \alpha_k^2\right)}, \qquad \sigma_k \in \mathbb{R}_+,$$

where $\alpha_k \in \mathbb{R}_+$ is a scale parameter. The $\mathcal{HC}$ distribution was cited by Gelman (2006) and Polson and Scott (2012) as an alternative to the traditionally used Inverse Gamma *prior*. Here, we set $\alpha_k = 25$ to epitomize noninformativeness about σ_k. Hence, the joint *prior* distribution for the variances of the random-intercepts can be expressed as

$$\pi(\sigma_1, \sigma_2) \propto \prod_{k=1}^{2} \frac{1}{\sigma_k^2 + 625}, \qquad \sigma_k \in \mathbb{R}_+.$$

From that, one can easily derive the joint distribution for the reparameterization (ξ_1, ξ_2) using the Jacobian transformation method. Then,

$$\pi(\xi_1, \xi_2) \propto \prod_{k=1}^{2} \frac{e^{\xi_k}}{e^{2\xi_k} + 625}, \qquad \xi_k \in \mathbb{R}.$$

For correlation parameters, a popular noninformative *prior* is the Uniform distribution on $(-1, 1)$. Thus, by assuming that $\rho \sim \mathcal{U}(-1, 1)$, the obtained distribution for the reparameterization υ can be written as

$$\pi(\upsilon) = \frac{e^{-\upsilon}}{(1 + e^{-\upsilon})^2}, \qquad \upsilon \in \mathbb{R},$$

which can be recognized as the pdf of the standard Logistic distribution.

6.3.2 Posterior *distribution and estimation*

Considering the outlined structure for the $\mathcal{ZMPL}$ regression model with random-intercepts, the full joint *posterior* distribution of the unknown vector $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta};\boldsymbol{y}) = c^{-1}\exp\{\ell(\boldsymbol{\theta};\boldsymbol{y})\}\,\pi(\boldsymbol{\theta}) = c^{-1}\exp\{\ell(\boldsymbol{\theta};\boldsymbol{y}) + \log[\pi(\boldsymbol{\theta})]\},$$

where ℓ corresponds to the log-likelihood function (6.9),

$$c = \int_{\mathbb{R}^d} \exp\{\ell(\boldsymbol{\theta};\boldsymbol{y}) + \log[\pi(\boldsymbol{\theta})]\}\mathrm{d}\boldsymbol{\theta} = \int_{\mathbb{R}^d} \exp\{-r(\boldsymbol{\theta};\boldsymbol{y})\}\mathrm{d}\boldsymbol{\theta}$$

is the marginal pmf of the observed vector $\boldsymbol{y}$ (normalizing constant) and

$$\pi(\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2m}\boldsymbol{\beta}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0\right)^{-1}\boldsymbol{\beta}\right\}\frac{e^{-\upsilon}}{(1+e^{-\upsilon})^2}\prod_{k=1}^{2}\frac{e^{\xi_k}}{e^{2\xi_k}+625}$$

is the *kernel* of the joint *prior* distribution of $\boldsymbol{\theta}$.

Any method to approximate the *posterior* distribution $\pi(\boldsymbol{\theta};\boldsymbol{y})$ will behave well if $c \in (0,\infty)$. Ignoring cases where c is not finite may lead to misleading inferences and conclusions. The normalizing constant is clearly greater than zero as the integrand is an exponential function. On the other hand, the finiteness of c depends either on the properness of $\pi(\boldsymbol{\theta})$ and the behavior of the log-likelihood function for limiting values on the d-dimensional parametric space of $\boldsymbol{\theta}$. Noticeably, the *prior* distribution $\pi(\boldsymbol{\theta})$ is proper on $\mathbb{R}^d$, since it is defined by the product of square-integrable densities. For a given value of $\boldsymbol{\theta}$, the function to be integrated when obtaining the normalizing constant can be approximated by using (6.10) as $\exp\{-r^L(\boldsymbol{\theta};\boldsymbol{y})\}$, where $-r^L(\boldsymbol{\theta};\boldsymbol{y}) = \ell^L(\boldsymbol{\theta};\boldsymbol{y}) + \log[\pi(\boldsymbol{\theta})]$. In this way, the Laplace approximation for c is given by

$$c^L \approx \frac{2\pi\exp\{-r^L(\tilde{\boldsymbol{\theta}};\boldsymbol{y})\}}{\sqrt{\det\tilde{\boldsymbol{\mathcal{V}}}_{\boldsymbol{\theta}}}},$$

where $\tilde{\boldsymbol{\theta}}$ is the unique minimum of r^L and $\tilde{\boldsymbol{\mathcal{V}}}_{\boldsymbol{\theta}}$ is the corresponding Hessian matrix evaluated at $\tilde{\boldsymbol{\theta}}$.

Clearly, if we manage to obtain a finite approximation for the normalizing constant, then the Hessian matrix of r^L is necessarily positive-definite. Thus, one may assume that, if the positive definiteness of $\boldsymbol{\mathcal{V}}_{\boldsymbol{\theta}}$ holds when considering a particular response vector $\boldsymbol{y}$, then the resulting *posterior* distribution is proper in the sense that it is nonnegative and integrates to 1. In this case, all Bayesian computations can be made, without loss of information, by considering the proportionality

$$\pi(\boldsymbol{\theta};\boldsymbol{y}) \propto \exp\{\ell(\boldsymbol{\theta};\boldsymbol{y}) + \log[\pi(\boldsymbol{\theta})]\},$$

which can be approximated by

$$\pi^L(\boldsymbol{\theta};\boldsymbol{y}) \approx \exp\left\{\ell^L(\boldsymbol{\theta};\boldsymbol{y}) + \log[\pi(\boldsymbol{\theta})]\right\}. \tag{6.11}$$

From the Bayesian point of view, inferences for the elements of $\boldsymbol{\theta}$ can be obtained from their marginal *posterior* distribution. However, deriving analytical expressions for these densities is infeasible, mainly due to the complexity of the associated log-likelihood function. In this case, to make inferences for $\boldsymbol{\theta}$, we must resort to a suitable iterative procedure to drawn pseudo-random samples from the approximate *posterior* density (6.11). Hence, aiming to generate N values for $\boldsymbol{\theta}$, we will adopt the Adaptive Metropolis (AM) algorithm proposed by Haario, Saksman and Tamminen (2001), which is based on the well-known Random-walk Metropolis (RwM) algorithm (METROPOLIS *et al.*, 1953; ROBERTS; GELMAN; GILKS, 1997). Both methods consider the multivariate Normal as proposal (candidate-generating) distribution, but in the AM algorithm, the covariance matrix of the Gaussian proposed values is adapted using all information so far gathered about the target distribution.

Let the sequence $\{\boldsymbol{\theta}_k\}_{k=0}^{t-1}$ be the history of the chains so far runned, where $\boldsymbol{\theta}_0$ is the initial state vector. At each state t, the scheme to obtain an update of $\boldsymbol{\theta}$ is based on proposing a candidate $\boldsymbol{\theta}'$ as

$$\boldsymbol{\theta}'|\boldsymbol{\theta}_{t-1} \sim \begin{cases} \mathcal{N}_d(\boldsymbol{\theta}_{t-1}, 0.10^2 d^{-1}\mathcal{I}_d), & \text{for } t \leqslant 2d \\ (1-\psi)\mathcal{N}_d(\boldsymbol{\theta}_{t-1}, 0.10^2 d^{-1}\mathcal{I}_d) + \psi\mathcal{N}_d(\boldsymbol{\theta}_{t-1}, 2.38^2 d^{-1}\mathcal{H}_{t-1}), & \text{otherwise,} \end{cases}$$

where $\psi \in (0,1)$ is a user-defined mixture parameter and $\mathcal{H}_{t-1}$ is the empirical covariance matrix of $\boldsymbol{\theta}'$, that is,

$$\mathcal{H}_{t-1} = \frac{1}{t}\sum_{j=0}^{t-1}(\boldsymbol{\theta}_j - \overline{\boldsymbol{\theta}}_t)(\boldsymbol{\theta}_j - \overline{\boldsymbol{\theta}}_t)^\top,$$

with $\overline{\boldsymbol{\theta}}_t = t^{-1}\sum_{i=0}^{t-1}\boldsymbol{\theta}_i$. The scaling factor applied on $\mathcal{H}_{t-1}$ is an effort to make our procedure resembles the approach of Roberts, Gelman and Gilks (1997), in which the use of a Gaussian proposal with fixed covariance matrix equal to $2.38^2 d^{-1}\mathcal{H}$ is optimal in a particular high-dimensional case. Besides, the use of a 2-component mixture to propose candidates after the initial period $(t \leqslant 2d)$ is an alternative to prevent the algorithm of getting stuck in certain regions of the parametric space, whose representative sequences lead to singular empirical covariance matrices (ROBERTS; ROSENTHAL, 2009).

At any state t, a move from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}'$ is accepted with probability

$$\alpha\left(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}'\right) = \min\left\{\frac{\pi\left(\boldsymbol{\theta}';\boldsymbol{y}\right)}{\pi\left(\boldsymbol{\theta}_{t-1};\boldsymbol{y}\right)}, 1\right\},$$

which is the same acceptance scheme used in the RwM algorithm. However, due to the adaptive nature of the process, the AM algorithm is classified as a non-Markovian method but, as established by Haario, Saksman and Tamminen (2001), it holds correct ergodic properties. In this way, even that the acceptance probability is not based on symmetry and reversibility conditions ($\boldsymbol{\theta}'$ depends on the whole history before t), we may consider this mechanism as suitable to accepting/rejecting candidates for $\boldsymbol{\theta}$ when applying the AM algorithm. Regarding the acceptance rate, we refer to Gelman *et al.* (1996), who suggest that when this quantity ranges between 15 and

40%, the obtained chains are most efficient in the sense of better representing a possible sample collection from the *posterior* distribution. To the best of our knowledge, achieving acceptance rates in this range is routine when fitting hierarchical models from a Bayesian point of view.

The procedure to generate pseudo-random samples from the approximate *posterior* distribution of $\boldsymbol{\theta}$ is summarized in Algorithm 9 (see Appendix D, Section D.2). To run it, one has to specify the size of the chains to be generated (N), the initial state vector $(\boldsymbol{\theta}_0)$, and the mixture parameter (ψ) beforehand. The convergence of the simulated sequences can be monitored by using trace and autocorrelation plots, as well as the run-length control method with half-width test (HEIDELBERGER; WELCH, 1983), the Geweke z-score diagnostic (GEWEKE, 1992), and the Brooks-Gelman-Rubin scale-reduction statistic (BROOKS; GELMAN, 1998). After diagnosing convergence, some samples can be discarded as burn-in. The strategy to decrease the correlation between and within generated chains is based on getting thinned steps, and so the final sample is supposed to have size $M \ll N$. A full descriptive summary of (6.11) can be obtained through Approximate Monte Carlo (AMC) estimators using the sequence $\{\boldsymbol{\theta}_t\}_{t=1}^{M}$. We choose the *posterior* expected value as the Bayesian point estimator for $\boldsymbol{\theta}$, that is,

$$\hat{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\theta}}^{L}(\boldsymbol{y}) \approx \frac{1}{M}\sum_{t=1}^{M}\boldsymbol{\theta}_t, \tag{6.12}$$

which is also known as the minimum mean square error estimator.

In the next section, we discuss the results of the Monte Carlo simulation studies that were performed as a way to assess the performance of the proposed Bayesian methodology. In Section 6.5, the usefulness, and the competitiveness of the proposed model are illustrated by using a real dataset. All computations were performed using the `R` environment (R Development Core Team, 2017).

6.3.3 Random-effects prediction

Beyond parameter estimation, any further inferential procedure based on mixed models depends on obtaining predictions for the random-effects. The most popular approach to predict a random-intercept b_i is based on minimizing the mean squared error of the predicted value (MCCULLOCH; SEARLE; NEUHAUS, 2008). In this case, one can easily verify that the predictor $\overline{b}_i$ that minimizes the overall mean squared error of prediction is given by

$$\overline{b}_i = \frac{\int_{\mathbb{R}} b_i \exp\{\hat{\ell}_i(\boldsymbol{y}_i, b_i)\} f(b_i)\,\mathrm{d}b_i}{\int_{\mathbb{R}} \exp\{\hat{\ell}_i(\boldsymbol{y}_i, b_i)\} f(b_i)\,\mathrm{d}b_i},$$

where $\hat{\ell}_i(\boldsymbol{y}_i, b_i) \equiv \ell_i(\hat{\boldsymbol{\theta}}; \boldsymbol{y}_i, b_i)$ is the estimated subject-specific log-likelihood function, with $\hat{\boldsymbol{\theta}}$ given by (6.12). There are few situations in which it is possible to obtain a closed-form solution for $\overline{b}_i$ (MCCULLOCH; NEUHAUS, 2011). In our case, the best predicted values for the random-intercepts of the proposed model can be obtained as

$$\overline{b}_{1i} = \frac{\overline{a}_{1i}\hat{\sigma}_1}{\overline{e}_i} \quad \text{and} \quad \overline{b}_{2i} = \frac{\hat{\sigma}_2(\overline{a}_{1i}\hat{\rho} + \overline{a}_{2i}\sqrt{1-\hat{\rho}^2})}{\overline{e}_i},$$

where $\overline{e}_i = \widehat{\mathrm{P}}_*(\boldsymbol{Y}_i = \boldsymbol{y}_i) \equiv \mathrm{P}_*(\boldsymbol{Y}_i = \boldsymbol{y}_i; \hat{\boldsymbol{\theta}})$ is the estimated marginal distribution of $\boldsymbol{Y}_i$ and

$$\begin{aligned}
\overline{a}_{ki} &= \int_{\mathbb{R}^2} u_{ki} \widehat{\mathrm{P}}_*(\boldsymbol{Y}_i = \boldsymbol{y}_i; \boldsymbol{u}_i) f(\boldsymbol{u}_i)\, \mathrm{d}\boldsymbol{u}_i \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} u_{ki} \exp\left\{ -\left[\frac{u_{1i}^2}{2} - \hat{\ell}_{1i}(\boldsymbol{y}_i, u_{1i}) + \frac{u_{2i}^2}{2} - \hat{\ell}_{2i}(\boldsymbol{y}_i, \boldsymbol{u}_i) \right] \right\} \mathrm{d}\boldsymbol{u}_i \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} u_{ki} \exp\{-\hat{r}_i(\boldsymbol{u}_i; \boldsymbol{y}_i)\} \mathrm{d}\boldsymbol{u}_i.
\end{aligned}$$

Unfortunately, we are not able to derive analytic predictors for b_{i1} and b_{i2} as the involved integrals are infeasible. Resorting to the Laplace method, one can approximate $\overline{a}_{ki}$ and $\overline{e}_i$ by

$$\overline{a}_{ki}^L \approx \frac{\tilde{u}_{ki} \exp\{-\hat{r}_i(\tilde{\boldsymbol{u}}_i; \boldsymbol{y}_i)\}}{\sqrt{\det \tilde{\boldsymbol{\mathcal{V}}}_i}} \quad \text{and} \quad \overline{e}_i^L \approx \frac{\exp\{-\hat{r}_i(\tilde{\boldsymbol{u}}_i; \boldsymbol{y}_i)\}}{\sqrt{\det \tilde{\boldsymbol{\mathcal{V}}}_i}},$$

where $\tilde{\boldsymbol{u}}_i = (\tilde{u}_{1i}, \tilde{u}_{2i})$ is the unique minimum of $\hat{r}_i$ and $\tilde{\boldsymbol{\mathcal{V}}}_i$ is the corresponding Hessian matrix evaluated at $\tilde{\boldsymbol{u}}_i$. Therefore, the random-intercepts b_{i1} and b_{i2} can be predicted by the approximations

$$\overline{b}_{1i}^L \approx \tilde{u}_{1i} \hat{\sigma}_1 \quad \text{and} \quad \overline{b}_{2i}^L \approx \hat{\sigma}_2 (\tilde{u}_{1i} \hat{\rho} + \tilde{u}_{2i} \sqrt{1 - \hat{\rho}^2}).$$

6.3.4 Posterior *predictive distribution*

In a Bayesian context, the *posterior* predictive distribution (ppd) is defined as the distribution of possible future (unobserved) values conditioned on the observed ones. Predicting probabilities for unobserved values within the framework of clustered samples depends on how these values are treated regarding the available data. We may assume that the out-of-sample values (i) belong to $n_1 \leqslant n$ observed subjects, or (ii) belong to n_2 unobserved subjects, with $n_2 \geqslant 1$ not necessarily equal to n.

Firstly, we suppose that $n_1 = 1$ in case (i). Thus, the partially observed response vector of the i-th subject holding out-of-sample data $(\boldsymbol{y}_i^u)$ can be expressed as $\boldsymbol{v}_i = (\boldsymbol{y}_i, \boldsymbol{y}_i^u)$. In this particular situation, the ppd is given by

$$\begin{aligned}
\mathrm{P}_\pi(\boldsymbol{Y}_i = \boldsymbol{v}_i; \boldsymbol{y}) = \int_{\mathbb{R}^d} \mathrm{P}_*(\boldsymbol{Y}_i = \boldsymbol{v}_i; \boldsymbol{\theta}) \pi(\boldsymbol{\theta}; \boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta} &= \mathbb{E}_{\boldsymbol{\theta}}[\mathrm{P}_*(\boldsymbol{Y}_i = \boldsymbol{v}_i; \boldsymbol{\theta}); \boldsymbol{y}] \\
&= \frac{1}{2\pi} \mathbb{E}_{\boldsymbol{\theta}} \left[\int_{\mathbb{R}^2} \exp\{-r_i(\boldsymbol{u}_i; \boldsymbol{v}_i, \boldsymbol{\theta})\} \mathrm{d}\boldsymbol{u}_i; \boldsymbol{y} \right].
\end{aligned}$$

Noticeably, the above equation has no closed-form as we are not able to compute the expectation over $\boldsymbol{\theta}$. Using the Laplace method for the inner term of the expectation, one can obtain an AMC estimator for the ppd as

$$\widehat{\mathrm{P}}_\pi^L(\boldsymbol{Y}_i = \boldsymbol{v}_i; \boldsymbol{y}) \approx \frac{1}{M} \sum_{t=1}^{M} \frac{\exp\{-r_i(\tilde{\boldsymbol{u}}_{it}; \boldsymbol{v}_i, \boldsymbol{\theta}_t)\}}{\sqrt{\det \tilde{\boldsymbol{\mathcal{V}}}_{it}}},$$

where $\tilde{\boldsymbol{u}}_{it}$ is the unique minimum of r_i at state t, and $\tilde{\boldsymbol{\mathcal{V}}}_{it}$ is the corresponding Hessian matrix evaluated at $\tilde{\boldsymbol{u}}_{it}$. For $n_1 > 1$, the full vector of partially observed responses can be expressed as

$\boldsymbol{v} = [(\boldsymbol{y}_1, \boldsymbol{y}_1^u), \ldots, (\boldsymbol{y}_{n_1}, \boldsymbol{y}_{n_1}^u)] = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_1})$. Hence, the corresponding ppd can be written as

$$\mathrm{P}_\pi(\boldsymbol{Y} = \boldsymbol{v}; \boldsymbol{y}) = \prod_{i=1}^{n_1} \mathrm{P}_\pi(\boldsymbol{Y}_i = \boldsymbol{v}_i; \boldsymbol{y}) = \frac{1}{(2\pi)^{n_1}} \prod_{i=1}^{n_1} \mathbb{E}_{\boldsymbol{\theta}} \left[\int_{\mathbb{R}^2} \exp\{-r_i(\boldsymbol{u}_i; \boldsymbol{v}_i, \boldsymbol{\theta})\} \mathrm{d}\boldsymbol{u}_i; \boldsymbol{y} \right],$$

where the first equality holds by the fact that outcomes for different subjects are assumed to be unconditionally independent. In this case, the AMC estimator for the ppd becomes

$$\widehat{\mathrm{P}}_\pi^L(\boldsymbol{Y} = \boldsymbol{v}; \boldsymbol{y}) \approx \frac{1}{M^{n_1}} \prod_{i=1}^{n_1} \sum_{t=1}^{M} \frac{\exp\{-r_i(\tilde{\boldsymbol{u}}_{it}; \boldsymbol{v}_i, \boldsymbol{\theta}_t)\}}{\sqrt{\det \tilde{\boldsymbol{\mathcal{V}}}_{it}}}, \tag{6.13}$$

which can also be used to estimate the ppd in case (ii), if one consider that $\boldsymbol{v}$ contains only unobserved responses, that is, $\boldsymbol{v} = (\boldsymbol{y}_1^u, \ldots, \boldsymbol{y}_{n_2}^u)$.

Algorithm 10 (see Appendix D, Section D.2) can be used to generate a single realization (y) from the ppd (in any of the presented cases) using the sequential-search scheme. To run it, one has to specify the sequence $\{\boldsymbol{\theta}_t\}_{t=1}^M$ obtained from Algorithm 9 after getting thinned steps. The procedure to generate a pseudo-random sample of size m consists of running the algorithm as often as necessary, say $m^* \geqslant m$ times. The sequential-search is a black-box algorithm and works with any computable probability vector. The main advantage of such a method is its great simplicity. On the other hand, it might be a bit slow when the occurrence probabilities for lower counts are close to zero, since the *while-loop* may have to be repeated very often in this case. We refer to Hörmann, Leydold and Derflinger (2013) for more details about this algorithm.

6.3.5 Influential points

Identifying influential observations is a crucial step in any statistical analysis. Usually, the presence of influential points impacts the inferential procedures and the subsequent conclusions considerably. In this way, this subsection is dedicated to present some case deletion Bayesian diagnostic measures that can be used to quantify the influence of observations from each subject in a given dataset.

The computation of divergence measures between *posterior* distributions is a useful way to quantify influence. According to Csiszár (1967), the φ-divergence measure between two densities f and g for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is defined by

$$d_\varphi = \int_{\boldsymbol{\Theta}} g(\boldsymbol{\theta}) \varphi \left[\frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})} \right] \mathrm{d}\boldsymbol{\theta},$$

where φ is a smooth convex, lower semicontinuous function such that $\varphi(1) = 0$. Some popular divergence measures can be obtained by choosing specific functions for φ. The well-known Kullback-Leibler (KL) divergence is obtained by considering $\varphi(z) = -\log(z)$. Cho *et al.* (2009) have considered such type of divergence measure in the context of survival models. A symmetric version of the KL divergence, the Jeffrey (J) divergence, can be obtained by specifying $\varphi(z) = (z-1)\log(z)$ and the variational divergence (L^1 norm) is obtained when $\varphi(z) = 0.50|z-1|$.

These divergence measures were considered by Garay *et al.* (2015) for case deletion of influential points in the context of censored linear regression models with scaled mixtures. In addition, the Chi-Square (CS) divergence is obtained by considering $\varphi(z) = (z-1)^2$ and the Hellinger (H) distance arises when $\varphi(z) = 0.50(\sqrt{z}-1)^2$. We refer to Peng and Dey (1995) for a detailed study on several types of φ-divergence.

Let $g(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta};\boldsymbol{y}_i)$ be the joint *posterior* distribution of $\boldsymbol{\theta}$ based only on the information provided by the i-th subject and let $f(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta};\boldsymbol{y}_i^j)$, where $\boldsymbol{y}_i^j = (y_{i1},\ldots,y_{ij-1},y_{ij+1},\ldots,y_{m_i})$ is the response vector of the i-th subject without the j-th observation. After some algebra (see Cho *et al.* (2009) for the KL divergence case), one can verify that the φ-divergence corresponds to

$$d_\varphi = \mathbb{E}_{\boldsymbol{\theta}}\left\{\varphi\left[\frac{\mathbb{E}_{\boldsymbol{\theta}}^{-1}\left[\mathrm{P}_*^{-1}\left(Y_{ij}=y_{ij};\boldsymbol{\theta},\boldsymbol{u}_i\right);\boldsymbol{y}_i\right]}{\mathrm{P}_*\left(Y_{ij}=y_{ij};\boldsymbol{\theta},\boldsymbol{u}_i\right)}\right];\boldsymbol{y}_i\right\},$$

where $\mathbb{E}_{\boldsymbol{\theta}}^{-1}[\mathrm{P}_*^{-1}(Y_{ij}=y_{ij};\boldsymbol{\theta},\boldsymbol{u}_i);\boldsymbol{y}_i]$ is the conditional predictive ordinate (CPO) statistic (GEISSER; EDDY, 1979) for the j-th observation of the i-th subject. Here, we are also not able to compute the inner expectation over $\boldsymbol{\theta}$ analytically and so, an AMC estimator for the CPO_{ij} is given by

$$\widehat{\mathrm{CPO}}_{ij}^{L} \approx \left[\frac{1}{M}\sum_{t=1}^{M}\mathrm{P}_*^{-1}\left(Y_{ij}=y_{ij};\boldsymbol{\theta}_t,\overline{\boldsymbol{u}}_i^L\right)\right]^{-1}, \tag{6.14}$$

where $\overline{\boldsymbol{u}}_i^L$ is the approximated predicted value of $\boldsymbol{u}_i$. According to Congdon (2005), the harmonic mean estimator (6.14) is stable when most of the individual log-likelihood values exceed -10. Using the estimated CPO, one can approximate the local influence of a particular y_{ij} on the joint *posterior* distribution (6.11) as

$$\hat{d}_\varphi^L \approx \frac{1}{M}\sum_{t=1}^{M}\varphi\left[\frac{\widehat{\mathrm{CPO}}_{ij}^{L}}{\mathrm{P}_*\left(Y_{ij}=y_{ij};\boldsymbol{\theta}_t,\overline{\boldsymbol{u}}_i^L\right)}\right].$$

One can notice that, if $\pi(\boldsymbol{\theta};\boldsymbol{y}_i^j) = \pi(\boldsymbol{\theta};\boldsymbol{y}_i)$, then there is no divergence caused by observation y_{ij}. In practice, however, it may not be elementary to define a threshold value for the divergence in order to decide about the magnitude of the influence (WEISS, 1996). A measure of calibration for the KL divergence was proposed by McCulloch (1989). The idea is based on the typical toy binary example of tossing a coin once and observing its upper face. This experiment can be described by $\mathrm{P}(Y=y;\zeta) = \zeta^y(1-\zeta)^{1-y}$, $y\in\{0,1\}$, where $\zeta\in[0,1]$ is the probability of success. Regardless of what success means, if the coin is unbiased, then $\mathrm{P}(Y=y;\zeta) = 0.50$. Thus, the φ-divergence between a (possibly) biased and an unbiased coin is given by

$$d_\varphi(\zeta) = \frac{\varphi(2\zeta)+\varphi[2(1-\zeta)]}{2},$$

from which one can conclude that the divergence between two *posteriors* distributions can be associated with the biasedness of a coin (PENG; DEY, 1995). By analogy, this implies that predict unobserved responses using $\pi(\boldsymbol{\theta};\boldsymbol{y}_i^j)$ instead of $\pi(\boldsymbol{\theta};\boldsymbol{y}_i)$ is equivalent to describe an

unobserved event as having probability ζ_{ij}, when the correct probability is 0.50. Considering some specific choices for φ, in Table 127 (see Appendix D, Section D.1) we present AMC estimators that can be used to compute the local influence of each y_{ij}. Besides, we also present the expression of $d_\varphi(\zeta)$ for each φ. For ease of notation, we assume $f_{ij}^t = \mathrm{P}(Y_{ij} = y_{ij}; \boldsymbol{\theta}_t, \overline{\boldsymbol{u}}_i^L)$.

The function $d_\varphi(\zeta)$ is symmetric about 0.50 and increases as ζ moves away from 0.50. In addition, $\inf_{\zeta\in(0,1)} d_\varphi(\zeta) = 0$, which is attained at $\zeta = 0.50$ since $d_\varphi(0.50) = \varphi(1) = 0$. Therefore, a general measure of calibration based on the φ-divergence can be obtained by solving

$$2d_\varphi(\zeta) - \varphi(2\zeta) - \varphi[2(1-\zeta)] = 0.$$

An estimator for the calibration measure (ζ_φ) associated with each φ-divergence type is also presented in Table 127. Clearly, depending on the form of φ, such an equation may not have a closed-form, which is the case of the J divergence. Besides, one can notice that $\zeta_{ij} \in [0.50, 1]$ and so, for $\zeta_{ij} \gg 0.50$, the j-th observation of the i-th subject may be considered an influential point. For example, if $\zeta_{ij} > 0.80$ is considered a significative bias, then y_{ij} will be classified as influential if $\hat{d}_{ij} > 0.223$ $(d_\varphi(0.80) \approx 0.223)$ under the KL divergence or yet if $\hat{d}_{ij} > 0.051$ $(d_\varphi(0.80) \approx 0.051)$ under the H divergence.

6.3.6 Model comparison and adequacy

There are several methods for Bayesian model selection that are useful to compare competing models. The most popular method is the Deviance Information Criterion (DIC), which was proposed to work simultaneously as a measure of fit and complexity of the model. The DIC criterion is defined as

$$\mathrm{DIC} = \mathbb{E}_{\boldsymbol{\theta}}[\mathrm{D}(\boldsymbol{\theta})] + \rho_{\mathrm{D}} = \underline{\mathrm{D}}(\boldsymbol{\theta}) + \rho_{\mathrm{D}},$$

where $\mathrm{D}(\boldsymbol{\theta}) = -2\ell(\boldsymbol{\theta}; \boldsymbol{y})$ is the *deviance* function and $\rho_{\mathrm{D}} = \underline{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\hat{\boldsymbol{\theta}})$ is the effective number of model parameters, with $\hat{\boldsymbol{\theta}}$ given by (6.12). A negative value for ρ_{D} may suggest that the log-likelihood function is non-concave, the *prior* distribution is misspecified, or the *posterior* expected value is not a good estimator for $\boldsymbol{\theta}$. On the other hand, when $\rho_{\mathrm{D}} \gg d$, then there is an indication of overfitting with estimate $\hat{\boldsymbol{\theta}}$.

For known reasons, we are not able to compute the expectation of $\mathrm{D}(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ analytically. Considering the Laplace approximation (6.10) for the log-likelihood function of $\boldsymbol{\theta}$, an AMC estimator for such a measure is given by

$$\widehat{\underline{\mathrm{D}}}^L(\boldsymbol{\theta}) \approx -\frac{2}{M}\sum_{t=1}^{M} \ell^L(\boldsymbol{\theta}_t; \boldsymbol{y}),$$

and so the DIC can be estimated by

$$\widehat{\mathrm{DIC}}^L \approx 2\widehat{\underline{\mathrm{D}}}^L(\boldsymbol{\theta}) - \mathrm{D}(\hat{\boldsymbol{\theta}}).$$

The Expected Akaike (EAIC) and the Expected Bayesian (EBIC) information criteria can also be used when comparing Bayesian models (CARLIN; LOUIS, 2001; BROOKS, 2002). Using the approximation for the expected value of $\mathrm{D}(\boldsymbol{\theta})$, these measures can be estimated by

$$\widehat{\mathrm{EAIC}}^{L} \approx \widehat{\underline{\mathrm{D}}}^{L}(\boldsymbol{\theta}) + 2d \quad \text{and} \quad \widehat{\mathrm{EBIC}}^{L} \approx \widehat{\underline{\mathrm{D}}}^{L}(\boldsymbol{\theta}) + d\log(m).$$

Another widely used criterion is derived from the CPO statistic, which is based on the cross-validation criterion to compare models. For the j-th observation of the i-th subject, the CPO can be estimated through Equation (6.14). A summary statistic of the estimated CPO's is the Log-Marginal Pseudo-Likelihood (LMPL) given by the sum of the logarithms of $\widehat{\mathrm{CPO}}_{ij}$'s. Regarding model comparison, we have that the lower the values of DIC, EAIC, EBIC, and NLMPL (negative LMPL), the better the fit.

The comparison procedure between two competing models can be formulated as a hypothesis test. In a Bayesian framework, one can quantify the strength of evidence in favor of one model among its competitors by computing the Bayes Factor. This measure is defined as the ratio between the marginal likelihood functions of competing models and can be well-approximated using the CPO statistic when marginalization is unfeasible (GELFAND; DEY, 1994). Any approximation for the Bayes Factor is denominated as Pseudo-Bayes Factor (PBF). For two competing models, say M_0 and M_1 (null and alternative hypothesis, respectively), one can evaluate the relative plausibility of the available data under M_1 versus M_0 using the Laplace approximation

$$\widehat{\mathrm{PBF}}^{L} \approx \frac{\prod_{i=1}^{n}\prod_{j=1}^{m_i}\widehat{\mathrm{CPO}}_{ij}^{L}(M_1)}{\prod_{i=1}^{n}\prod_{j=1}^{m_i}\widehat{\mathrm{CPO}}_{ij}^{L}(M_0)},$$

which is an adapted AMC estimator for the PBF, designed to fit our case. The interpretation of this ratio is straightforward. If $\widehat{\mathrm{PBF}} > 1$, then there is evidence that M_1 is more plausible (more strongly supported by the data) than M_0. On the other hand, when $\widehat{\mathrm{PBF}}$ approaches to zero, then there is strong evidence against M_1 in favor of M_0. A widely used table with scales for interpretation of PBF is provided by Kass and Raftery (1995). One of the main features of the PBF is that, unlike the Bayes Factor, it can be used even when improper *prior* distributions are adopted.

In addition to comparing, researchers are often interested in verifying the adequacy of the fitted models. Congdon (2005) suggests that the CPO can also be used as a metric to assess model fit. The procedure consists of obtaining the scaled CPOs (dividing each one by the maximum value) and then check for the existence of values lower than 0.01. If the majority of SCPOs (scaled CPOs) is above this value, the model fit can be considered adequate. Another effective way to evaluate model adequacy is based on the use of measures derived from the ppd. For instance, if any observation is extremely unlikely relative to the ppd, the adequacy of the obtained fit might be questionable. A widespread discrepancy measure between model and data was proposed by Gelman *et al.* (2004). In our case, we need a slightly adapted version of such a

measure, which is given by

$$\mathrm{T}(\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{u}) = -2\sum_{i=1}^{n}\sum_{j=1}^{m_i}\log\left[\mathrm{P}_*\big(Y_{ij}=y_{ij};\boldsymbol{\theta},\boldsymbol{u}_i\big)\right].$$

The Bayesian p-value (*posterior* predictive p-value), proposed by Rubin (1984), is defined as

$$p_{\mathrm{B}} = \mathrm{P}\left[\mathrm{T}(\boldsymbol{y}_r,\boldsymbol{\theta},\boldsymbol{u}) \geqslant \mathrm{T}(\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{u})\,;\boldsymbol{y}\right],$$

where $\boldsymbol{y}_r$ denotes the response vector that might have been observed if the conditions generating $\boldsymbol{y}$ were reproduced. This predictive measure can be empirically estimated as the relative number of times that $\mathrm{T}(\boldsymbol{y}_r,\hat{\boldsymbol{\theta}},\overline{\boldsymbol{u}}^L)$ exceeds $\mathrm{T}(\boldsymbol{y},\hat{\boldsymbol{\theta}},\overline{\boldsymbol{u}}^L)$ out of B simulations, where $\overline{\boldsymbol{u}}^L$ is the approximated predicted value of $\boldsymbol{u}=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)$. Each replica $\boldsymbol{y}_r$ is a pseudo-random sample from the approximate ppd (6.13), which can be obtained by running Algorithm 10 at least m times. In general, the model fit becomes suspect if the discrepancy is of practical relevance, and the associated Bayesian p-value is close either to 0 or 1 (GELMAN *et al.*, 2004). A large (small) value of p_{B}, say greater than 0.95 (lower than 0.05), indicates model misspecification (lack-of-fit), that is, the observed behavior would be unlikely to be seen if we replicate the response vector using the fitted model.

6.3.7 Residual analysis

The residual analysis plays an essential role in the task of validating the results obtained from a regression model. In general, residual metrics are responsible for indicating departures from the underlying model assumptions by quantifying the portion of data variability that was not explained by the fitted model. Assessing the adequacy of a regression model using residual metrics is a common practice nowadays due to the availability of statistical packages providing diagnostic tools for well-established models. However, deriving appropriate residuals is not always an easy task for non-normal models that accommodate overdispersion and mixed-effects. In this way, we will consider here a popular residual metric proposed by Dunn and Smyth (1996), the randomized quantile residuals (RQRs), which can be straightforwardly used in our context to assess the appropriateness of the proposed model when fitted to real data.

For obvious reasons, we focus on the definition of RQRs for discrete random variables, particularly for clustered data. In this case, the RQR associated to the j-th observation of the i-th subject is defined as $r_{ij} = \Phi^{-1}(u_{ij})$, where Φ denotes the cdf of the standard Normal distribution and u_{ij} is a Uniform random variable defined on $(a_{ij},b_{ij}]$, with $a_{ij} = \lim_{y\uparrow y_{ij}}\mathrm{F}(y_{ij})$ and $b_{ij} = \mathrm{F}(y_{ij})$, where $\mathrm{F}(y_{ij})$ is the cdf of the current model. In our case, we may obtain an AMC estimator for the RQR by approximating r_{ij} as

$$\hat{r}_{ij}^{L} \approx \Phi^{-1}(u_{ij}),$$

with $u_{ij} \sim \mathcal{U}(\lim_{y \uparrow y_{ij}} \hat{\mathrm{F}}_*(y_{ij}), \hat{\mathrm{F}}_*(y_{ij})]$. Here, $\hat{\mathrm{F}}_*(y_{ij}) \equiv \mathrm{F}_*(y_{ij}; \hat{\mu}_{ij}, \hat{\omega}_{ij})$ is an estimate for the probability of $Y_{ij} \leqslant y_{ij}$ using cdf (6.2), where $\hat{\mu}_{ij}$ and $\hat{\omega}_{ij}$ depend on the fitted model as

$$\hat{\mu}_{ij} = \hat{\mu}_{ij}^L(\boldsymbol{y}) \approx \log(\boldsymbol{x}_{1ij}^\intercal \hat{\boldsymbol{\beta}}_1 + \tilde{u}_{1i}^L \hat{\sigma}_1),$$

and, for a given link function g_2,

$$\hat{\omega}_{ij} = \hat{\omega}_{ij}^L(\boldsymbol{y}) \approx g_2^{-1}\left[\boldsymbol{x}_{2ij}^\intercal \hat{\boldsymbol{\beta}}_2 + \hat{\sigma}_2\left(\tilde{u}_{1i}^L \hat{\rho} + u_{2i}^L \sqrt{1-\hat{\rho}^2}\right)\right].$$

The main assumption for this metric is that $\hat{r}_{ij} \sim \mathcal{N}(0,1)$ must hold, whichever the variability degree of $\hat{\mu}_{ij}$ and $\hat{\omega}_{ij}$. In this case, after model fitting, one has to evaluate if these residuals are normally distributed around zero, which can be made by applying standard adherence tests and by using graphical techniques as histograms and Half-Normal probability plots. An excellent alternative for checking whether RQRs are consistent with the fitted model is the inclusion of simulated envelopes in their Half-Normal plot. Thus, if a significative subset of estimated residuals falls outside the envelope bands, then the adequacy of the fitted model must be questioned, and further investigation on the corresponding observations are necessary. An algorithm for obtaining simulated envelopes for a Half-Normal plot is provided by Moral, Hinde and Demétrio (2017).

6.4 Simulation study

The empirical properties of an estimator can be accessed through Monte Carlo simulations. In this way, we have performed an intensive simulation study aiming to evaluate the performance of the proposed Bayesian approach in some specific situations. The simulation process was carry out by generating 500 pseudo-random samples of sizes $n = 25, 50,$ and 100 from the $\mathcal{ZMPL}$ distribution under the regression framework presented in Section 6.2, with $\mu_{ij} = \exp\{\beta_{10} + \beta_{11} x_{1ij} + b_{1i}\}$ and $\omega_{ij} = [1 + \exp\{-(\beta_{20} + \beta_{21} x_{2ij} + b_{2i})\}]^{-1}$ (*logit* link function). The product-Normal parameterization (6.7) was considered for the random-effects. For each subject, we have generated the same number of observations $(m_i = 4)$ as to simulate clustered data (with individual heterogeneity) in a balanced experiment. Also, we have assumed that $x_{1ij} = x_{2ij} = x_{ij}$ is a within-subjects fixed covariate $(x_{ij} = x_{ij'}, j \neq j')$, with $x_{i\cdot} \sim \mathcal{U}(0,1)$.

In order to address both zero-inflated and zero-deflated cases in the simulations, we have considered six scenarios for each kind of zero modification. Table 40 presents the actual parameter values that were considered in our study. For the zero-inflated (zero-deflated) case, the samples were generated from the $\mathcal{ZMPL}$ distribution by considering that $p_{ij} \in (0,1)$ $(p_{ij} \in [1, \mathrm{P}^{-1}(Y_{ij} > 0; \mu_{ij})])$ for all pairs (i,j). Here, the parameters were chosen by taking into account that zero-inflated (zero-deflated) samples have, naturally, the proportion of zeros greater (lower) than expected under an ordinary count distribution and, therefore, the artificial datasets were generated with expected value far from zero (close to zero).

Table 40 – Actual parameter values for simulation of zero-modified artificial datasets.

Case	Scenario	β_{10}	β_{11}	β_{20}	β_{21}	σ_1	σ_2	ρ
Inflation	1	1.50	2.50	1.00	−1.50	0.50	1.50	−0.50
	2							0.00
	3							0.50
	4					1.50	0.50	−0.50
	5							0.00
	6							0.50
Deflation	1	−2.00	−1.00	1.50	−2.00	0.50	1.50	−0.50
	2							0.00
	3							0.50
	4					1.50	0.50	−0.50
	5							0.00
	6							0.50

Source: Elaborated by the author.

Algorithm 11 (see Appendix D, Section D.2) can be used to generate a pseudo-random response vector for a single subject (under the proposed regression framework with a sole covariate) using the inverse transform method. The extension for the use of more covariates is straightforward. To run it, one has to specify the vector of model parameters to be used in the Monte Carlo simulations. The procedure to generate a full response vector $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ consists of running the algorithm n times. To apply the inverse transform method for sampling, we have to obtain the quantile function of the $\mathcal{ZMPL}$ distribution, that is, the inverse of cdf (6.2). However, since a hurdle model is defined by the mixture of a zero-truncated distribution and a binary probability model, we may use the quantile function of the $\mathcal{ZTPL}$ distribution to generate positive observations from (6.1), and these values become zero with probability $1-\omega_{ij}$. This quantile function was derived by Jodrá (2010) and is given by

$$Q(u,\theta) = \left\lceil -\frac{\theta^2+3\theta+1}{\theta} - \frac{1}{\log(\theta+1)} W_{-1}\left(\frac{(\theta^2+3\theta+1)\log(\theta+1)}{\theta(\theta+1)^{(\theta^2+3\theta+1)\theta^{-1}}}(u-1) \right) \right\rceil,$$

for $\theta \in \mathbb{R}_+$ and $u \in (0,1)$. Here, $\lceil a \rceil$ is the ceiling function of $a \in \mathbb{R}$, that is, $\lceil a \rceil = \inf_{a \in \mathbb{R}} \{k \in \mathbb{Z} : k \geqslant a\}$ and W_{-1} denotes the negative branch of the LambertW function (see Corless *et al.* (1996) for further details). Since pmf (6.1) is parameterized in terms of the expected value (μ) of the untruncated $\mathcal{PL}$ distribution, we should express parameter θ as

$$\theta = -\frac{(\mu-1) - \sqrt{(\mu-1)^2+8\mu}}{2\mu}.$$

To apply the proposed Bayesian methodology to each scenario, we have considered the AM algorithm to generate pseudo-random samples from the approximate *posterior* distribution (6.11). For the proposal distribution, the mixture parameter was chosen as $\psi = 0.95$. A total

of $N = 100{,}000$ values were generated for $\boldsymbol{\theta}$. After running Algorithm 9, the first 20,000 samples were discarded as the burn-in period, and then one out every 40 generated values were kept, resulting in sequences of size $M = 2{,}000$ for each element of $\boldsymbol{\theta}$. Using Geweke's z-score diagnostic, the stationarity of the chains was revealed. When running the simulations, the acceptance rates varied between 15 and 25%.

Since $\boldsymbol{\theta}$ is a reparameterization of the proposed model, we may obtain pseudo-random samples of the original parameters by applying the inverse function (respect to those defining the transformations) on the generated chains. For example, given a sequence $\{\xi_{1t}\}_{t=1}^{M}$ of ξ_1, a chain for σ_1 is obtained as $\{e^{\xi_{1t}}\}_{t=1}^{M}$. The adopted Bayesian point estimator for the original parameters is the same as provided by Equation (6.12). The performance of such an estimator was studied by assessing its bias (B), its mean squared error (MSE), and its mean absolute percentage error (MAPE). Using samples of the original parameters, one may obtain AMC estimators for these measures as

$$\widehat{\mathrm{B}}_{\hat{\gamma}}^{L} \approx \frac{1}{500}\sum_{j=1}^{500}\left(\hat{\gamma}_j - \gamma\right), \quad \widehat{\mathrm{MSE}}_{\hat{\gamma}}^{L} \approx \frac{1}{500}\sum_{j=1}^{500}\left(\hat{\gamma}_j - \gamma\right)^2 \quad \text{and} \quad \widehat{\mathrm{MAPE}}_{\hat{\gamma}}^{L} \approx \frac{1}{500}\sum_{j=1}^{500}\left|\frac{\hat{\gamma}_j - \gamma}{\gamma}\right|,$$

where $\gamma = \beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_1, \sigma_2$ or ρ, and $\{\hat{\gamma}_j\}_{j=1}^{500}$ is the sequence of *posterior* means based on chains of size 2,000. Noticeably, the MAPE cannot be computed for $\hat{\rho}$ when its actual value is zero. The variance of $\hat{\gamma}$ was estimated as the difference between MSE and the square of the bias. Besides, the coverage probability (CP) of the 95% highest *posterior* density intervals (HPDIs) was estimated by the approximation

$$\widehat{\mathrm{CP}}_{\gamma}^{L} \approx \frac{1}{500}\sum_{j=1}^{500}\delta_j(\gamma),$$

where $\delta_j(\gamma)$ assumes 1 if the j-th HPDI contains the actual value γ and zero otherwise. We have also estimated the below (BNCP) and above (ANCP) noncoverage probabilities of the HPDIs. These measures are computed analogously to CP. The BNCP and ANCP may be useful to determine asymmetrical behaviors as they provide the probabilities of finding the actual value of γ on the tails of its *posterior* distribution.

Due to the massive amount of outputs, the obtained results were made available in Appendix D (Section D.3). We have noticed in our study that the overall accuracy of the Bayesian estimators improved with increasing sample sizes since the estimated MSEs and MAPEs have decreased as n increased. However, the square root of the ratio between the MSEs and variances slowly approached to 1 due to the slight instability of the estimated biases, which can be explained by the complexity of the proposed model. On the other hand, we have observed that the estimated CP of the HPDIs was converging to 95% in most cases, but, in some zero-deflated scenarios, some of the CPs for the random-effects parameters reached 100% when the actual value of ρ was either -0.50 or zero. Additionally, one can notice that, in the zero-inflated cases, the approximate *posterior* densities of β_{10} and β_{11} are more asymmetric since the low

amount of positive observations on the generated samples led to higher uncertainty about the *posterior* behavior of these coefficients.

Considering the predefined scenarios, we conclude that our simulation study provides favorable indications about the suitability of the adopted Bayesian methodology to estimate the parameters of the proposed model. We believe that in a similar procedure with a different set of actual values, the overall behavior of the estimators should resemble the results that we have described here. Besides, the adopted methodology would also be reliable if one or more than one covariates (possibly of other nature) were included in the linear predictors of μ_{ij} and ω_{ij}.

6.5 Real data application

In this section, we present the analysis of a real dataset using the proposed model. For comparison purposes, we have also fitted mixed-effects models based on the $\mathcal{ZMP}$ and zero-modified Negative Binomial ($\mathcal{ZMNB}$) distributions (parameterized by μ and ω as well). The ordinary $\mathcal{P}$, $\mathcal{PL}$, and Negative Binomial ($\mathcal{NB}$) models were also considered in our analyses. In the example, the relationship between effects and parameters were established by a log-linear structure for μ and a logistic regression model for ω, as described in Section 6.2. From the general formulation, three distinct approaches were treated in the modeling procedure. Firstly, we have assumed that μ and ω share the same random-effect (Case 1). The second and third approaches only differ by the fact that the random-effects were assumed to be uncorrelated (Case 2) and correlated (Case 3).

To fit the models in all cases, we have adopted the same Bayesian procedures detailed in the previous section. For the dispersion parameter (ϕ) of the $\mathcal{NB}$ and $\mathcal{ZMNB}$ models, we have assigned a noninformative Inverse Half-Cauchy *prior* distribution with $\alpha = 25$, that is, $\pi(\phi) \propto (625\phi^2 + 1)^{-1}$. Finally, using the Laplace approximations presented in Section 6.3, we were able to summarize the *posterior* distributions, to compare and validate the fitted models, and to make appropriate inferences based on the available data.

Table 41 – Descriptive summary of grooming counting data per rat.

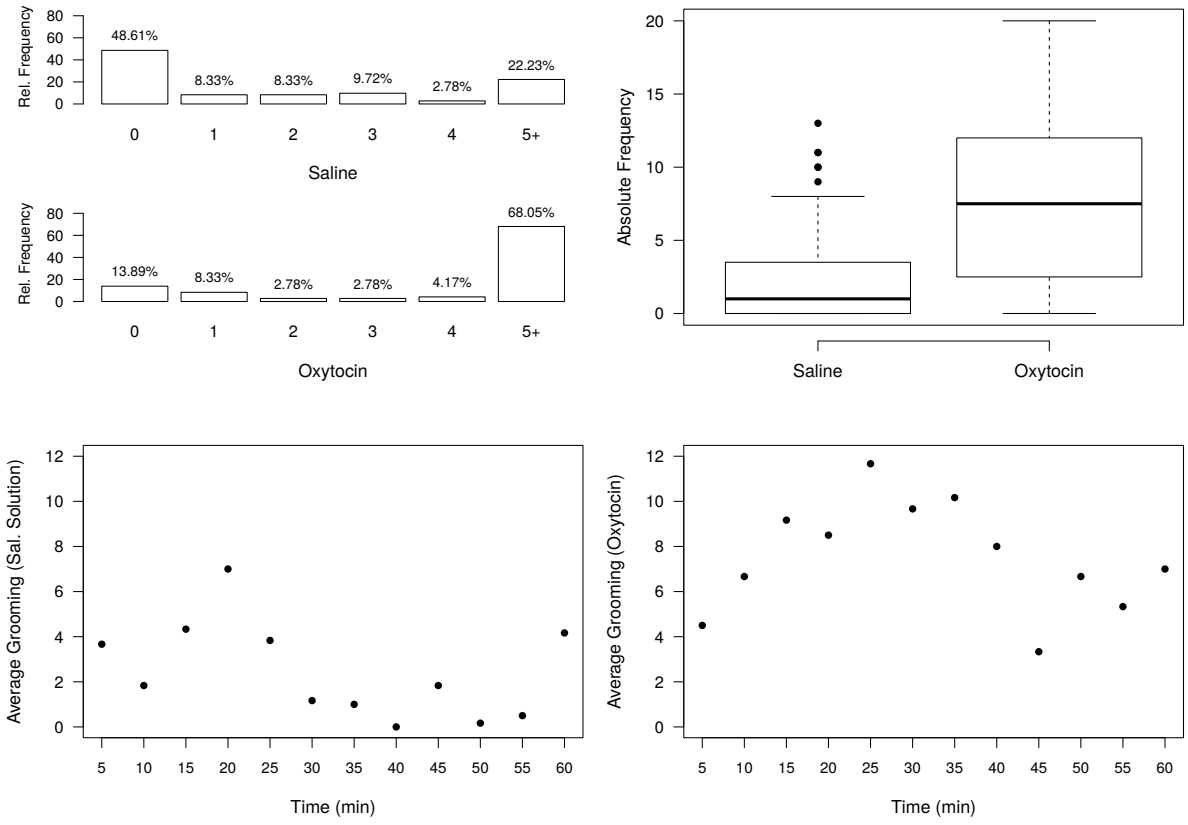
Rat	m_i	$\bar{y}_i$	s_i	$s_i^2/\bar{y}_i$	m_i^0/m_i	$y_i^{\max}$
1	24	4.125	5.519	7.385	0.250	20
2	24	6.667	6.267	5.891	0.250	19
3	24	3.750	3.937	4.133	0.333	13
4	24	6.042	5.775	5.519	0.292	16
5	24	4.458	4.995	5.597	0.417	16
6	24	5.000	4.782	5.574	0.333	14
Overall	144	5.007	5.278	5.563	0.312	20

Source: Elaborated by the author.

We have considered data from a neurophysiologic experiment realized in the Neurophysiology and Neuroethology Experimental Laboratory (Medical School of Ribeirão Preto, University of São Paulo, Brazil) with six male rats of the War species. This experiment was designed to analyze the number of grooming movements (grooming counting) of the rats at different times after receiving two treatments: saline solution (placebo) and oxytocin. The treatments were administered in the following order: first, the animal received saline and then, oxytocin. After the application of each treatment, the grooming countings were measured 12 times every five minutes, totalizing 24 observations per rat.

The full dataset was made available by Achcar, Coelho-Barros and Martinez (2008) and is composed by $m = 144$ observations from a discrete random variable Y_{ij} denoting the j-th grooming counting registered for the i-th rat $(i = 1,\ldots,6)$. The described experiment is fully balanced as the number of observation are the same across rats (24). Using this data, we aimed to analyze if the administration of oxytocin affects the occurrence and persistence of grooming movements, and also to investigate the behavior of these components over time. Subject-specific random-effects were considered as an attempt to overlap the variation among the repeated measures of each rat.

Figure 26 – Descriptive summary (per treatment) of grooming counting data.



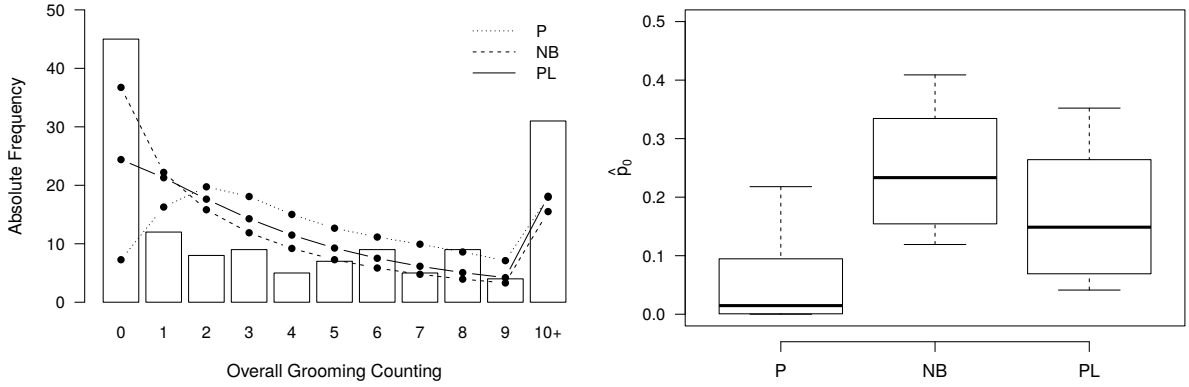
Source: Elaborated by the author.

A descriptive summary of the number of grooming movements is presented in Table 41.

A preliminary analysis suggests that the response variable is hugely overdispersed (overall and within-rat), with average grooming counting higher than 3.70 for all rats. However, the average number of movements was lower when the rats were under the effect of saline solution (2.46). The overall number of zeros in the sample is $m^0 = 45$. The maximum number of movements was observed for Rat 1 ($y_1^{\max} = 20$), followed by Rat 2, with $y_1^{\max} = 19$. Noticeably, Rat 5 was less responsive to both treatments, once presenting no grooming movements in approximately 42% of the time ($m_5^0 = 10$). In general, all rats have responded with at least one movement in more than 58% of the time.

Figure 26 presents the frequency distribution of the number of movements per treatment (upper-panels) and the average grooming counting over time (lower-panels). These illustrations are indicating that the behavior of the response variable differs considerably between treatments, and also that the average of grooming movements are not constant over time. Particularly, the occurrence was higher at around 20-35min and have lowered in the last quarter of each treatment. Besides, when oxytocin was administered, the average persistence was higher (7.56), but only Rats 2 and 6 have responded with at least one movement at all times. Such a fact suggests that exits an individual heterogeneity between the rats, which can also be noticed by analyzing the variability measures presented in Table 41.

Figure 27 – *Posterior* descriptive summary for the $\mathcal{P}$, $\mathcal{NB}$, and $\mathcal{PL}$ fitted models.



Source: Elaborated by the author.

The frequency distribution presented in Figure 26 (upper-panels) does not provide much evidence about the nature of a possible zero modification in the number of grooming movements. However, since the within-rat averages are relatively far from zero and the indices of dispersion are greater than 1, it is plausible to believe that there is a portion of zero-valued observations that exceed what would be expected by an ordinary discrete distribution. Therefore, we have searched for empirical evidence of zero inflation by fitting the $\mathcal{P}$, $\mathcal{NB}$, and $\mathcal{PL}$ distributions to the available data. Here, the expected value of these distributions was modeled by the following log-linear structure

$$\log\left(\mu_{ij}\right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x^*_{2ij} + b_i,$$

where $b_i \sim \mathcal{N}(0, \sigma_b^2)$ is a rat-specific random-intercept, x_{1ij} is the administered treatment (0: saline solution (baseline); 1: oxytocin), and $x_{2ij}^* \in (0,1)$ is the scaled logarithm of x_{2ij} (SLT), with x_{2ij} denoting the time (in minutes).

The obtained results are summarized in Figure 27. The estimated frequency distribution (left-panel) is suggesting that the sample is zero-inflated. Noticeably, the $\mathcal{P}$ model does not provide a suitable fit as it cannot handle the overall overdispersion of the response variable. Under this model, the expected number of zeros is much lower than the real one (45) as the probability of non-occurrence of grooming movements (p_0) was underestimated. Besides, although accommodating extra-$\mathcal{P}$ variability, the $\mathcal{NB}$ and $\mathcal{PL}$ models are also not able to handle heterogeneity due to zero modification. Notwithstanding, we have also performed a formal test, as described in Subsection 6.2.1. In this case, the *posterior* estimate for parameter ν from the $\mathcal{ZAPL}$ model was -0.989 (-1.354, -0.632), which indicates the presence of an excessive amount of zeros regarding what would be expected under the $\mathcal{PL}$ distribution.

Table 42 – Comparison criteria and adequacy measures for the fitted models.

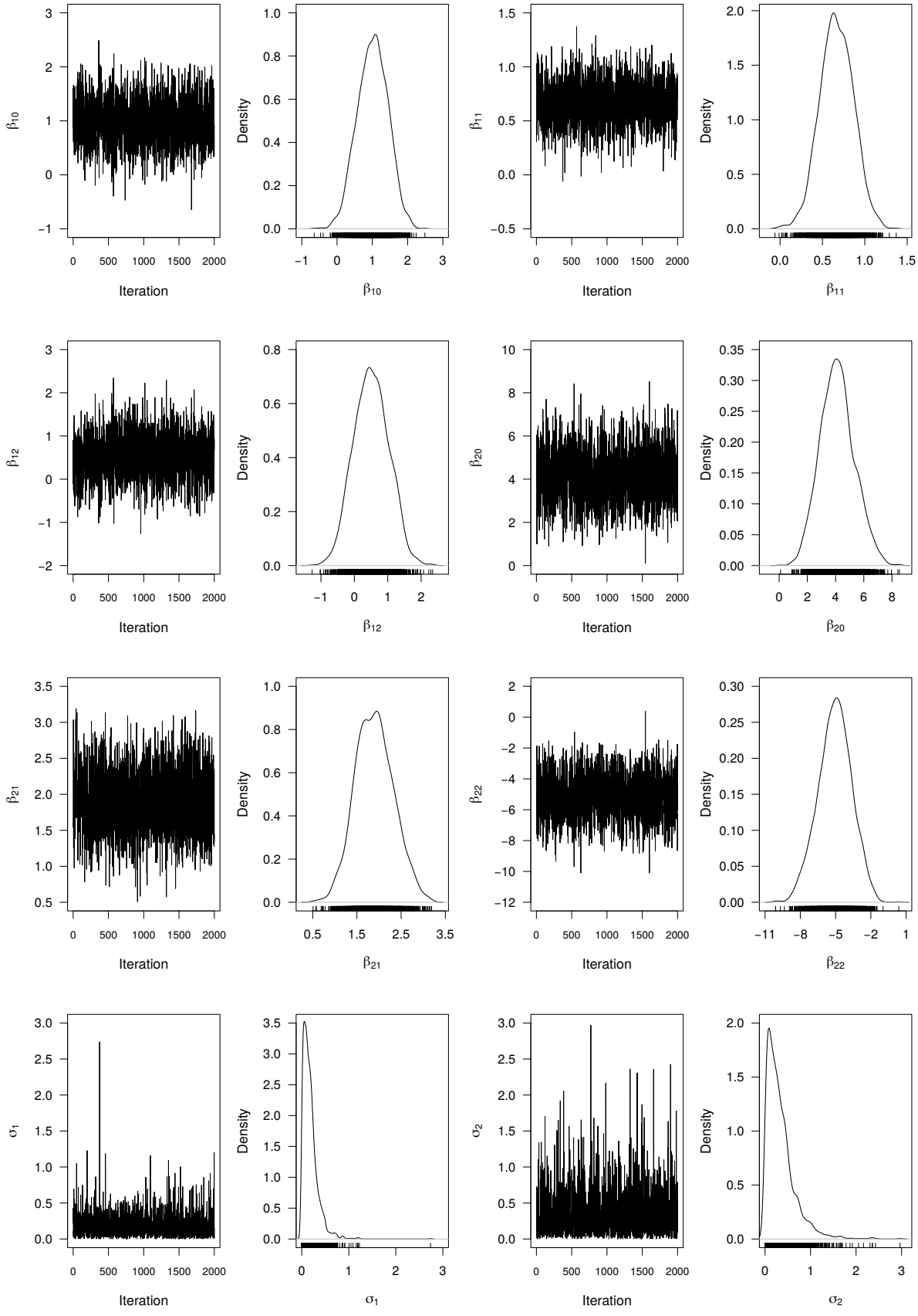
Model	Case	ρ_{D} (d)	DIC	EAIC	EBIC	NLMPL	% SCPO > 0.01	p_{B}
$\mathcal{ZMP}$	1	6.91 (7)	782.93	790.02	810.81	384.98	89.60	0.000
	2	8.12 (8)	785.44	793.32	817.08	383.92	90.30	0.000
	3	8.04 (9)	785.62	795.58	822.31	384.18	90.30	0.000
$\mathcal{ZMNB}$	1	8.07 (8)	712.87	720.80	744.56	353.67	96.50	0.454
	2	9.16 (9)	715.58	724.42	751.14	352.54	96.50	0.474
	3	9.25 (10)	715.89	726.64	756.34	352.74	96.50	0.474
$\mathcal{ZMPL}$	1	7.19 (7)	712.14	718.96	739.75	353.19	97.20	0.508
	2	8.26 (8)	714.76	722.51	746.27	352.11	97.20	0.514
	3	8.37 (9)	715.16	724.79	751.52	352.36	96.50	0.530

Source: Elaborated by the author.

As we have noticed, the response variable combines three different aspects: overdispersion, individual heterogeneity, and zero modification. In this case, a natural attempt to better overlapping all these sources of variation is to consider zero-modified mixed models. Particularly, for the available dataset, we have fitted an occurrence-persistence model based on the $\mathcal{ZMPL}$ distribution. In this framework, the probability of a grooming movement being practiced by the i-th rat is given by ω_{ij} and the average persistence of grooming counting is given by μ_{ij}. To describe these parameters, we have considered the following structures

$$\log\left(\mu_{ij}\right) = \beta_{10} + \beta_{11}x_{1ij} + \beta_{12}x_{2ij}^* + b_{1i} \quad \text{and} \quad \text{logit}\left(\omega_{ij}\right) = \beta_{20} + \beta_{21}x_{1ij} + \beta_{22}x_{2ij}^* + b_{2i},$$

Figure 28 – Trace plots and marginal *posterior* distributions of parameters of the $\mathcal{ZMPL}_2$ model.



Source: Elaborated by the author.

where $\boldsymbol{b}_i = (b_{1i}, b_{2i})^\top$ is the vector of rat-specific random-intercepts. We have adopted the product-Normal parameterization (6.7) for the elements of $\boldsymbol{b}_i$, which caracterizes Case 3. From that, Case 1 is obtained by assuming $b_{1i} = b_{2i} = b_i$ and Case 2 is obtained by setting $\rho = 0$. The same strucutre was adopted to fit occurrence-persistence models based on the $\mathcal{ZMP}$ and $\mathcal{ZMNB}$ distributions.

The comparison procedure between the zero-modified models is presented in Table 42. One can notice that the $\mathcal{ZMP}$ model has performed poorly in all cases. On the other hand, both $\mathcal{ZMNB}$ and $\mathcal{ZMPL}$ models have performed quite well, with the latter being slightly better. In practical terms, we may choose any of these two distributions to represent the available data. However, we argue that the proposed model is most parsimonious as it has a fewer number of parameters to handle the positive observations. Besides, when looking at the PBF estimates presented in Table 43, one may conclude that there is no evidence that the $\mathcal{ZMNB}$ model is more plausible than the $\mathcal{ZMPL}$, especially under Cases 2 and 3. The estimates for the PBF when M_0: $\mathcal{ZMP}$ were always much higher than 100, reassuring that this model cannot be considered for the analysis of the available data.

In Table 42, we have also reported the scaled CPOs and Bayesian p-values for the fitted models. Noticeably, there is no indication of lack-of-fit for the $\mathcal{ZMPL}$-based models, since over 96% of the SCPOs were higher than 0.01 and the corresponding *posterior* values of p_{B} are close to 0.50. Therefore, based on the presented results, we have chosen the proposed model under Case 2 ($\mathcal{ZMPL}_2$) for the subsequent analyses. This choice is also supported by the fact that, under Case 3, the correlation between b_{1i} and b_{2i} was estimated as approximately zero, suggesting that there is no significant association between the occurrence and the average persistence of grooming movements.

Table 43 – *Posterior* estimates for the Pseudo-Bayes Factor ($\mathcal{ZMPL}$ and $\mathcal{ZMNB}$-based models).

M_1	M_0	$\mathcal{ZMPL}$			$\mathcal{ZMNB}$		
	Case	1	2	3	1	2	3
$\mathcal{ZMPL}$	1	1.000	0.341	0.434	1.617	0.522	0.636
	2	2.935	1.000	1.275	4.746	1.532	1.867
	3	2.302	0.784	1.000	3.721	1.202	1.464

Source: Elaborated by the author.

Figures 28 depicts the post-burn-in trace plots (history of the chains) and the marginal *posterior* distributions of parameters of the $\mathcal{ZMPL}_2$ model. The acceptance rate in the AM algorithm was approximately 21%. The normality assumption for the distribution of the regression coefficients is quite reasonable, even in the presence of slight tails on the estimated densities. Besides, there exists evidence of symmetry since the *posterior* means and medians are very close to each other. All chains have passed in the Heidelberger-Welch stationarity and half-width tests

with p-values ranging from 0.229 to 0.992. For the Geweke z-score diagnostic, the corresponding p-values ranged from 0.216 to 0.599, suggesting no significant difference on the *posterior* means across regions of the chains.

Table 44 presents the *posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPL}_2$ fitted model. From the displayed results, one can make some conclusions. Firstly, one can notice that the HPDI of parameters β_{11} and β_{21} do not contain the value zero, which constitutes the administered treatment as a relevant covariate to describe the probability of a grooming movement being practiced as well as the average persistence of this movement. Under effect of oxytocin, the rats had a higher *posterior* probability of practicing any grooming movement than when they were under the effect of saline solution: odds ratio (conditional on b_{2i}) of $e^{1.879} \approx 6.67$ $(2.37, 14.32)$. In this case, the rats also had a higher average persistence of movements (as measured on the logarithmic scale by β_{11}). Conversely, the covariate SLT is less predictive as it only contributed to explaining the decrease of the probability of grooming occurrence over time, mainly when the rats were under the effect of saline solution. Since $\hat{\sigma}_2 > \hat{\sigma}_1$, we may conclude that the heterogeneity between rats affects the probability of occurrence more than the average persistence of grooming movements.

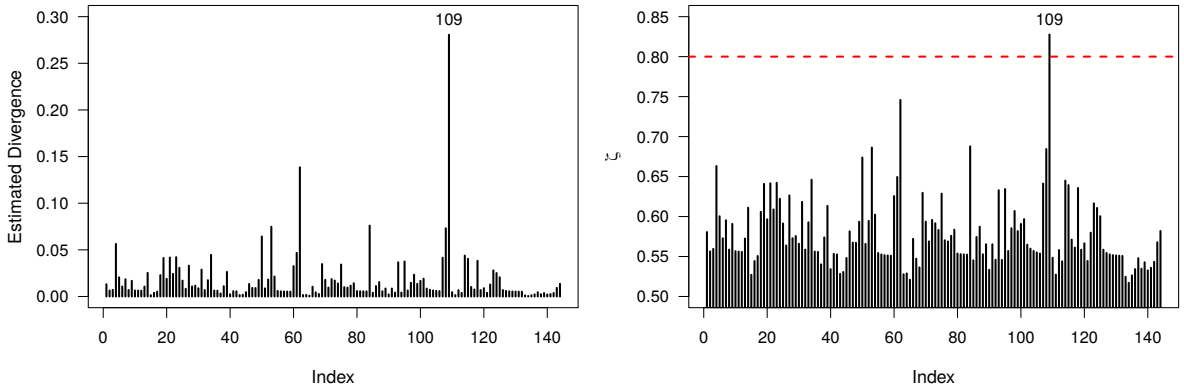
Table 44 – *Posterior* parameter estimates and 95% HPDIs from $\mathcal{ZMPL}_2$ fitted model.

Component	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Persistence	β_{10} (Intercept)	1.010	1.022	0.429	0.191	1.801
	β_{11} (Oxytocin)	0.668	0.663	0.198	0.292	1.051
	β_{12} (SLT)	0.507	0.497	0.522	−0.509	1.455
Occurrence	β_{20} (Intercept)	4.122	4.082	1.208	1.705	6.422
	β_{21} (Oxytocin)	1.897	1.888	0.430	1.110	2.794
	β_{22} (SLT)	−5.055	−5.000	1.415	−7.877	−2.370
Variance	σ_1 (Rat)	0.177	0.140	0.167	0.000	0.471
	σ_2 (Rat)	0.339	0.261	0.319	0.000	0.957

Source: Elaborated by the author.

A sensitivity analysis to verify the existence of influential points is presented in Figure 29. We have estimated all divergence measures presented in Table 127 but, since the obtained results led to the same conclusions, here we are only reporting the KL divergence and its calibration for each observation. We have considered that influential points were those whose calibration (ζ) exceeded 0.80. Based on Figure 29b, one can visualize the existence of a single influential point (109), corresponding to the first measure (five initial minutes) for Rat 5 after receiving oxytocin (zero grooming movements). In order to assess the influence of point 109, the $\mathcal{ZMPL}_2$ model was estimated once again, without using such observation. Besides, as we aimed to obtain more precise estimates, the covariate SLT was no longer considered in the log-linear model for μ_{ij}.

Figure 29 – Sensitivity analysis for diagnostic of influential points.



(a) KL divergence (b) Calibration KL divergence

Source: Elaborated by the author.

Table 45 – *Posterior* parameter estimates and 95% HPDIs from fitted $\mathcal{ZMPL}_2$ model (without influential point).

Parameter	Mean	Median	Std. Dev.	95% HPDI	
				Lower	Upper
β_{10} (Intercept)	1.377	1.377	0.197	1.010	1.765
	(36.337%)	(34.736%)	(−54.079%)	(428.796%)	(−1.999%)
β_{11} (Oxytocin)	0.704	0.705	0.200	0.313	1.086
	(5.389%)	(6.335%)	(1.010%)	(7.192%)	(3.330%)
β_{20} (Intercept)	5.079	5.025	1.329	2.412	7.585
	(23.217%)	(23.101%)	(10.017%)	(41.466%)	(18.110%)
β_{21} (Oxytocin)	2.057	2.039	0.444	1.282	2.988
	(8.434%)	(7.998%)	(3.256%)	(15.495%)	(6.943%)
β_{22} (SLT)	−6.205	−6.141	1.566	−9.278	−3.314
	(22.750%)	(22.820%)	(10.671%)	(17.786%)	(39.831%)
σ_1 (Rat)	0.179	0.138	0.164	0.000	0.486
	(1.130%)	(−1.429%)	(−1.796%)	(0.000%)	(3.185%)
σ_2 (Rat)	0.317	0.237	0.304	0.000	0.895
	(−6.490%)	(−9.195%)	(−4.702%)	(0.000%)	(−6.479%)

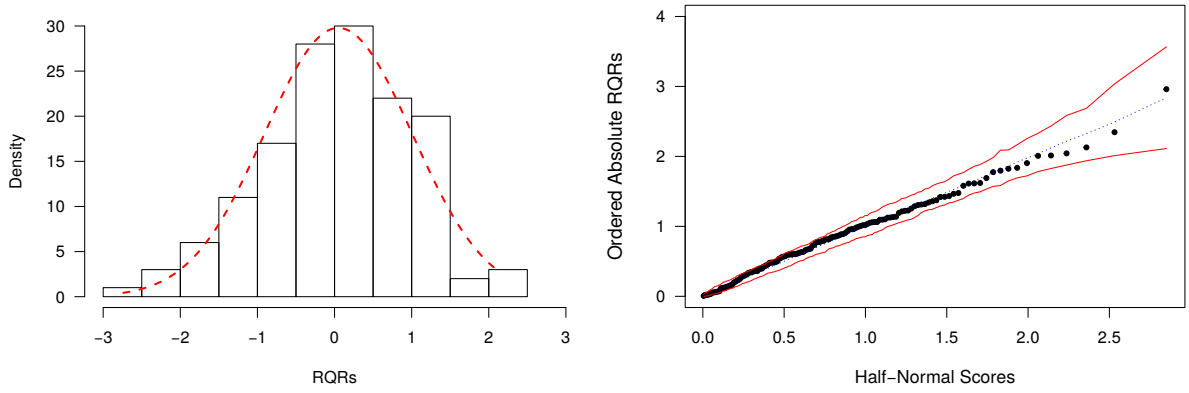
Source: Elaborated by the author.

The *posterior* parameter estimates without observation 109 and the respective variation percentage regarding the previous results (with the full dataset) are presented in Table 45. The acceptance rate in the AM algorithm was approximately 18% in this case. After performing tests for diagnosing convergence, the stationarity of the chains was revealed. Overall, the removal of the influential point was responsible to grew the estimates of β_{10}, β_{20}, and β_{22} significantly. However, the interpretability of parameters (and related functions) remains unchanged, being

just a matter of magnitude. For example, now we may conclude that grooming movements were $e^{2.057} \approx 7.82$ $(2.60, 17.02)$ times more likely to occur when the rats were under the effect of oxytocin and that the heterogeneity between rats in the probability of occurrence has decreased around 6.50%.

There are some indications that the fit summarized in Table 45 is adequate since the corresponding Bayesian p-value is 0.484, and over 97% of the SCPOs were higher than 0.01. Figure 30 depicts additional evidence based on the RQRs that also supports the $\mathcal{ZMPL}_2$ model. The normality assumption for the residuals can be easily verified by the behavior of its frequency distribution (left-panel). Also, the Half-Normal probability plot indicates that the fit without the influential point was very satisfactory since all estimated residuals are lying within the simulated envelope (right-panel).

Figure 30 – Frequency distribution and Half-Normal plot with simulated envelope for the RQRs.



Source: Elaborated by the author.

Based on the *posterior* estimates displayed on Table 45, the components of the fitted $\mathcal{ZMPL}_2$ model can be expressed by

$$\hat{\mu}_{ij} = \exp\left\{1.377 + 0.704x_{1ij} + \overline{b}_{1i}\right\},$$

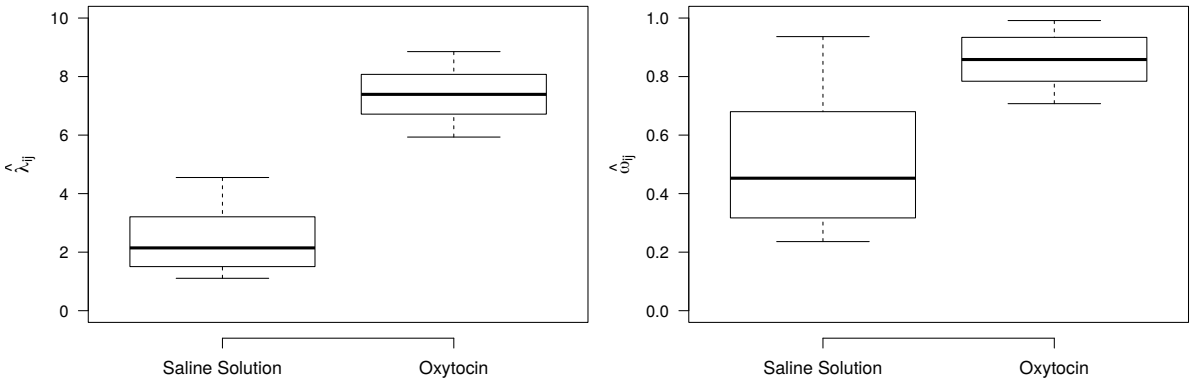
and

$$\hat{\omega}_{ij} = \frac{\exp\left\{5.079 + 2.057x_{1ij} - 6.205x^*_{2ij} + \overline{b}_{2i}\right\}}{1 + \exp\left\{5.079 + 2.057x_{1ij} - 6.205x^*_{2ij} + \overline{b}_{2i}\right\}},$$

where $\overline{b}_{1i} \approx 0.179\tilde{u}_{1i}$ and $\overline{b}_{2i} \approx 0.317\tilde{u}_{2i}$ are predictions for the random-intercepts (see Subsection 6.3.3) associated to the i-th rat.

From the fitted model, one may also obtain Bayesian estimates for the zero modification parameter as $\hat{p}_{ij} = \hat{\omega}_{ij}\mathrm{P}^{-1}(Y_{ij} > 0; \hat{\mu}_{ij})$ and consequently, for the overall mean as $\hat{\lambda}_{ij} = \hat{\mu}_{ij}\hat{p}_{ij}$. Figure 31 present Bayesian estimates, per treatment, for parameters λ_{ij} and ω_{ij}. One can notice that, through the $\mathcal{ZMPL}_2$ model, we were able to correctly characterize the probability of occurrence as well as the average persistence of grooming movements. Moreover, the estimates

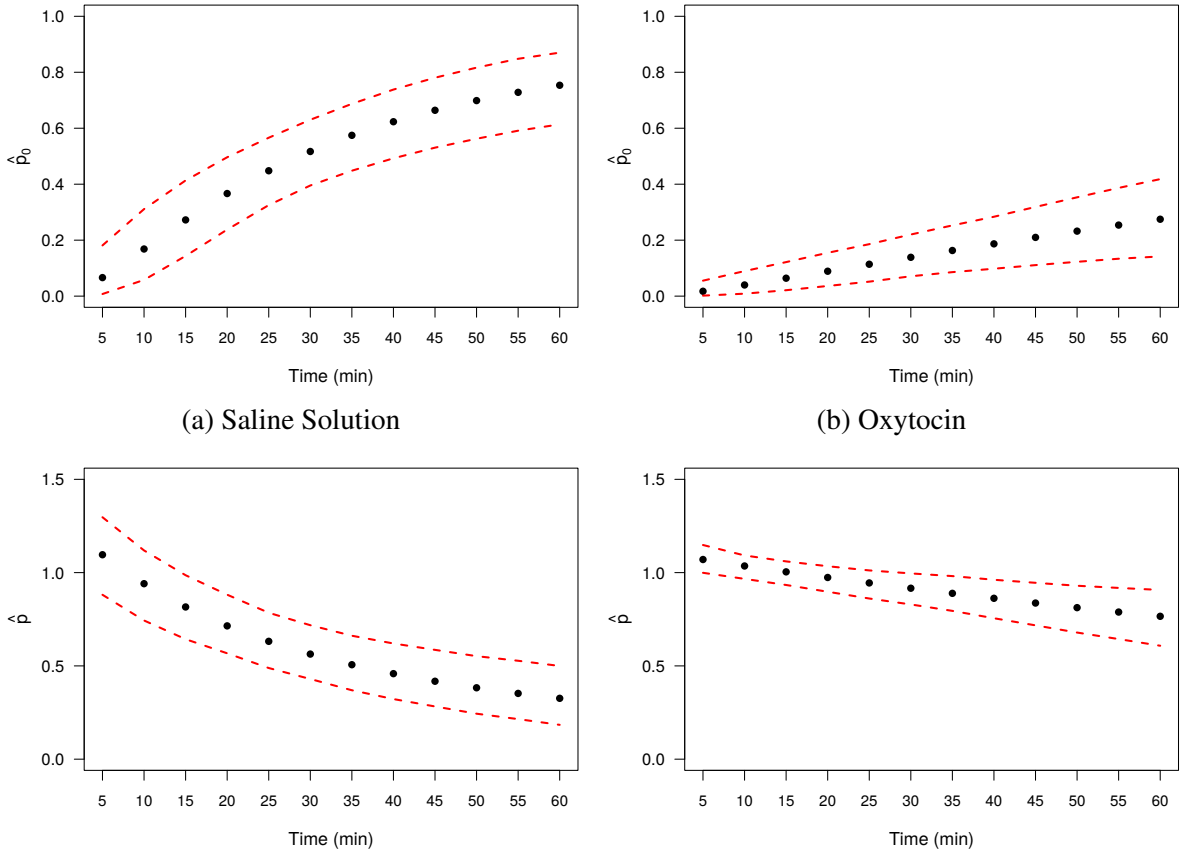
Figure 31 – *Posterior* estimates (per treatment) of representative parameters λ_{ij} and ω_{ij}.



Source: Elaborated by the author.

are not showing great variability between treatments (small caveat for $\hat{\omega}_{ij}$ under saline solution), which indicates that the obtained results are greatly consistent with the observed data.

Figure 32 – *Posterior* estimates (per treatment) of parameters p_0 and p over time.



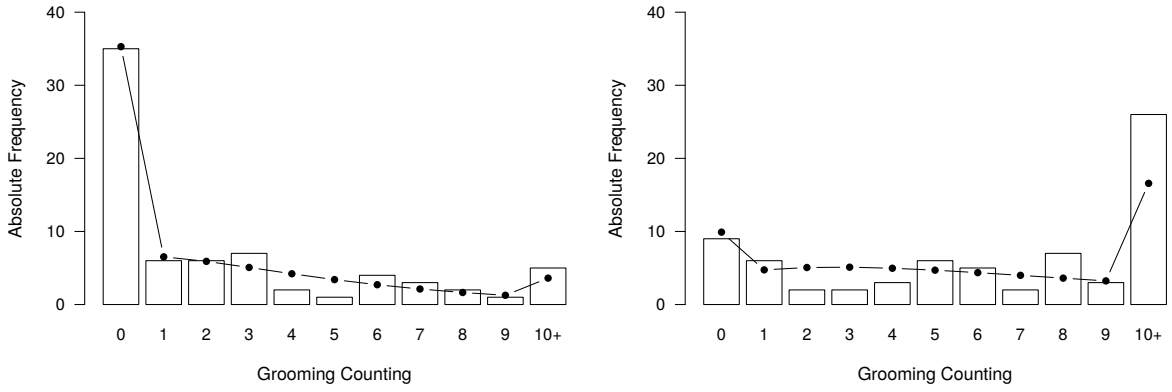
(a) Saline Solution (b) Oxytocin

(c) Saline Solution (d) Oxytocin

Source: Elaborated by the author.

Figure 32 highlights the evolution, per treatment, of the average estimated probability of non-occurrence of grooming movements (upper-panels) and of parameter p (lower-panels). The point estimates are denoted by the black dots, and the corresponding 95% HPDIs are represented by the red dashed lines. Noticeably, the occurrence of movements had become less likely over time, especially when the rats were under the effect of saline solution. These results are consistent with the behavior of grooming counting, as illustrated in Figure 26. Besides, although inferences for parameter p confirm that the response variable is majorly zero-inflated within-rat, the $\mathcal{ZMPL}_2$ model was capable of identifying that the persistence of movements was probably a bit higher than expected in the five initial minutes of both treatments. The use of zero-modified models is even more justified in situations where the nature of the zero modification changes over time (or across individuals).

Figure 33 – *Posterior* expected frequencies (per treatment) under the $\mathcal{ZMPL}_2$ fitted model.



(a) Saline Solution (b) Oxytocin

Source: Elaborated by the author.

As a final assessment of the suitability of the $\mathcal{ZMPL}_2$ fitted model, we present, in Figure 33, the *posterior* expected frequencies, per treatment, for the number of grooming movements practiced by the six rats in a neurophysiologic experiment. One can notice that the proposed model was able to handle all three sources of variation present in the response variable parsimoniously, then providing an excellent fit to the available data. The results presented in this section show that the proposed model is highly flexible to fit zero-modified data with a hierarchical structure, being also competitive with well-established models in the literature. In this way, we consider the $\mathcal{ZMPL}$ model as a great addition to the class of zero-modified regression models with mixed-effects structure.

6.6 Concluding remarks

Over the last years, the model-based approach for data analysis was becoming increasingly challenging, with more complex data structures being revealed. Particularly, when handling

count data, there are structures not supported by some standard distributions, which led many researchers to cast doubt about the usefulness of such models in some cases. Nowadays, the most prevalent data structures ruling the nature of discrete phenomena are the overdispersion, the zero modification, and the individual heterogeneity due to clusterization or existence of repeated measures. There are many practical situations in which those three characteristics are present in a count dataset, and still, there are few models supporting all at once.

In this context, we aimed to introduce a robust alternative to handle all these data structures simultaneously. Our approach was based on proposing an extended mixed-effects regression model for the analysis of zero-modified hierarchical count data. The proposed model is based on the hurdle version of the $\mathcal{PL}$ distribution, in which the zero modification is characterized by a binary probability model for the zero-valued observations and a zero-truncated distribution for the positive ones. One of the main aspects of the $\mathcal{ZMPL}$ model under such formulation is that any estimator for the expected value of the untruncated $\mathcal{PL}$ distribution depends only on the positive values, which implies that the observed zeros are only responsible for explaining, interchangeably, the zero probability or the degree of zero modification in a given dataset. In this chapter, all inferential procedures were conducted under a fully Bayesian approach. We have considered an adaptation of the *g-prior* method for the fixed-effects, and $\mathcal{HC}$ *prior* distributions for the random-effects parameters. The AM algorithm was used for sampling values from the approximate *posterior* distribution of model parameters. Intensive Monte Carlo simulation studies were performed, and the obtained results have allowed us to assess the empirical properties of the Bayesian estimators and then conclude about the suitability of the adopted methodology to the predefined scenarios.

The proposed model was considered for the analysis of a real dataset on the number of grooming movements practiced by six rats in a neurophysiologic experiment. The choice of the mixed-effects $\mathcal{ZMPL}$ regression model is justifiable since the response variable was identified as overdispersed, zero-inflated, and with individual heterogeneity due to repeated measures over time. A sensitivity analysis was conducted by using some standard divergence measures, where a single locally influent point was identified and then removed before further inferential procedures. The adequacy of the fitted models was evaluated by using the Bayesian p-value, the SCPOs, and the RQRs.

The main conclusion one can make from the fitted models is that the administered treatment is statistically relevant to describe either the probability of occurrence as well as the average persistence of grooming movements. Besides, inferences for the zero-modified parameter (p) indicate that the distribution of the response variable is majorly zero-inflated over time, with this conclusion being extended for the observed zeros within-rat. When looking at the *posterior* comparison criteria and the PBF estimates, we have noticed that the $\mathcal{ZMPL}$ model with independent random-intercepts has outperformed its competitors. Therefore, based on the obtained results, we conclude by stating that the proposed model can be considered an excellent

addition to the class of mixed-effects regression models as it has proven to be flexible and competitive when it comes to modeling zero-modified count data with a hierarchical structure.

CHAPTER

7

CONCLUSIONS AND PERSPECTIVES

In this doctoral thesis, we have introduced a new class of discrete models as an alternative for the analysis of zero-modified count data. The proposed class is composed of hurdle versions of the Poisson-Lindley, Poisson-Shanker, and Poisson-Sujatha distributions, as presented in Chapter 2. Unlike the traditional formulation of zero-modified models, in the hurdle configuration, the positive observations are entirely represented by zero-truncated versions of the baseline distributions. Since the adopted uniparametric Poisson mixtures can accommodate different levels of overdispersion, the proposed models showed to be natural candidates for modeling zero inflation/deflation, which are structures that typically induces an additional layer of extra-Poisson variability on the data.

In the sense of extending the applicability of the theoretical models, we have developed a fixed-effects regression framework, in which an overdispersed and zero-modified response variable could also be modeled in the presence of covariates. In order to gain interpretability, the hurdle models were reparameterized in terms of the expected value of the baseline distributions, and so, the proposed regression structure can overlap the probability of zero-valued observations being generated as well as the average number of positive observations per individual. Such a regression framework was considered to all models of the proposed class, as presented in Chapters 3-5.

The final contribution of this work was the development of an even more flexible regression framework, which allowed for the inclusion of both fixed and random-effects in the linear predictors of the hurdle models. This extension arose naturally from the structure presented in Chapters 3-5, and its proposition is justifiable as the mixed-effects approach provides a comprehensive way to handle more complex data structure involving the presence of individual heterogeneity due to clustering or repeated measurements. Particularly, our approach was based on the use of scalar random-effects (random-intercepts) to quantify the within-subjects heterogeneity. Although the proposed mixed-effects framework could be extended to all models of the proposed class, in this work, we have only applied such a structure to the hurdle version of

the Poisson-Lindley distribution, as presented in Chapter 6.

All estimation procedures in this work were conducted under a fully Bayesian perspective. In Chapter 2, we have considered *Jeffreys' priors* for both parameters (θ, ω) of the theoretical models. From that, it was possible to obtain a known *posterior* distribution only for parameter ω. Thus, in order to generate pseudo-random samples from the *posterior* distribution of θ and make inferences, we have resorted to a rejection sampling scheme. In Chapters 3-6, we have considered adaptations of the *g-prior* method for the fixed-effects of both developed regression structures. In Chapters 3-4, the variance scale factors were arbitrarily chosen, and in Chapters 5-6, we have considered that the amount of information provided by the *g-prior* should be the same as the amount of information contained in a single observation. In all these cases, we were not able to obtain a closed-form for the *posterior* distributions, and so we have adopted the Random-walk Metropolis algorithm with different structures for the proposal density as the alternative to generate pseudo-random samples from the *posterior* distribution of the fixed-effects (Chapters 5-6). Additionally, in Chapter 6, we have considered noninformative *priors* for the random-effects parameters, and the Adaptive Metropolis algorithm was used to approximate the full *posterior* density of the proposed mixed-effects regression model.

Several Monte Carlo simulation studies were performed in order to assess the empirical properties of the Bayesian estimators and, thus, evaluate the performance of the proposed methodologies. We have learned from the obtained results in each chapter that, for both theoretical and regression models, the overall accuracy of the estimators improved with increasing sample sizes since the estimated biases, mean squared errors, and mean absolute percentage error have decreased as the sample sizes increased. Besides, in most of the studied cases, the coverage of the 95% highest *posterior* density intervals (or the Bayesian credible intervals, when used) became closer to the nominal ones as the sample sizes increased. A specific pathological case involving zero inflation was considered in Chapter 2 to illustrate the instability of the so adopted methodology to estimate the model parameters in such a case. Additionally, in Chapter 2, we have also conducted simulation studies aiming to evaluate the reliability of the Bayesian comparison criteria (DIC, EAIC, EBIC, and LMPL) and to quantify the possible loss of efficiency that may occur when the proposed models are used for the analysis of datasets without zero modification.

The usefulness of the proposed class of zero-modified models was illustrated by using real datasets with different structures. In Chapter 2, we have modeled three overdispersed samples (without covariates), being two of them identified as zero-deflated. In Chapter 3, we have analyzed zero-inflated data on the notification of fetal deaths registered in the Bahia State (Brazil) in 2014 using the zero-modified Poisson-Shanker regression model. The human development index was found statistically relevant to describe either the probability of non-notification as well as the average number of fetal deaths. In Chapter 4, we have enhanced the modeling of the Takeover Bids dataset (also considered in Chapter 2) by including the *white knight* (a management action of inviting a friendly third-part to enter the bidding) as a covariate in the fixed-effects regression

model. In Chapter 5, a heavily zero-inflated dataset on the number of cytogenetic chromosomal aberrations was analyzed using a regression framework based on the hurdle version of the Poisson-Sujatha distribution. The dose of ionizing radiation was considered as the only covariate to describe both parameters of the theoretical model. Finally, in Chapter 6, we have used the zero-modified Poisson-Lindley model with a mixed-effects structure to modeling the number of grooming movements practiced by six rats of the War species after receiving two treatments (saline solution and oxytocin). In this study, the response variable was identified as overdispersed, zero-inflated, and with individual heterogeneity due to repeated measures over time. The data analysis highlighted that the administered treatment was statistically relevant to describe either the probability of occurrence as well as the average persistence of grooming movements.

Beyond parameter estimation, we have performed sensitivity analyses to identify influent points. In order to validate the fitted models, we have computed Bayesian p-values, the randomized quantile residuals, among other measures. The competitiveness of the proposed models was proven in all examples and, despite having a strong competitor (the zero-modified Negative Binomial distribution), which frequently provides similar fits, we argue that the hurdle versions of the Poisson-Lindley/Shanker/Sujatha distributions are most parsimonious as they have a fewer number of parameters to handle the positive observations. We hence conclude that the proposed zero-modified models can be considered an excellent addition to the literature of discrete distributions as they are flexible and competitive when it comes to modeling overdispersed and zero-modified count data, even in the presence of individual heterogeneity arising from clustering or repeated (correlated) measurements.

7.1 Perspectives

Some works that can provide continuity to the current research project are listed in the following.

- Expand the class of zero-modified models by considering other Poisson mixtures as baseline distributions;
- Adopt different *prior* distributions for the vector of fixed-effects from the proposed regression models;
- Adopt different distributions for the random-intercepts in the proposed mixed-effects regression model;
- Extend the proposed mixed-effects structure to allow the inclusion of bivariate random-effects (intercepts and slopes);
- Develop a flexible and user-friendly `R` package for the analysis of zero-modified count data using the models proposed in this thesis.

BIBLIOGRAPHY

ACHCAR, J. A.; COELHO-BARROS, E. A.; MARTINEZ, E. Z. Statistical analysis for longitudinal counting data in the presence of a covariate considering different "frailty" models. **Brazilian Journal of Probability and Statistics**, JSTOR, v. 22, n. 2, p. 183–205, 2008. Citation on page 172.

AL-KHASAWNEH, M. F. Estimating the Negative Binomial dispersion parameter. **Asian Journal of Mathematics & Statistics**, v. 3, n. 1, p. 1–15, 2010. Citation on page 36.

ALFÒ, M.; MARUOTTI, A. Two-part regression models for longitudinal zero-inflated count data. **Canadian Journal of Statistics**, Wiley Online Library, v. 38, n. 2, p. 197–216, 2010. Citation on page 32.

ANGERS, J. F.; BISWAS, A. A Bayesian analysis of zero-inflated generalized Poisson model. **Computational Statistics & Data Analysis**, Elsevier, v. 42, n. 1, p. 37–46, 2003. Citation on page 30.

BAETSCHMANN, G.; WINKELMANN, R. A dynamic hurdle model for zero-inflated count data. **Communications in Statistics - Theory and Methods**, Taylor & Francis, v. 46, n. 14, p. 7174–7187, 2017. Citation on page 31.

BAHN, G. D.; MASSENBURG, R. Deal with excess zeros in the discrete dependent variable, the number of homicide in Chicago census tract. In: **Joint Statistical Meetings of the American Statistical Association**. [S.l.: s.n.], 2008. p. 3905–3912. Citation on page 30.

BAKOUCH, H. S. A Weighted Negative Binomial-Lindley distribution with applications to dispersed data. **Anais da Academia Brasileira de Ciências**, SciELO Brasil, v. 90, n. 3, p. 2617–2642, 2018. Citation on page 30.

BARNARD, J.; MCCULLOCH, R.; MENG, X. L. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. **Statistica Sinica**, JSTOR, v. 10, n. 4, p. 1281–1311, 2000. Citation on page 158.

BARRETO-SOUZA, W.; SIMAS, A. B. General mixed Poisson regression models with varying dispersion. **Statistics and Computing**, Springer, v. 26, n. 6, p. 1263–1280, 2016. Citation on page 29.

BATEMAN, H.; ERDÉLYI, A. **Higher transcendental functions**. [S.l.]: McGraw-Hill, New York, 1953. Citation on page 46.

BATES, D. Computational methods for mixed models. **Vignette for `lme4` package**, 2011. Citation on page 152.

BAZÁN, J. L.; TORRES-AVILÉS, F.; SUZUKI, A. K.; LOUZADA, F. Power and reversal power links for binary regressions: An application for motor insurance policyholders. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 33, n. 1, p. 22–34, 2017. Citations on pages 69, 98, 122, and 151.

BERGER, J. O.; BERNARDO, J. M.; SUN, D. Overall objective *priors*. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 10, n. 1, p. 189–221, 2015. Citation on page 46.

BERTOLI, W.; CONCEIÇÃO, K. S.; ANDRADE, M. G.; LOUZADA, F. A Bayesian approach for some zero-modified Poisson mixture models. **Statistical Modelling**, SAGE Publications Sage India: New Delhi, India. Advance online publication. DOI:10.1177/1471082x19841984, 2019. Citations on pages 119, 120, and 148.

____. Bayesian approach for the zero-modified Poisson-Lindley regression model. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 33, n. 4, p. 826–860, 2019. Citation on page 148.

BERTOLI, W.; RIBEIRO, A. M. T.; CONCEIÇÃO, K. S.; ANDRADE, M. G.; LOUZADA, F. On zero-modified Poisson-Sujatha distribution to model overdispersed count data. **Austrian Journal of Statistics**, v. 47, n. 3, p. 1–19, 2018. Citation on page 119.

BEUF, K. D.; SCHRIJVER, J. D.; THAS, O.; CRIEKINGE, W. V.; IRIZARRY, R. A.; CLEMENT, L. Improved base-calling and quality scores for 454 sequencing based on a hurdle Poisson model. **BMC Bioinformatics**, BioMed Central Ltd, v. 13, n. 1, p. 303, 2012. Citation on page 30.

BOHARA, A. K.; KRIEG, R. G. A zero-inflated Poisson model of migration frequency. **International Regional Science Review**, SAGE Publications, v. 19, n. 3, p. 211–222, 1996. Citation on page 30.

BONAT, W. H.; RIBEIRO JR., P. J.; ZEVIANI, W. M. Likelihood analysis for a class of Beta mixed models. **Journal of Applied Statistics**, Taylor & Francis, v. 42, n. 2, p. 252–266, 2015. Citation on page 32.

BOVÉ, D. S.; HELD, L. *Hyper-g priors* for generalized linear models. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 6, n. 3, p. 387–410, 2011. Citations on pages 71, 124, 125, and 157.

BRADIE, B. **A friendly introduction to numerical analysis**. [S.l.]: Upper Saddle River, NJ: Pearson Prentice Hall, 2006. Citation on page 201.

BROOKS, S. P. **Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde**. 2002. 616–639 p. Citations on pages 130, 166, 203, and 248.

BROOKS, S. P.; GELMAN, A. General methods for monitoring convergence of iterative simulations. **Journal of Computational and Graphical Statistics**, Taylor & Francis Group, v. 7, n. 4, p. 434–455, 1998. Citations on pages 126 and 161.

BULMER, M. G. On fitting the Poisson-Lognormal distribution to species-abundance data. **Biometrics**, JSTOR, p. 101–110, 1974. Citation on page 29.

BÜRGIN, R.; RITSCHARD, G. Tree-based varying coefficient regression for longitudinal ordinal responses. **Computational Statistics & Data Analysis**, Elsevier, v. 86, p. 65–80, 2015. Citation on page 32.

BUU, A.; LI, R.; TAN, X.; ZUCKER, R. A. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. **Statistics in Medicine**, Wiley Online Library, v. 31, n. 29, p. 4074–4086, 2012. Citation on page 32.

CARLIN, B. P.; LOUIS, T. A. Bayes and empirical Bayes methods for data analysis. **Statistics and Computing**, Springer, v. 7, n. 2, p. 153–154, 1997. Citation on page 85.

____. **Bayes and empirical Bayes methods for data analysis**. [S.l.]: Chapman & Hall/CRC Boca Raton, FL, 2001. Citations on pages 130, 166, 203, and 248.

CHEN, M. H.; IBRAHIM, J. G. Conjugate *priors* for generalized linear models. **Statistica Sinica**, v. 13, n. 2, p. 461–476, 2003. Citations on pages 100, 124, and 156.

CHO, H.; IBRAHIM, J. G.; SINHA, D.; ZHU, H. Bayesian case influence diagnostics for survival models. **Biometrics**, Wiley Online Library, v. 65, n. 1, p. 116–124, 2009. Citations on pages 73, 128, 163, and 164.

COHEN, A. C. An extension of a truncated Poisson distribution. **Biometrics**, JSTOR, v. 16, n. 3, p. 446–450, 1960. Citation on page 30.

CONCEIÇÃO, K. S.; ANDRADE, M. G.; LOUZADA, F. Zero-modified Poisson model: Bayesian approach, influence diagnostics, and an application to a Brazilian leptospirosis notification data. **Biometrical Journal**, Wiley Online Library, v. 55, n. 5, p. 661–678, 2013. Citations on pages 31 and 82.

____. On the zero-modified Poisson model: Bayesian analysis and *posterior* divergence measure. **Computational Statistics**, Springer, v. 29, n. 5, p. 959–980, 2014. Citations on pages 31, 45, 46, 99, 124, and 154.

CONCEIÇÃO, K. S.; LOUZADA, F.; ANDRADE, M. G.; HELOU, E. S. Zero-modified Power Series distribution and its hurdle distribution version. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 87, n. 9, p. 1842–1862, 2017. Citations on pages 31, 44, and 97.

CONCEIÇÃO, K. S.; SUZUKI, A. K.; ANDRADE, M. G.; LOUZADA, F. A Bayesian approach for a zero modified Poisson model to predict match outcomes applied to the 2012-13 La Liga season. **Brazilian Journal of Probability and Statistics**, Brazilian Statistical Association, v. 31, n. 4, p. 746–764, 2017. Citation on page 31.

CONGDON, P. **Bayesian models for categorical data**. [S.l.]: John Wiley & Sons, 2005. Citations on pages 128, 164, and 166.

COOPER, N. J.; LAMBERT, P. C.; ABRAMS, K. R.; SUTTON, A. J. Predicting costs over time using Bayesian Markov chain Monte Carlo methods: An application to early inflammatory polyarthritis. **Health Economics**, Wiley Online Library, v. 16, n. 1, p. 37–56, 2007. Citation on page 152.

CORLESS, R. M.; GONNET, G. H.; HARE, D. E. G.; JEFFREY, D. J.; KNUTH, D. E. On the LambertW function. **Advances in Computational Mathematics**, Springer, v. 5, n. 1, p. 329–359, 1996. Citation on page 169.

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society. Series B (Methodological)**, Wiley Online Library, v. 30, n. 2, p. 248–265, 1968. Citation on page 74.

CSISZÁR, I. Information-type measures of difference of probability distributions and indirect observations. **Studia Scientiarum Mathematicarum Hungarica**, v. 2, p. 299–318, 1967. Citations on pages 72, 128, 163, and 245.

DATASUS. Secretaria Executiva do Ministério da Saúde, 2016. (Departamento de Informática do SUS). Available: <http://www.datasus.gov.br>. Citation on page 81.

DEB, P.; TRIVEDI, P. K. The structure of demand for health care: Latent class versus two-part models. **Journal of Health Economics**, Elsevier, v. 21, n. 4, p. 601–625, 2002. Citation on page 30.

DIETZ, E.; BÖHNING, D. On estimation of the Poisson parameter in zero-modified Poisson models. **Computational Statistics & Data Analysis**, Elsevier, v. 34, n. 4, p. 441–459, 2000. Citations on pages 30, 45, 46, 54, 99, 124, and 154.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citations on pages 74, 131, 167, and 248.

EVANS, M.; SWARTZ, T. **Approximating integrals via Monte Carlo and deterministic methods**. [S.l.]: Oxford University Press, 2000. Citation on page 156.

FABIO, L. C.; PAULA, G. A.; CASTRO, M. A Poisson mixed model with nonnormal random effect distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 6, p. 1499–1510, 2012. Citation on page 151.

GALVIS, D. M.; BANDYOPADHYAY, D.; LACHOS, V. H. Augmented mixed Beta regression models for periodontal proportion data. **Statistics in Medicine**, Wiley Online Library, v. 33, n. 21, p. 3759–3771, 2014. Citation on page 32.

GAO, J.; PAN, Y.; HABER, M. Assessment of observer agreement for matched repeated binary measurements. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 5, p. 1052–1060, 2012. Citations on pages 31 and 32.

GARAY, A. M.; BOLFARINE, H.; LACHOS, V. H.; CABRAL, C. R. B. Bayesian analysis of censored linear regression models with scale mixtures of Normal distributions. **Journal of Applied Statistics**, Taylor & Francis, v. 42, n. 12, p. 2694–2714, 2015. Citations on pages 73 and 164.

GEISSER, S. **Predictive inference: An introduction**. [S.l.]: Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 1993. Citation on page 246.

GEISSER, S.; EDDY, W. F. A predictive approach to model selection. **Journal of the American Statistical Association**, Taylor & Francis Group., v. 74, n. 365, p. 153–160, 1979. Citations on pages 128, 164, and 203.

GELFAND, A. E.; DEY, D. K. Bayesian model choice: Asymptotics and exact calculations. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 501–514, 1994. Citation on page 166.

GELMAN, A. *Prior* distributions for variance parameters in hierarchical models. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 1, n. 3, p. 515–533, 2006. Citation on page 158.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. [S.l.]: Chapman & Hall/CRC, Boca Raton, FL, 2004. Citations on pages 130, 166, 167, and 204.

GELMAN, A.; ROBERTS, G. O.; GILKS, W. R.; WALTER, R. Efficient Metropolis jumping rules. **Bayesian Statistics**, v. 5, n. 599-608, p. 42, 1996. Citation on page 160.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Citeseer, v. 7, n. 4, p. 457–472, 1992. Citation on page 101.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of *posterior* moments. **Bayesian Statistics**, Clarendon Press, v. 4, p. 641–649, 1992. Citations on pages 72, 101, 126, and 161.

GHITANY, M. E.; AL-MUTAIRI, D. K. Estimation methods for the discrete Poisson-Lindley distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis Group, v. 79, n. 1, p. 1–9, 2009. Citation on page 38.

GHITANY, M. E.; AL-MUTAIRI, D. K.; NADARAJAH, S. Zero-truncated Poisson-Lindley distribution and its application. **Mathematics and Computers in Simulation**, Elsevier, v. 79, n. 3, p. 279–287, 2008. Citations on pages 41, 95, and 148.

GHOSH, S. K.; MUKHOPADHYAY, P.; LU, J. C. Bayesian analysis of zero-inflated regression models. **Journal of Statistical Planning and Inference**, Elsevier, v. 136, n. 4, p. 1360–1375, 2006. Citation on page 30.

GOSSET, W. S. The probable error of a mean. **Biometrika**, JSTOR, v. 6, n. 1, p. 1–25, 1908. Citation on page 55.

GUPTA, M.; IBRAHIM, J. G. An information matrix *prior* for Bayesian analysis in generalized linear models with high dimensional data. **Statistica Sinica**, NIH Public Access, v. 19, n. 4, p. 1641–1663, 2009. Citations on pages 70, 100, 124, 125, 156, and 157.

GUPTA, R.; SZCZESNIAK, R. D.; MACALUSO, M. Modeling repeated count measures with excess zeros in an epidemiological study. **Annals of Epidemiology**, Elsevier, v. 25, n. 8, p. 583–589, 2015. Citation on page 32.

GURMU, S.; TRIVEDI, P. K. Excess zeros in count models for recreational trips. **Journal of Business & Economic Statistics**, Taylor & Francis Group, v. 14, n. 4, p. 469–477, 1996. Citation on page 30.

HAARIO, H.; SAKSMAN, E.; TAMMINEN, J. An adaptive Metropolis algorithm. **Bernoulli**, Bernoulli Society for Mathematical Statistics and Probability, v. 7, n. 2, p. 223–242, 2001. Citation on page 160.

HALL, D. B. Zero-inflated Poisson and Binomial regression with random effects: A case study. **Biometrics**, Wiley Online Library, v. 56, n. 4, p. 1030–1039, 2000. Citation on page 32.

HANSEN, M. H.; YU, B. Minimum description length model selection criteria for generalized linear models. **Lecture Notes - Monograph Series**, JSTOR, v. 40, p. 145–163, 2003. Citations on pages 125 and 157.

HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. **Operations Research**, INFORMS, v. 31, n. 6, p. 1109–1144, 1983. Citations on pages 126 and 161.

HEILBRON, D. C. Zero-altered and other regression models for count data with added zeros. **Biometrical Journal**, Wiley Online Library, v. 36, n. 5, p. 531–547, 1994. Citations on pages 122, 151, and 153.

HEILBRON, D. C.; GIBSON, D. R. Shared needle use and health beliefs concerning AIDS: Regression modeling of zero-heavy count data. Poster session. In: **Proceedings of the Sixth International Conference on AIDS, San Francisco, CA**. [S.l.: s.n.], 1990. Citation on page 30.

HEIMERS, A.; BREDE, H. J.; GIESEN, U.; HOFFMANN, W. Chromosome aberration analysis and the influence of mitotic delay after simulated partial-body exposure with high doses of sparsely and densely ionising radiation. **Radiation and Environmental Biophysics**, Springer, v. 45, n. 1, p. 45–54, 2006. Citation on page 138.

HÖRMANN, W.; LEYDOLD, J.; DERFLINGER, G. **Automatic nonuniform random variate generation**. [S.l.]: Springer Science & Business Media, 2013. Citations on pages 51, 75, 103, 133, and 163.

HOSSEINI, F.; EIDSVIK, J.; MOHAMMADZADEH, M. Approximate Bayesian inference in spatial GLMM with Skew-Normal latent variables. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 4, p. 1791–1806, 2011. Citation on page 151.

HU, M. C.; PAVLICOVA, M.; NUNES, E. V. Zero-inflated and hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. **The American Journal of Drug and Alcohol Abuse**, Taylor & Francis Group, v. 37, n. 5, p. 367–375, 2011. Citation on page 30.

HUNGER, M.; DÖRING, A.; HOLLE, R. Longitudinal Beta regression models for analyzing health-related quality of life scores over time. **BMC Medical Research Methodology**, BioMed Central, v. 12, n. 1, p. 144, 2012. Citation on page 32.

JAGGIA, S.; THOSAR, S. Multiple bids as a consequence of target management resistance: A count data approach. **Review of Quantitative Finance and Accounting**, Springer, v. 3, n. 4, p. 447–457, 1993. Citations on pages 55 and 109.

JEFFREYS, H. An invariant form for the *prior* probability in estimation problems. **Proceedings of the Royal Society of London. Series A (Mathematical and Physical Sciences)**, The Royal Society London, v. 186, n. 1007, p. 453–461, 1946. Citation on page 46.

JEYASEELAN, L.; ANTONYSAMY, B.; JOHN, J. K. Use of Markov model in a longitudinal study with repeated measures of ordinal outcome. **Journal of Clinical Epidemiology**, Elsevier, v. 49, p. S18, 1996. Citation on page 32.

JODRÁ, P. Computer generation of random variables with Lindley or Poisson-Lindley distribution via the LambertW function. **Mathematics and Computers in Simulation**, Elsevier, v. 81, n. 4, p. 851–859, 2010. Citation on page 169.

KACHMAN, S. D.; STROUP, W. W. Generalized linear mixed models: An application. In: **Conference on Applied Statistics in Agriculture**. [S.l.]: Kansas State University, 1994. Citation on page 31.

KARLIS, D.; XEKALAKI, E. Mixed Poisson distributions. **International Statistical Review**, Wiley Online Library, v. 73, n. 1, p. 35–58, 2005. Citation on page 29.

KASS, R. E.; RAFTERY, A. E. Bayes factors. **Journal of the American Statistical Association**, Taylor & Francis, v. 90, n. 430, p. 773–795, 1995. Citation on page 166.

KASS, R. E.; WASSERMAN, L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. **Journal of the American Statistical Association**, Taylor & Francis, v. 90, n. 431, p. 928–934, 1995. Citations on pages 125 and 157.

KERMAN, J. A closed-form approximation for the median of the Beta distribution. **arXiv preprint arXiv:1111.0433**, 2011. Citation on page 47.

KOMÁREK, A.; LESAFFRE, E. Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 52, n. 7, p. 3441–3458, 2008. Citation on page 151.

KUŞ, C.; AKDOĞAN, Y.; ASGHARZADEH, A.; KıNACı, İ.; KARAKAYA, K. Binomial-Discrete Lindley distribution. **Communications Faculty of Sciences, University of Ankara. Series A1 (Mathematics and Statistics)**, v. 68, n. 1, p. 401–411, 2019. Citation on page 30.

LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. **Technometrics**, Taylor & Francis Group, v. 34, n. 1, p. 1–14, 1992. Citation on page 30.

LANDERMAN, L. R.; MUSTILLO, S. A.; LAND, K. C. Modeling repeated measures of dichotomous data: Testing whether the within-person trajectory of change varies across levels of between-person factors. **Social Science Research**, Elsevier, v. 40, n. 5, p. 1456–1464, 2011. Citations on pages 31 and 32.

LI, Y.; SCHAFER, D. W. Likelihood analysis of the multivariate ordinal *probit* regression model for repeated ordinal responses. **Computational Statistics & Data Analysis**, Elsevier, v. 52, n. 7, p. 3474–3492, 2008. Citation on page 32.

LINDLEY, D. V. Fiducial distributions and Bayes' theorem. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 102–107, 1958. Citations on pages 36 and 92.

MARIN, J. M.; ROBERT, C. **Bayesian core: A practical approach to computational Bayesian statistics**. [S.l.]: Springer Science & Business Media, 2007. Citations on pages 125 and 157.

MCCULLOCH, C. E.; NEUHAUS, J. M. Prediction of random effects in linear and generalized linear models under model misspecification. **Biometrics**, Wiley Online Library, v. 67, n. 1, p. 270–279, 2011. Citations on pages 151 and 161.

MCCULLOCH, C. E.; SEARLE, S.; NEUHAUS, J. M. **Generalized, linear, and mixed models**. [S.l.]: Wiley Online Library, 2008. Citation on page 161.

MCCULLOCH, R. E. Local model influence. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 84, n. 406, p. 473–478, 1989. Citations on pages 74, 129, 164, and 246.

MCDOWELL, A. From the help desk: Hurdle models. **The Stata Journal**, SAGE Publications, v. 3, n. 2, p. 178–184, 2003. Citation on page 30.

MCLACHLAN, G. J.; PEEL, D. **Finite mixture models**. [S.l.]: John Wiley & Sons, 2004. Citation on page 29.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The Journal of Chemical Physics**, American Institute of Physics, v. 21, n. 6, p. 1087–1092, 1953. Citations on pages 126 and 160.

MIN, Y.; AGRESTI, A. Random effect models for repeated measures of zero-inflated count data. **Statistical Modelling**, Sage Publications Sage CA: Thousand Oaks, CA, v. 5, n. 1, p. 1–19, 2005. Citations on pages 31, 32, and 151.

MOLENBERGHS, G.; VERBEKE, G. **Models for discrete longitudinal data**. [S.l.]: Springer Series in Statistics. Springer-Verlag, NY, 2005. Citation on page 29.

MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-Normal plots and overdispersed models in R: The hnp package. **Journal of Statistical Software**, Foundation for Open Access Statistics, v. 81, n. 10, p. 1–23, 2017. Citations on pages 131, 168, and 249.

MOUATASSIM, Y.; EZZAHID, E. H. Poisson regression and zero-inflated Poisson regression: Application to private health insurance data. **European Actuarial Journal**, Springer, v. 2, n. 2, p. 187–204, 2012. Citation on page 30.

MULLAHY, J. Specification and testing of some modified count data models. **Journal of Econometrics**, Elsevier, v. 33, n. 3, p. 341–365, 1986. Citations on pages 30, 44, 67, and 96.

NEELON, B. H.; O'MALLEY, A. J.; NORMAND, S. L. T. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. **Statistical Modelling**, SAGE Publications Sage India: New Delhi, India, v. 10, n. 4, p. 421–439, 2010. Citations on pages 32 and 153.

NEELON, B. H.; O'MALLEY, A. J.; SMITH, V. A. Modeling zero-modified count and semicontinuous data in health services research - Part 1: Background and overview. **Statistics in Medicine**, Wiley Online Library, v. 35, n. 27, p. 5070–5093, 2016. Citation on page 32.

____. Modeling zero-modified count and semicontinuous data in health services research - Part 2: Case studies. **Statistics in Medicine**, Wiley Online Library, v. 35, n. 27, p. 5094–5112, 2016. Citation on page 32.

NGATCHOU-WANDJI, J.; PARIS, C. On the zero-inflated count models with application to modelling annual trends in incidences of some occupational allergic diseases in France. **Journal of Data Science**, v. 9, p. 639–659, 2011. Citation on page 30.

OLIVEIRA, M.; EINBECK, J.; HIGUERAS, M.; AINSBURY, E.; PUIG, P.; ROTHKAMM, K. Zero-inflated regression models for radiation-induced chromosome aberration data: A comparative study. **Biometrical Journal**, Wiley Online Library, v. 58, n. 2, p. 259–279, 2016. Citations on pages 30 and 139.

PENG, F.; DEY, D. K. Bayesian analysis of outlier problems using divergence measures. **Canadian Journal of Statistics**, Wiley Online Library, v. 23, n. 2, p. 199–213, 1995. Citations on pages 128, 129, 164, 245, and 246.

POLSON, N. G.; SCOTT, J. G. On the Half-Cauchy *prior* for a global scale parameter. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 7, n. 4, p. 887–902, 2012. Citation on page 158.

R Development Core Team. **R: A language and environment for statistical computing**. Vienna, Austria, 2017. Available: <http://www.R-project.org>. Citations on pages 49, 72, 102, 127, and 161.

RANA, S.; ROY, S.; DAS, K. Analysis of ordinal longitudinal data under nonignorable missingness and misreporting: An application to Alzheimer's disease study. **Journal of Multivariate Analysis**, Elsevier, v. 166, p. 62–77, 2018. Citation on page 32.

RIDOUT, M.; DEMÉTRIO, C. G. B.; HINDE, J. Models for count data with many zeros. In: **Proceedings of the XIXth International Biometric Conference**. [S.l.: s.n.], 1998. v. 19, p. 179–192. Citation on page 30.

ROBERT, C.; CASELLA, G. **Monte Carlo statistical methods**. [S.l.]: Springer Science & Business Media, 2013. Citation on page 48.

ROBERTS, G. O.; GELMAN, A.; GILKS, W. R. Weak convergence and optimal scaling of random walk Metropolis algorithms. **The Annals of Applied Probability**, Institute of Mathematical Statistics, v. 7, n. 1, p. 110–120, 1997. Citations on pages 101, 126, and 160.

ROBERTS, G. O.; ROSENTHAL, J. S. Examples of adaptive MCMC. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 18, n. 2, p. 349–367, 2009. Citation on page 160.

RODRIGUES, J. Bayesian analysis of zero-inflated distributions. **Communications in Statistics - Theory and Methods**, Taylor & Francis Group, v. 32, n. 2, p. 281–289, 2003. Citation on page 30.

RUBIN, D. B. Bayesianly justifiable and relevant frequency calculations for the applied statistician. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 12, n. 4, p. 1151–1172, 1984. Citations on pages 130, 167, and 204.

RULI, E.; SARTORI, N.; VENTURA, L. Improved Laplace approximation for marginal likelihoods. **Electronic Journal of Statistics**, The Institute of Mathematical Statistics and the Bernoulli Society, v. 10, n. 2, p. 3986–4009, 2016. Citation on page 156.

SAFFARI, S. E.; ADNAN, R.; GREENE, W. Parameter estimation on hurdle Poisson regression model with censored data. **Jurnal Teknologi**, v. 57, n. 1, p. 189–198, 2012. Citations on pages 44, 67, and 96.

SANKARAN, M. The discrete Poisson-Lindley distribution. **Biometrics**, JSTOR, v. 26, n. 1, p. 145–149, 1970. Citations on pages 29, 38, 92, and 148.

SHABAN, S. A. On the discrete Poisson-Inverse Gaussian distribution. **Biometrical Journal**, Wiley Online Library, v. 23, n. 3, p. 297–303, 1981. Citation on page 29.

SHANKER, R. Shanker distribution and its applications. **International Journal of Statistics and Applications**, Scientific & Academic Publishing, v. 5, n. 6, p. 338–348, 2015. Citation on page 37.

____. Sujatha distribution and its applications. **Statistics in Transition - New Series**, Główny Urząd Statystyczny - Zakład Wydawnictw Statystycznych, v. 17, n. 3, p. 1–20, 2016. Citation on page 37.

____. The discrete Poisson-Amarendra distribution. **International Journal of Statistical Distributions and Applications**, Science Publishing Group, v. 2, n. 2, p. 14–21, 2016. Citation on page 30.

____. The discrete Poisson-Shanker distribution. **Jacobs Journal of Biostatistics**, Jacobs Publishers, v. 1, n. 1, p. 1–7, 2016. Citations on pages 30, 38, and 64.

____. The discrete Poisson-Sujatha distribution. **International Journal of Probability and Statistics**, Scientific & Academic Publishing, v. 5, n. 1, p. 1–9, 2016. Citations on pages 30, 38, and 119.

____. A zero-truncated Poisson-Shanker distribution and its applications. **International Journal of Statistics and Applications**, v. 7, n. 3, p. 159–169, 2017. Citation on page 41.

SHANKER, R.; FESSHAYE, H. On zero-truncation of Poisson, Poisson-Lindley and Poisson-Sujatha distributions and their applications. **Biometrics & Biostatistics International Journal**, v. 3, n. 4, p. 125–135, 2016. Citations on pages 41, 95, and 120.

SHANKER, R.; MISHRA, A. A two parameter Poisson-Lindley distribution. **International Journal of Statistics and Systems**, v. 9, n. 1, p. 79–85, 2014. Citation on page 30.

____. A quasi Poisson-Lindley distribution. **Journal of Indian Statistical Association**, v. 54, n. 1 & 2, p. 113–125, 2016. Citation on page 30.

SHANKER, R.; SHARMA, S.; SHANKER, U.; SHANKER, R.; LEONIDA, T. A. The discrete Poisson-Janardan distribution with applications. **International Journal of Soft Computing and Engineering**, Citeseer, v. 4, n. 2, p. 31–33, 2014. Citation on page 29.

SHANKER, R.; SHUKLA, K. K. On Poisson-Weighted Lindley distribution and its applications. **Journal of Scientific Research**, v. 11, n. 1, p. 1–13, 2018. Citation on page 30.

SMALL, C. G. **Expansions and asymptotics for statistics**. [S.l.]: Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 2010. Citations on pages 47 and 156.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit (with discussion). **Journal of the Royal Statistical Society. Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citation on page 203.

STUDENT. On the error of counting with a haemocytometer. **Biometrika**, v. 5, n. 3, p. 351–360, 1907. Citation on page 55.

SYVERSVEEN, A. R. Noninformative Bayesian *priors*. Interpretation and problems with construction and applications. **Preprint Statistics**, v. 3, n. 3, p. 1–11, 1998. Citation on page 46.

UMBACH, D. On inference for a mixture of a Poisson and a degenerate distribution. **Communications in Statistics - Theory and Methods**, Taylor & Francis Group, v. 10, n. 3, p. 299–306, 1981. Citation on page 30.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. [S.l.]: Springer Series in Statistics. Springer-Verlag, NY, 2000. Citation on page 31.

VON NEUMANN, J. Various techniques used in connection with random digits. **National Bureau of Standards Applied Mathematics Series**, v. 12, n. 5, p. 36–38, 1951. Citation on page 48.

WAGH, Y. S.; KAMALJA, K. K. Zero-inflated models and estimation in zero-inflated Poisson distribution. **Communications in Statistics - Simulation and Computation**, Taylor & Francis, v. 47, n. 8, p. 2248–2265, 2018. Citation on page 30.

WANG, K.; YAU, K. K. W.; LEE, A. H. A zero-inflated Poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 68, n. 3, p. 195–203, 2002. Citation on page 32.

WANG, X.; GEORGE, E. I. Adaptive Bayesian criteria in variable selection for generalized linear models. **Statistica Sinica**, JSTOR, v. 17, n. 2, p. 667–690, 2007. Citations on pages 125 and 157.

WEISS, R. An approach to Bayesian sensitivity analysis. **Journal of the Royal Statistical Society. Series B (Methodological)**, Wiley Online Library, v. 58, n. 4, p. 739–750, 1996. Citations on pages 129 and 164.

WILLIAMSON, D. S.; BANGDIWALA, S. I.; MARSHALL, S. W.; WALLER, A. E. Repeated measures analysis of binary outcomes: Applications to injury research. **Accident Analysis & Prevention**, Elsevier, v. 28, n. 5, p. 571–579, 1996. Citations on pages 31 and 32.

XU, X. S.; SAMTANI, M. N.; DUNNE, A.; NANDY, P.; VERMEULEN, A.; DE RIDDER, F. Mixed-effects Beta regression for modeling continuous bounded outcome scores using NONMEM when data are not on the boundaries. **Journal of Pharmacokinetics and Pharmacodynamics**, Springer, v. 40, n. 4, p. 537–544, 2013. Citation on page 32.

YAU, K. K. W.; LEE, A. H. Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. **Statistics in Medicine**, Wiley Online Library, v. 20, n. 19, p. 2907–2920, 2001. Citation on page 32.

YIN, T.; BAPST, B.; VON BORSTEL, U. U.; SIMIANER, H.; KÖNIG, S. Genetic analyses of binary longitudinal health data in small low input dairy cattle herds using generalized linear mixed models. **Livestock Science**, Elsevier, v. 162, p. 31–41, 2014. Citations on pages 31 and 32.

ZAMANI, H.; ISMAIL, N. Negative Binomial-Lindley distribution and its application. **Journal of Mathematics and Statistics**, Science Publications, v. 6, n. 1, p. 4–9, 2010. Citation on page 29.

ZELLNER, A. On assessing *prior* distributions and Bayesian regression analysis with *g-prior* distributions. **Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti**, North Holland, Amsterdam, v. 6, p. 233–243, 1986. Citations on pages 70, 99, 124, and 156.

ZHU, H.; LUO, S.; DESANTIS, S. M. Zero-inflated count models for longitudinal measurements with heterogeneous random effects. **Statistical Methods in Medical Research**, SAGE Publications Sage UK: London, England, v. 26, n. 4, p. 1774–1786, 2017. Citation on page 32.

ZORN, C. J. W. Evaluating zero-inflated and hurdle Poisson specifications. **Midwest Political Science Association**, v. 18, n. 20, p. 1–16, 1996. Citation on page 30.

APPENDIX

A

SUPPLEMENT FOR CHAPTER 2

A.1 Maximum likelihood estimation

For the proposed models in this paper, there is no closed-form solution for the MLE of parameter θ. Therefore, standard numerical optimization algorithms such as the Newton-Raphson, the Bisection, and the Regula-Falsi methods can be used to obtain estimates in the frequentist approach (BRADIE, 2006). The MLE of ω can be easily obtained as $\hat{\omega} = n_{+}n^{-1}$ by solving the score function

$$U(\omega, \boldsymbol{y}) = \frac{\partial \ell_n(\omega; \boldsymbol{y})}{\partial \omega} = \frac{n}{\omega} - \frac{n_0}{\omega(1-\omega)}.$$

Using the fact that $\mathbb{E}_Y(n_0) = n(1-\omega)$, one can easily verify that $\hat{\omega}$ is an unbiased estimator for ω. Now, the Fisher information of ω is given by

$$\mathcal{I}_\omega(\omega) = -\mathbb{E}_Y\left[\frac{\partial^2 \ell_n(\omega; \boldsymbol{Y})}{\partial \omega^2}\right] = \mathbb{E}_Y\left[\frac{n}{\omega^2} + \frac{n_0(2\omega - 1)}{\omega^2(1-\omega)^2}\right] = \frac{n}{\omega(1-\omega)}. \tag{A.1}$$

By the maximum likelihood theory, a consistent estimator for the variance of $\hat{\omega}$ is given by

$$\widehat{\mathbb{V}}(\hat{\omega}) = \frac{\hat{\omega}(1-\hat{\omega})}{n} = \frac{n_0 n_{+}}{n^3}.$$

The Fisher information of parameter θ depends on the selected element of $\mathcal{F}_3$. In general, it can be expressed as

$$\mathcal{I}_\theta(\theta) = -\mathbb{E}_Y\left[\frac{\partial^2 \ell_{n_{+}}(\theta; \boldsymbol{Y})}{\partial \theta^2}\right], \tag{A.2}$$

and therefore, a consistent estimator for the variance of $\hat{\theta}$ can be obtained by the inverse of (A.2) evaluated at $\hat{\theta}$. When considering the $\mathcal{ZMPL}$ distribution, the estimator for the asymptotic

variance of $\hat{\theta}$ is given by

$$\widehat{\mathbb{V}}\left(\hat{\theta}\right) = \frac{\hat{\theta}^2+3\hat{\theta}+1}{n_+}\left\{\hat{\theta}^2\left[\left(\hat{\theta}+1\right)^2\zeta\left[\left(\hat{\theta}+1\right)^{-1},1,\hat{\theta}\right]-\frac{\hat{\theta}^3+5\hat{\theta}^2+8\hat{\theta}+2}{\hat{\theta}\left(\hat{\theta}+2\right)}\right]-\frac{\hat{\theta}^4-\hat{\theta}^3-8\hat{\theta}^2-10\hat{\theta}-2}{\hat{\theta}^2\left(\hat{\theta}^2+3\hat{\theta}+1\right)}\right\}^{-1}.$$

Under the $\mathcal{ZMPS}_h$ distribution, we have

$$\widehat{\mathbb{V}}\left(\hat{\theta}\right) = \frac{\hat{\theta}^3+\hat{\theta}^2+2\hat{\theta}+1}{n_+}\left\{-\frac{4\hat{\theta}^6+3\hat{\theta}^5-9\hat{\theta}^3-6\hat{\theta}^2-6\hat{\theta}-2}{\hat{\theta}^2\left(\hat{\theta}^3+\hat{\theta}^2+2\hat{\theta}+1\right)}+\hat{\theta}^2\left(2\hat{\theta}+1\right)^2\left(\hat{\theta}+1\right)\left[\zeta\left[\left(\hat{\theta}+1\right)^{-1},1,\hat{\theta}^2+\hat{\theta}\right]-\frac{\hat{\theta}+1}{\hat{\theta}\left(\hat{\theta}^2+\hat{\theta}+1\right)}\right]\right\}^{-1},$$

and finally, under the $\mathcal{ZMPS}_u$ distribution, the variance of $\hat{\theta}$ can be estimated by

$$\widehat{\mathbb{V}}\left(\hat{\theta}\right) = \frac{\hat{\theta}^4+4\hat{\theta}^3+10\hat{\theta}^2+7\hat{\theta}+2}{n_+}\left\{\frac{\hat{\theta}^3\left(2\hat{\theta}+4\right)^2}{\left(\hat{\theta}+1\right)\left(\hat{\theta}^2+4\hat{\theta}+9\right)}\Psi\left(\hat{\theta}\right)-\frac{4\hat{\theta}^8+15\hat{\theta}^7+10\hat{\theta}^6-61\hat{\theta}^5-173\hat{\theta}^4-254\hat{\theta}^3-200\hat{\theta}^2-72\hat{\theta}-12}{\hat{\theta}^2\left(\hat{\theta}^4+4\hat{\theta}^3+10\hat{\theta}^2+7\hat{\theta}+2\right)}\right\}^{-1},$$

where $\Psi(\theta)$ is a non-linear and strictly decreasing function which has an extensive and very complicated expression, also depending on the Lerch-Phi function. For this reason, such a function is omitted here, but some of its values are presented in Table 46.

Table 46 – Function $\Psi(\theta)$ for different values of θ.

θ	0.50	1.00	1.50	3.00	5.00	10.00
$\Psi(\theta)$	2.5179	1.7972	1.5499	1.2945	1.1856	1.0971

Source: Elaborated by the author.

Now, the zero modification parameter can be estimated by the equation

$$\hat{p} = s\left(\hat{\theta},\hat{\omega}\right) = \frac{n_+}{n\left[1-f\left(0;\hat{\theta}\right)\right]},$$

for $f(\cdot;\theta)\in\mathcal{F}_2$. The variance of $\hat{p}$ can be estimated using the delta-method. Since $\hat{\theta}$ and $\hat{\omega}$ are orthogonal, the covariance between them is equal to 0. Hence,

$$\begin{aligned}\widehat{\mathbb{V}}(\hat{p}) &\approx \widehat{\mathbb{V}}\left(\hat{\theta}\right)\left[\frac{\partial}{\partial\theta}s(\theta,\omega)\right]^2+\widehat{\mathbb{V}}(\hat{\omega})\left[\frac{\partial}{\partial\omega}s(\theta,\omega)\right]^2\\ &\approx \frac{n_+}{n^2\left[1-f\left(0;\hat{\theta}\right)\right]^2}\left\{n_+\frac{\widehat{\mathbb{V}}\left(\hat{\theta}\right)d^2\left(\hat{\theta}\right)}{\left[1-f\left(0;\hat{\theta}\right)\right]^2}+\frac{\left(n-n_+\right)}{n}\right\},\end{aligned}$$

where $d(\hat{\theta}) = \frac{\mathrm{d}}{\mathrm{d}\theta} f(0;\theta)$ evaluated at $\hat{\theta}$. The expression of function d for the $\mathcal{PL}$, the $\mathcal{PS}_h$ and the $\mathcal{PS}_u$ distributions are given, respectively, by

$$\frac{\hat{\theta}\left(\hat{\theta}+4\right)}{\left(\hat{\theta}+1\right)^4}, \quad \frac{\hat{\theta}\left(\hat{\theta}^4+\hat{\theta}^3+5\hat{\theta}^2+3\hat{\theta}+2\right)}{\left(\hat{\theta}^2+1\right)^2\left(\hat{\theta}+1\right)^3} \quad \text{and} \quad \frac{\hat{\theta}^2\left(\hat{\theta}^4+6\hat{\theta}^3+25\hat{\theta}^2+32\hat{\theta}+24\right)}{\left(\hat{\theta}^2+\hat{\theta}+2\right)^2\left(\hat{\theta}+1\right)^4}.$$

A.2 Model comparison and suitability

There are several methods for Bayesian model selection, which are useful to compare competing models that were fitted on the same dataset. One of the most used criteria is the deviance information criterion (DIC), which was proposed to work simultaneously as a measure of fit and complexity of the model. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ a random sample of size n from $f(\cdot;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and $\boldsymbol{y} = (y_1, \ldots, y_n)$ its observed values. The DIC criterion is given by

$$\mathrm{DIC} = \bar{\mathrm{D}}(\boldsymbol{\theta}) + \rho_{\mathrm{D}} = 2\bar{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\tilde{\boldsymbol{\theta}}),$$

where $\bar{\mathrm{D}}(\boldsymbol{\theta}) = -2\mathbb{E}\{\log[f(\boldsymbol{Y};\boldsymbol{\theta})]\}$ is the *posterior* expectation of the *deviance*. In this case, the *deviance* is evaluated at some estimate $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$, being the *posterior* conditional mean a natural choice for $\tilde{\boldsymbol{\theta}}$. In connection with a measure of model complexity, the criterion considers the measure $\rho_{\mathrm{D}} = \bar{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\tilde{\boldsymbol{\theta}})$, which correspond to the effective number of parameters in the model. One can notice that the computation of $\bar{\mathrm{D}}(\boldsymbol{\theta})$ is a complex numerical problem. In this case, a MC estimator for such a measure is given by

$$\bar{\mathrm{D}} = -\frac{2}{M}\sum_{j=1}^{M} \log\left[f\left(\boldsymbol{y};\boldsymbol{\theta}^{(j)}\right)\right],$$

and hence, the DIC can be approximated by

$$\widehat{\mathrm{DIC}} = 2\bar{\mathrm{D}} - \mathrm{D}(\tilde{\boldsymbol{\theta}}).$$

Using the estimate $\bar{\mathrm{D}}$, one can define other criteria that can be considered when comparing models. The expected Akaike information criterion (EAIC) and the expected Bayesian information criterion (EBIC) can be estimated as

$$\widehat{\mathrm{EAIC}} = \bar{\mathrm{D}} + 2q \quad \text{and} \quad \widehat{\mathrm{EBIC}} = \bar{\mathrm{D}} + q\log(n),$$

where q is the total number of estimated parameters in the model. See Carlin and Louis (2001), Spiegelhalter *et al.* (2002) and Brooks (2002) for further details on these comparison criteria. Another widely used criterion is derived from the conditional predictive ordinate (CPO) statistic (GEISSER; EDDY, 1979), which is based on the cross validation criterion to compare models. For the i-th observation of $\boldsymbol{y}$, the CPO_i is given by

$$\mathrm{CPO}_i = \int_{\boldsymbol{\Theta}} f(y_i;\boldsymbol{\theta})\,\pi\left(\boldsymbol{\theta};\boldsymbol{y}_{(-i)}\right)\mathrm{d}\boldsymbol{\theta},$$

where $\boldsymbol{y}_{(-i)}$ is the sample vector after removal of the i-th observation. In practice, there are several situations in which the CPO_i does not have a closed-form. In this case, a MC estimator for such a measure is given by

$$\widehat{\text{CPO}}_i = \left[\frac{1}{M} \sum_{k=1}^{M} \frac{1}{f\left(y_i; \boldsymbol{\theta}^{(k)}\right)} \right]^{-1}.$$

A summary statistic of the CPO measure is the log-marginal pseudo-likelihood (LMPL) given by the sum of the logarithms of $\widehat{\text{CPO}}_i$'s. In terms of model comparison, we have that the lower the value of DIC, EAIC, and EBIC, the better the fit. On the other hand, for the latter criterion, we have that the higher the value of LMPL, the better the fit. Also, one can notice that for all the presented criteria, the evaluation of the likelihood function is quite important task. In our case, this function is available in closed-form and is given by (2.6).

Once the best model has been determined, its fit to the data should be evaluated. To assess model suitability, one can use a discrepancy measure based on the ppd. For instance, if any observed value is extremely relative to the ppd, the adequacy of the model-fit might be questionable. By letting $\boldsymbol{y}$ as the observed vector, the discrepancy measure between model and data is computed as a summary statistic given by

$$T(\boldsymbol{y}, \boldsymbol{\theta}) = -2 \sum_{i=1}^{n} \log\left[f(y_i; \boldsymbol{\theta})\right].$$

The Bayesian p-value (*posterior* predictive p-value), proposed by Rubin (1984), is defined as

$$p_{\text{B}} = P\left[T(\boldsymbol{y}_s, \boldsymbol{\theta}) \geqslant T(\boldsymbol{y}, \boldsymbol{\theta}); Y = y\right],$$

where $\boldsymbol{y}_s$ denotes the simulated vector from (2.15). This measure is calculated as the number of times $T(\boldsymbol{y}_s, \boldsymbol{\theta})$ exceeds $T(\boldsymbol{y}, \boldsymbol{\theta})$ out of N simulated drawns. Gelman *et al.* (2004) states that a model becomes suspect if the discrepancy is of practical relevance and its p-value is close to 0 or 1. A very large (small) p-value, say greater than 0.95 (lower than 0.05), indicates model misspecification, that is, the observed behaviour would be unlikely to be seen if we replicate the data under the true model.

A.3 Results from the simulation studies

This section contains a full report regarding the obtained results in each simulation study proposed in the paper. The goals and particularities of each study are presented in Section 2.6 of the manuscript.

A.3.1 Study 1

A.3.1.1 Zero-inflated artificial data

The following tables illustrate the results obtained in the first simulation study, considering Scenarios 1-4 of the zero-inflated case as described in Section 2.6 (Table 6)

Table 47 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 0.50$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	θ	0.1080	0.0606	0.0723	36.2519
	ω	0.0034	0.0029	0.0029	26.2943
	p	0.0171	0.0048	0.0051	28.1300
100	θ	0.0455	0.0205	0.0226	20.6406
	ω	0.0021	0.0014	0.0014	18.7453
	p	0.0083	0.0024	0.0024	19.7549
200	θ	0.0252	0.0095	0.0102	14.1610
	ω	0.0006	0.0007	0.0007	12.7316
	p	0.0039	0.0011	0.0011	13.5143
500	θ	0.0094	0.0029	0.0030	8.5523
	ω	−0.0006	0.0003	0.0003	7.8953
	p	0.0005	0.0004	0.0004	8.3193
			Scenario 2		
50	θ	0.0274	0.0088	0.0095	14.8179
	ω	−0.0033	0.0048	0.0048	8.4442
	p	0.0096	0.0088	0.0089	9.1931
100	θ	0.0133	0.0042	0.0044	10.0023
	ω	0.0003	0.0021	0.0021	5.5269
	p	0.0071	0.0039	0.0040	6.2365
200	θ	0.0074	0.0018	0.0019	6.6842
	ω	0.0003	0.0010	0.0010	3.9307
	p	0.0041	0.0020	0.0020	4.5188
500	θ	0.0037	0.0006	0.0006	3.9766
	ω	0.0001	0.0005	0.0005	2.5635
	p	0.0020	0.0008	0.0008	2.7969

Source: Elaborated by the author.

Table 48 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 5.00$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 3		
50	θ	1.0677	14.8947	16.0346	58.5396
	ω	0.0166	0.0006	0.0009	57.7185
	p	0.1926	0.1711	0.2082	127.1419
100	θ	2.6378	31.2366	38.1944	86.2703
	ω	0.0057	0.0003	0.0004	39.9392
	p	0.1759	0.1412	0.1721	121.5846
200	θ	4.4031	67.4253	86.8126	114.9686
	ω	0.0017	0.0002	0.0002	27.6028
	p	0.1908	0.1510	0.1874	122.8802
500	θ	2.7461	47.9365	55.4774	74.9845
	ω	0.0004	0.0001	0.0001	17.2483
	p	0.1023	0.0600	0.0705	71.3279
			Scenario 4		
50	θ	4.0657	57.0500	73.5797	107.4576
	ω	0.0032	0.0028	0.0028	27.7432
	p	0.6478	1.8494	2.2691	108.3521
100	θ	3.6657	82.2925	95.7300	93.6567
	ω	0.0017	0.0014	0.0014	19.7314
	p	0.5323	1.7420	2.0254	87.0143
200	θ	1.8732	43.0111	46.5200	53.6183
	ω	0.0002	0.0006	0.0006	13.2571
	p	0.2692	0.7824	0.8549	50.5409
500	θ	0.5331	2.4133	2.6975	22.9489
	ω	−0.0007	0.0002	0.0002	8.1805
	p	0.0756	0.0573	0.0630	22.5011

Source: Elaborated by the author.

Table 49 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-inflated case with $\theta = 0.50$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	0.6080	0.6160	0.2462	0.2960	1.1175
	ω	0.1664	0.1540	0.0539	0.0805	0.2767
	p	0.2171	0.2291	0.0693	0.1014	0.3793
100	θ	0.5455	0.5535	0.1432	0.3366	0.8333
	ω	0.1651	0.1527	0.0374	0.1005	0.2421
	p	0.2083	0.2203	0.0490	0.1244	0.3130
200	θ	0.5252	0.5332	0.0975	0.3767	0.7103
	ω	0.1635	0.1511	0.0265	0.1162	0.2173
	p	0.2039	0.2159	0.0332	0.1430	0.2750
500	θ	0.5094	0.5174	0.0539	0.4149	0.6176
	ω	0.1624	0.1500	0.0173	0.1315	0.1958
	p	0.2005	0.2125	0.0200	0.1612	0.2437
Scenario 2						
50	θ	0.5274	0.5354	0.0938	0.3792	0.7111
	ω	0.6485	0.6361	0.0693	0.5157	0.7704
	p	0.8096	0.8216	0.0938	0.6331	0.9874
100	θ	0.5133	0.5213	0.0648	0.4083	0.6353
	ω	0.6522	0.6398	0.0458	0.5576	0.7410
	p	0.8071	0.8191	0.0624	0.6822	0.9315
200	θ	0.5074	0.5154	0.0424	0.4323	0.5907
	ω	0.6521	0.6397	0.0316	0.5853	0.7161
	p	0.8041	0.8161	0.0447	0.7160	0.8919
500	θ	0.5037	0.5117	0.0245	0.4557	0.5551
	ω	0.6520	0.6396	0.0224	0.6098	0.6929
	p	0.8020	0.8140	0.0283	0.7463	0.8575

Source: Elaborated by the author.

Table 50 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-inflated case with $\theta = 5.00$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	6.0677	6.0790	3.8594	1.0101	22.0838
	ω	0.0545	0.0683	0.0245	0.0122	0.1287
	p	0.3926	0.3946	0.4136	0.0449	1.5928
100	θ	7.6378	7.6491	5.5890	1.4136	27.0622
	ω	0.0437	0.0575	0.0173	0.0141	0.0899
	p	0.3759	0.3779	0.3758	0.0560	1.4323
200	θ	9.4031	9.4144	8.2113	2.0209	31.7021
	ω	0.0397	0.0535	0.0141	0.0178	0.0701
	p	0.3908	0.3928	0.3886	0.0758	1.3746
500	θ	7.7461	7.7574	6.9236	2.6041	20.7141
	ω	0.0384	0.0522	0.0100	0.0235	0.0567
	p	0.3023	0.3043	0.2449	0.0990	0.8165
Scenario 2						
50	θ	9.0657	9.0770	7.5531	1.9686	30.4054
	ω	0.1550	0.1688	0.0529	0.0726	0.2628
	p	1.4478	1.4498	1.3599	0.2960	4.9892
100	θ	8.6657	8.6770	9.0715	2.5008	25.5350
	ω	0.1536	0.1674	0.0374	0.0912	0.2289
	p	1.3323	1.3343	1.3198	0.3829	3.9335
200	θ	6.8732	6.8845	6.5583	2.8888	15.7468
	ω	0.1521	0.1659	0.0245	0.1063	0.2045
	p	1.0692	1.0712	0.8845	0.4488	2.4440
500	θ	5.5331	5.5444	1.5535	3.3652	8.9952
	ω	0.1512	0.1650	0.0141	0.1213	0.1837
	p	0.8756	0.8776	0.2394	0.5295	1.4302

Source: Elaborated by the author.

Table 51 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 0.50$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
			Scenario 1			Scenario 2	
50	θ	4.00	93.40	2.60	3.60	93.00	3.40
	ω	2.00	94.20	3.80	3.40	94.00	2.60
	p	2.40	95.40	2.20	5.00	92.80	2.20
100	θ	1.80	96.00	2.20	3.20	94.40	2.40
	ω	2.60	94.00	3.40	3.20	94.80	2.00
	p	2.40	94.80	2.80	3.00	95.80	1.20
200	θ	3.00	94.40	2.60	3.40	94.60	2.00
	ω	1.20	96.60	2.20	2.20	95.40	2.40
	p	2.00	96.40	1.60	2.00	96.00	2.00
500	θ	2.20	95.40	2.40	1.20	96.20	2.60
	ω	2.00	95.20	2.80	2.60	95.20	2.20
	p	2.80	94.20	3.00	3.00	94.80	2.20

Source: Elaborated by the author.

Table 52 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPL}$ model, $\theta = 5.00$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
			Scenario 3			Scenario 4	
50	θ	0.00	91.20	8.80	1.60	93.40	5.00
	ω	5.20	94.80	0.00	3.20	92.00	4.80
	p	3.20	96.00	0.80	3.00	91.20	5.80
100	θ	0.00	93.20	6.80	5.60	91.00	3.40
	ω	3.40	96.60	0.00	2.40	94.00	3.60
	p	2.60	94.60	2.80	5.00	92.00	3.00
200	θ	2.40	92.20	5.40	3.20	94.20	2.60
	ω	0.80	96.20	3.00	1.20	96.80	2.00
	p	3.20	92.00	4.80	3.80	93.60	2.60
500	θ	3.00	93.00	4.00	2.80	95.20	2.00
	ω	2.00	95.60	2.40	1.40	96.40	2.20
	p	4.00	91.00	5.00	3.20	94.40	2.40

Source: Elaborated by the author.

Table 53 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 0.50$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	θ	0.0848	0.0402	0.0474	30.9839
	ω	0.0039	0.0029	0.0029	25.2611
	p	0.0174	0.0045	0.0048	27.0907
100	θ	0.0377	0.0153	0.0167	18.5781
	ω	0.0026	0.0015	0.0015	18.2788
	p	0.0088	0.0022	0.0023	19.1378
200	θ	0.0214	0.0072	0.0076	12.9463
	ω	0.0009	0.0007	0.0007	12.5279
	p	0.0043	0.0011	0.0011	13.1377
500	θ	0.0084	0.0024	0.0025	7.9089
	ω	−0.0003	0.0003	0.0003	7.6647
	p	0.0009	0.0004	0.0004	8.0367
			Scenario 2		
50	θ	0.0226	0.0070	0.0076	13.3959
	ω	−0.0039	0.0046	0.0046	7.8656
	p	0.0090	0.0078	0.0079	8.7908
100	θ	0.0107	0.0034	0.0035	9.2027
	ω	−0.0003	0.0021	0.0021	5.3950
	p	0.0062	0.0035	0.0035	6.0171
200	θ	0.0055	0.0015	0.0015	6.0411
	ω	−0.0004	0.0010	0.0010	3.7801
	p	0.0029	0.0018	0.0018	4.3054
500	θ	0.0029	0.0005	0.0005	3.6587
	ω	−0.0001	0.0004	0.0004	2.4419
	p	0.0015	0.0007	0.0007	2.6882

Source: Elaborated by the author.

Table 54 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 5.00$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 3		
50	θ	−0.3298	8.3335	8.4422	44.6405
	ω	0.0170	0.0005	0.0008	60.0575
	p	0.1139	0.1030	0.1160	97.1146
100	θ	1.5402	21.4820	23.8543	69.5817
	ω	0.0062	0.0003	0.0003	40.4844
	p	0.1272	0.1014	0.1176	104.3386
200	θ	4.1031	61.3412	78.1763	111.2650
	ω	0.0017	0.0002	0.0002	29.4944
	p	0.1725	0.1358	0.1656	117.6234
500	θ	3.5390	79.4106	91.9354	92.1415
	ω	0.0005	0.0001	0.0001	18.2152
	p	0.1200	0.0869	0.1013	81.7623
			Scenario 4		
50	θ	3.6419	48.2332	61.4965	101.3040
	ω	0.0046	0.0025	0.0025	28.8019
	p	0.5587	1.4186	1.7307	100.0529
100	θ	4.2056	94.2202	111.9070	106.9576
	ω	0.0026	0.0012	0.0013	20.3847
	p	0.5749	1.8067	2.1373	95.2501
200	θ	2.2536	43.1892	48.2678	62.2680
	ω	0.0006	0.0006	0.0006	14.0695
	p	0.2952	0.6539	0.7410	54.7191
500	θ	0.6656	3.1470	3.5900	27.0701
	ω	−0.0002	0.0002	0.0002	8.7723
	p	0.0888	0.0688	0.0767	25.0646

Source: Elaborated by the author.

Table 55 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-inflated case with $\theta = 0.50$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	0.5848	0.5893	0.2005	0.3045	1.0112
	ω	0.1728	0.1633	0.0539	0.0852	0.2846
	p	0.2174	0.2080	0.0671	0.1035	0.3750
100	θ	0.5377	0.5422	0.1237	0.3459	0.7907
	ω	0.1714	0.1619	0.0387	0.1056	0.2496
	p	0.2088	0.1994	0.0469	0.1264	0.3112
200	θ	0.5214	0.5259	0.0849	0.3844	0.6866
	ω	0.1698	0.1603	0.0265	0.1215	0.2242
	p	0.2043	0.1949	0.0332	0.1445	0.2735
500	θ	0.5084	0.5129	0.0490	0.4210	0.6066
	ω	0.1685	0.1590	0.0173	0.1372	0.2025
	p	0.2009	0.1915	0.0200	0.1624	0.2431
Scenario 2						
50	θ	0.5226	0.5271	0.0837	0.3861	0.6870
	ω	0.6717	0.6622	0.0678	0.5401	0.7906
	p	0.8090	0.7996	0.0883	0.6405	0.9767
100	θ	0.5107	0.5152	0.0583	0.4139	0.6210
	ω	0.6753	0.6658	0.0458	0.5820	0.7623
	p	0.8062	0.7968	0.0592	0.6876	0.9235
200	θ	0.5055	0.5100	0.0387	0.4363	0.5813
	ω	0.6752	0.6657	0.0316	0.6093	0.7377
	p	0.8029	0.7935	0.0424	0.7192	0.8857
500	θ	0.5029	0.5074	0.0224	0.4586	0.5499
	ω	0.6754	0.6659	0.0200	0.6339	0.7156
	p	0.8015	0.7921	0.0265	0.7486	0.8539

Source: Elaborated by the author.

Table 56 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-inflated case with $\theta = 5.00$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	4.6702	4.6747	2.8868	0.8488	16.9624
	ω	0.0514	0.0419	0.0224	0.0108	0.1240
	p	0.3139	0.3045	0.3209	0.0367	1.2388
100	θ	6.5402	6.5447	4.6349	1.2405	23.4675
	ω	0.0406	0.0311	0.0173	0.0125	0.0855
	p	0.3272	0.3178	0.3184	0.0487	1.2367
200	θ	9.1031	9.1076	7.8321	1.8537	31.7380
	ω	0.0361	0.0266	0.0141	0.0155	0.0654
	p	0.3725	0.3631	0.3685	0.0689	1.3261
500	θ	8.5390	8.5435	8.9113	2.5188	25.0256
	ω	0.0349	0.0254	0.0100	0.0208	0.0526
	p	0.3200	0.3106	0.2948	0.0949	0.9228
Scenario 2						
50	θ	8.6419	8.6464	6.9450	1.7958	29.8891
	ω	0.1422	0.1327	0.0500	0.0637	0.2470
	p	1.3587	1.3493	1.1910	0.2672	4.7205
100	θ	9.2056	9.2101	9.7067	2.4013	28.8794
	ω	0.1403	0.1308	0.0346	0.0808	0.2131
	p	1.3749	1.3655	1.3441	0.3645	4.2288
200	θ	7.2536	7.2581	6.5718	2.7725	17.9926
	ω	0.1382	0.1287	0.0245	0.0944	0.1887
	p	1.0952	1.0858	0.8086	0.4298	2.6290
500	θ	5.6656	5.6701	1.7740	3.2407	9.8288
	ω	0.1374	0.1279	0.0141	0.1088	0.1688
	p	0.8888	0.8794	0.2623	0.5133	1.5163

Source: Elaborated by the author.

Table 57 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 0.50$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	θ	3.40	94.00	2.60	4.00	92.80	3.20
	ω	2.60	94.20	3.20	3.80	92.40	3.80
	p	2.40	96.20	1.40	4.20	93.20	2.60
100	θ	3.00	95.00	2.00	3.20	94.40	2.40
	ω	2.40	95.00	2.60	3.20	94.60	2.20
	p	2.80	94.80	2.40	3.00	95.20	1.80
200	θ	2.80	95.00	2.20	3.00	95.40	1.60
	ω	1.60	96.20	2.20	2.20	95.60	2.20
	p	2.40	95.80	1.80	2.60	94.80	2.60
500	θ	2.20	95.80	2.00	1.20	96.00	2.80
	ω	2.00	95.00	3.00	2.40	95.80	1.80
	p	2.60	94.40	3.00	2.80	94.60	2.60

Source: Elaborated by the author.

Table 58 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 5.00$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	θ	0.00	89.40	10.60	0.20	94.40	5.40
	ω	4.20	95.80	0.00	3.00	94.40	2.60
	p	2.20	96.80	1.00	2.60	91.60	5.80
100	θ	0.00	92.80	7.20	5.80	91.00	3.20
	ω	3.40	96.60	0.00	3.60	93.00	3.40
	p	0.80	95.00	4.20	5.00	92.00	3.00
200	θ	2.00	92.20	5.80	4.20	93.20	2.60
	ω	1.40	94.20	4.40	2.40	94.40	3.20
	p	3.80	90.80	5.40	4.00	92.20	3.80
500	θ	4.40	92.00	3.60	2.80	94.80	2.40
	ω	2.80	95.20	2.00	2.20	95.80	2.00
	p	3.40	92.20	4.40	2.20	95.40	2.40

Source: Elaborated by the author.

Table 59 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 0.50$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	θ	0.0564	0.0226	0.0258	23.3386
	ω	0.0037	0.0032	0.0032	24.2683
	p	0.0096	0.0038	0.0039	24.6519
100	θ	0.0253	0.0088	0.0094	14.1633
	ω	0.0026	0.0016	0.0016	17.7310
	p	0.0054	0.0019	0.0020	17.9916
200	θ	0.0155	0.0042	0.0045	9.9899
	ω	0.0014	0.0008	0.0008	11.9265
	p	0.0031	0.0009	0.0009	12.0260
500	θ	0.0066	0.0015	0.0016	6.3728
	ω	−0.0001	0.0003	0.0003	7.5151
	p	0.0006	0.0004	0.0004	7.6590
			Scenario 2		
50	θ	0.0142	0.0042	0.0044	10.4302
	ω	−0.0065	0.0042	0.0042	6.7847
	p	−0.0015	0.0051	0.0051	7.0153
100	θ	0.0056	0.0020	0.0020	7.0637
	ω	−0.0032	0.0019	0.0019	4.7174
	p	−0.0011	0.0023	0.0023	4.8888
200	θ	0.0034	0.0009	0.0009	4.6752
	ω	−0.0012	0.0010	0.0010	3.3366
	p	0.0001	0.0012	0.0012	3.4981
500	θ	0.0017	0.0003	0.0003	3.0550
	ω	−0.0008	0.0004	0.0004	2.0563
	p	−0.0003	0.0005	0.0005	2.1356

Source: Elaborated by the author.

Table 60 – Empirical properties of the Bayesian estimators using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 5.00$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 3		
50	θ	0.8989	11.9023	12.7103	51.0886
	ω	0.0155	0.0006	0.0009	54.7119
	p	0.1682	0.1451	0.1734	116.1277
100	θ	2.3809	26.7883	32.4568	77.7677
	ω	0.0053	0.0004	0.0004	38.2572
	p	0.1652	0.1345	0.1618	116.7277
200	θ	3.7150	56.7867	70.5876	98.3843
	ω	0.0017	0.0002	0.0002	26.4489
	p	0.1726	0.1450	0.1748	113.4977
500	θ	2.0941	37.3556	41.7409	59.9110
	ω	0.0003	0.0001	0.0001	16.4346
	p	0.0863	0.0595	0.0669	63.4104
			Scenario 4		
50	θ	3.1828	42.1853	52.3154	87.1116
	ω	0.0035	0.0029	0.0029	26.2223
	p	0.5437	1.4252	1.7207	94.5984
100	θ	2.6997	56.4750	63.7636	72.7466
	ω	0.0023	0.0014	0.0015	18.7618
	p	0.4311	1.3626	1.5485	74.7180
200	θ	1.4198	24.7754	26.7912	42.6647
	ω	0.0008	0.0007	0.0007	12.7431
	p	0.2313	0.5806	0.6341	45.0595
500	θ	0.4012	1.5921	1.7531	18.4887
	ω	−0.0006	0.0003	0.0003	7.8841
	p	0.0638	0.0490	0.0531	20.4783

Source: Elaborated by the author.

Table 61 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-inflated case with $\theta = 0.50$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	0.5564	0.5568	0.1503	0.3314	0.8744
	ω	0.1882	0.1984	0.0566	0.0964	0.3030
	p	0.2096	0.2057	0.0616	0.1063	0.3419
100	θ	0.5253	0.5257	0.0938	0.3702	0.7205
	ω	0.1871	0.1973	0.0400	0.1185	0.2675
	p	0.2054	0.2015	0.0436	0.1294	0.2954
200	θ	0.5155	0.5159	0.0648	0.4049	0.6452
	ω	0.1860	0.1962	0.0283	0.1356	0.2421
	p	0.2031	0.1992	0.0300	0.1477	0.2653
500	θ	0.5066	0.5070	0.0387	0.4360	0.5845
	ω	0.1844	0.1946	0.0173	0.1519	0.2194
	p	0.2006	0.1967	0.0200	0.1649	0.2390
Scenario 2						
50	θ	0.5142	0.5146	0.0648	0.4044	0.6426
	ω	0.7315	0.7417	0.0648	0.6051	0.8410
	p	0.7985	0.7946	0.0714	0.6577	0.9247
100	θ	0.5056	0.5060	0.0447	0.4274	0.5928
	ω	0.7349	0.7451	0.0436	0.6455	0.8154
	p	0.7989	0.7950	0.0480	0.6997	0.8902
200	θ	0.5034	0.5038	0.0300	0.4477	0.5637
	ω	0.7369	0.7471	0.0316	0.6743	0.7950
	p	0.8001	0.7962	0.0346	0.7306	0.8655
500	θ	0.5017	0.5021	0.0173	0.4660	0.5392
	ω	0.7372	0.7474	0.0200	0.6979	0.7747
	p	0.7997	0.7958	0.0224	0.7561	0.8417

Source: Elaborated by the author.

Table 62 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-inflated case with $\theta = 5.00$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	5.8989	5.9115	3.4500	1.1673	20.6402
	ω	0.0564	0.0501	0.0245	0.0131	0.1315
	p	0.3682	0.3764	0.3809	0.0433	1.4863
100	θ	7.3809	7.3935	5.1757	1.6252	25.1183
	ω	0.0462	0.0399	0.0200	0.0155	0.0934
	p	0.3652	0.3734	0.3667	0.0564	1.3757
200	θ	8.7150	8.7276	7.5357	2.2390	27.9090
	ω	0.0425	0.0362	0.0141	0.0197	0.0739
	p	0.3726	0.3808	0.3808	0.0767	1.2756
500	θ	7.0941	7.1067	6.1119	2.8360	17.3709
	ω	0.0412	0.0349	0.0100	0.0257	0.0601
	p	0.2863	0.2945	0.2439	0.1010	0.7348
Scenario 2						
50	θ	8.1828	8.1954	6.4950	2.1716	25.7631
	ω	0.1669	0.1606	0.0539	0.0809	0.2773
	p	1.3437	1.3519	1.1938	0.2979	4.4706
100	θ	7.6997	7.7123	7.5150	2.7085	20.7617
	ω	0.1657	0.1594	0.0374	0.1010	0.2430
	p	1.2311	1.2393	1.1673	0.3875	3.4284
200	θ	6.4198	6.4324	4.9775	3.1480	13.2801
	ω	0.1642	0.1579	0.0265	0.1167	0.2180
	p	1.0313	1.0395	0.7620	0.4647	2.2119
500	θ	5.4012	5.4138	1.2618	3.5952	8.1502
	ω	0.1629	0.1566	0.0173	0.1320	0.1963
	p	0.8638	0.8720	0.2214	0.5439	1.3539

Source: Elaborated by the author.

Table 63 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 0.50$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	θ	3.60	94.00	2.40	4.00	92.80	3.20
	ω	3.80	91.80	4.40	4.60	92.20	3.20
	p	3.80	93.40	2.80	4.60	92.40	3.00
100	θ	3.20	94.60	2.20	3.40	93.60	3.00
	ω	2.40	93.80	3.80	3.00	94.40	2.60
	p	2.60	94.40	3.00	2.80	95.00	2.20
200	θ	2.00	95.80	2.20	2.60	95.80	1.60
	ω	2.40	95.00	2.60	3.60	94.80	1.60
	p	2.80	95.20	2.00	3.20	94.40	2.40
500	θ	2.40	96.00	1.60	1.80	96.40	1.80
	ω	2.40	95.60	2.00	2.60	95.20	2.20
	p	2.60	95.00	2.40	3.00	95.20	1.80

Source: Elaborated by the author.

Table 64 – Coverage probabilities (%) of the BCIs using zero-inflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 5.00$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	θ	0.00	94.40	5.60	1.60	93.60	4.80
	ω	2.40	97.60	0.00	2.20	94.00	3.80
	p	2.80	96.00	1.20	2.40	93.00	4.60
100	θ	0.00	94.40	5.60	4.40	92.60	3.00
	ω	3.20	96.80	0.00	2.60	94.20	3.20
	p	3.20	93.80	3.00	3.80	92.80	3.40
200	θ	2.40	93.00	4.60	3.40	94.20	2.40
	ω	1.40	94.80	3.80	1.00	96.60	2.40
	p	2.80	93.00	4.20	4.00	92.80	3.20
500	θ	4.00	92.40	3.60	2.60	95.40	2.00
	ω	2.00	95.00	3.00	1.80	95.20	3.00
	p	3.20	92.00	4.80	3.00	94.40	2.60

Source: Elaborated by the author.

A.3.1.2 Zero-deflated artificial data

The following tables illustrate the results obtained in the first simulation study, considering scenarios 1-4 of the zero-deflated case as described in Section 2.6 (Table 6).

Table 65 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 2.50$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	θ	0.5872	2.2849	2.6296	38.7108
	ω	−0.0036	0.0056	0.0056	12.3448
	p	0.2543	0.4621	0.5268	33.6599
100	θ	0.2517	0.5681	0.6315	22.0659
	ω	0.0009	0.0027	0.0027	8.6409
	p	0.1165	0.1219	0.1355	19.5120
200	θ	0.1320	0.2202	0.2376	14.6123
	ω	0.0009	0.0012	0.0012	5.7611
	p	0.0646	0.0541	0.0583	13.3916
500	θ	0.0615	0.0700	0.0738	8.5006
	ω	0.0006	0.0005	0.0005	3.7450
	p	0.0305	0.0197	0.0206	8.0638
			Scenario 2		
50	θ	0.2958	0.6334	0.7209	24.1878
	ω	−0.0059	0.0033	0.0034	5.5602
	p	0.2167	0.3921	0.4391	19.8135
100	θ	0.1347	0.2598	0.2780	15.3883
	ω	−0.0030	0.0013	0.0013	3.4556
	p	0.0979	0.1568	0.1664	12.4787
200	θ	0.0697	0.1086	0.1134	9.9585
	ω	−0.0020	0.0008	0.0008	2.6564
	p	0.0498	0.0706	0.0731	8.4424
500	θ	0.0298	0.0367	0.0376	6.1903
	ω	−0.0016	0.0003	0.0003	1.7113
	p	0.0191	0.0241	0.0245	5.2450

Source: Elaborated by the author.

Table 66 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 6.00$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 3		
50	θ	5.1449	104.0101	130.4798	110.0221
	ω	0.0024	0.0038	0.0038	21.9213
	p	1.1601	6.0880	7.4338	108.0458
100	θ	4.0243	124.7600	140.9553	87.1485
	ω	0.0022	0.0020	0.0020	15.8825
	p	0.8806	5.7641	6.5396	83.5732
200	θ	1.9601	51.2473	55.0892	48.1096
	ω	0.0008	0.0009	0.0009	10.9028
	p	0.4331	2.3479	2.5355	46.7987
500	θ	0.5535	2.8691	3.1754	20.6722
	ω	−0.0004	0.0004	0.0004	6.7408
	p	0.1217	0.1561	0.1709	21.2245
			Scenario 4		
50	θ	4.0865	105.3742	122.0737	89.6096
	ω	0.0015	0.0051	0.0051	14.7628
	p	1.5751	17.0394	19.5203	86.5442
100	θ	1.7836	47.6696	50.8508	46.6722
	ω	0.0023	0.0028	0.0028	10.8051
	p	0.7057	8.4348	8.9327	46.1133
200	θ	0.8054	6.4923	7.1410	27.1256
	ω	−0.0001	0.0012	0.0012	7.0980
	p	0.3077	0.8990	0.9937	27.0022
500	θ	0.3258	1.2229	1.3290	14.6293
	ω	−0.0001	0.0005	0.0005	4.5344
	p	0.1257	0.1985	0.2143	14.7487

Source: Elaborated by the author.

Table 67 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-deflated case with $\theta = 2.50$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	3.0872	3.1010	1.5116	1.5622	5.8608
	ω	0.4780	0.4798	0.0748	0.3453	0.6123
	p	1.6543	1.6465	0.6798	0.8902	3.0310
100	θ	2.7517	2.7655	0.7537	1.7443	4.2541
	ω	0.4825	0.4843	0.0520	0.3868	0.5790
	p	1.5165	1.5087	0.3491	1.0023	2.2782
200	θ	2.6320	2.6458	0.4693	1.9193	3.5687
	ω	0.4826	0.4844	0.0346	0.4141	0.5514
	p	1.4646	1.4568	0.2326	1.0988	1.9444
500	θ	2.5615	2.5753	0.2646	2.1044	3.1041
	ω	0.4822	0.4840	0.0224	0.4386	0.5259
	p	1.4305	1.4227	0.1404	1.1956	1.7090
Scenario 2						
50	θ	2.7958	2.8096	0.7959	1.7079	4.4682
	ω	0.8198	0.8216	0.0574	0.7067	0.9092
	p	2.6167	2.6089	0.6262	1.7507	3.9608
100	θ	2.6347	2.6485	0.5097	1.8749	3.6565
	ω	0.8227	0.8245	0.0361	0.7436	0.8898
	p	2.4979	2.4901	0.3960	1.8921	3.3185
200	θ	2.5697	2.5835	0.3295	2.0261	3.2393
	ω	0.8236	0.8254	0.0283	0.7685	0.8727
	p	2.4498	2.4420	0.2657	2.0163	2.9869
500	θ	2.5298	2.5436	0.1916	2.1795	2.9285
	ω	0.8241	0.8259	0.0173	0.7897	0.8560
	p	2.4191	2.4113	0.1552	2.1396	2.7389

Source: Elaborated by the author.

Table 68 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPL}$ model (zero-deflated case with $\theta = 6.00$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	11.1449	11.1514	10.1985	2.6134	36.4116
	ω	0.2269	0.2156	0.0616	0.1260	0.3481
	p	2.5601	2.5554	2.4674	0.5852	8.4908
100	θ	10.0243	10.0308	11.1696	3.2102	27.9620
	ω	0.2267	0.2154	0.0447	0.1515	0.3121
	p	2.2806	2.2759	2.4009	0.7325	6.3537
200	θ	7.9601	7.9666	7.1587	3.6697	16.9878
	ω	0.2253	0.2140	0.0300	0.1707	0.2851
	p	1.8331	1.8284	1.5323	0.8456	3.9126
500	θ	6.5535	6.5600	1.6938	4.2039	10.1489
	ω	0.2241	0.2128	0.0200	0.1888	0.2614
	p	1.5217	1.5170	0.3951	0.9725	2.3654
Scenario 2						
50	θ	10.0865	10.0930	10.2652	3.0750	28.8456
	ω	0.3863	0.3750	0.0714	0.2605	0.5207
	p	3.9751	3.9704	4.1279	1.2230	11.3772
100	θ	7.7836	7.7901	6.9043	3.4684	16.9435
	ω	0.3871	0.3758	0.0529	0.2956	0.4830
	p	3.1057	3.1010	2.9043	1.3970	6.7629
200	θ	6.8054	6.8119	2.5480	3.9472	11.6039
	ω	0.3847	0.3734	0.0346	0.3190	0.4525
	p	2.7077	2.7030	0.9482	1.5792	4.6128
500	θ	6.3258	6.3323	1.1058	4.5316	8.7902
	ω	0.3848	0.3735	0.0224	0.3427	0.4276
	p	2.5257	2.5210	0.4455	1.8143	3.5074

Source: Elaborated by the author.

Table 69 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 2.50$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	θ	4.80	91.80	3.40	3.40	93.40	3.20
	ω	3.20	92.40	4.40	3.20	94.00	2.80
	p	4.00	92.20	3.80	3.60	93.60	2.80
100	θ	3.20	94.20	2.60	2.60	95.00	2.40
	ω	4.00	93.60	2.40	2.00	95.80	2.20
	p	3.60	93.60	2.80	3.00	94.40	2.60
200	θ	4.00	93.80	2.20	3.20	95.00	1.80
	ω	2.00	95.60	2.40	2.60	94.40	3.00
	p	3.60	93.20	3.20	2.80	94.40	2.80
500	θ	3.80	94.00	2.20	2.00	94.80	3.20
	ω	1.60	95.80	2.60	1.80	94.80	3.40
	p	4.40	93.00	2.60	2.60	94.20	3.20

Source: Elaborated by the author.

Table 70 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPL}$ model, $\theta = 6.00$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	θ	4.80	90.60	4.60	3.60	92.40	4.00
	ω	3.20	93.40	3.40	3.40	93.00	3.60
	p	4.60	91.20	4.20	4.40	91.20	4.40
100	θ	5.60	90.80	3.60	3.20	92.80	4.00
	ω	3.40	92.60	4.00	4.60	91.20	4.20
	p	5.40	90.60	4.00	3.00	94.00	3.00
200	θ	4.00	93.40	2.60	2.40	93.80	3.80
	ω	2.20	94.60	3.20	2.40	94.40	3.20
	p	4.80	92.40	2.80	2.60	93.20	4.20
500	θ	3.00	95.20	1.80	2.20	94.60	3.20
	ω	4.20	92.40	3.40	2.00	94.80	3.20
	p	2.60	94.60	2.80	4.00	92.60	3.40

Source: Elaborated by the author.

Table 71 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 2.50$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	θ	0.6431	2.6288	3.0424	40.7919
	ω	−0.0026	0.0052	0.0052	13.0551
	p	0.2704	0.5382	0.6113	35.5518
100	θ	0.2778	0.6263	0.7034	22.7574
	ω	0.0008	0.0026	0.0026	9.2075
	p	0.1246	0.1419	0.1574	20.8770
200	θ	0.1470	0.2489	0.2705	14.9533
	ω	0.0003	0.0012	0.0012	6.2788
	p	0.0685	0.0649	0.0696	14.4486
500	θ	0.0630	0.0794	0.0833	8.9057
	ω	0.0001	0.0005	0.0005	3.9387
	p	0.0299	0.0237	0.0246	8.6947
			Scenario 2		
50	θ	0.3842	1.1667	1.3143	28.1364
	ω	−0.0061	0.0040	0.0041	6.5609
	p	0.2752	0.7434	0.8192	23.9563
100	θ	0.1623	0.3541	0.3805	16.8785
	ω	−0.0037	0.0018	0.0018	4.5426
	p	0.1136	0.2357	0.2486	14.4426
200	θ	0.0774	0.1230	0.1290	10.8767
	ω	−0.0017	0.0010	0.0010	3.2379
	p	0.0557	0.0912	0.0943	9.8702
500	θ	0.0353	0.0398	0.0410	6.3614
	ω	−0.0012	0.0003	0.0004	1.9701
	p	0.0241	0.0301	0.0307	5.7739

Source: Elaborated by the author.

Table 72 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 6.00$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 3		
50	θ	5.2632	103.0505	130.7522	114.6161
	ω	0.0025	0.0036	0.0036	23.4158
	p	1.1039	5.4860	6.7046	107.4200
100	θ	4.7599	160.4740	183.1310	100.7584
	ω	0.0019	0.0018	0.0019	16.7770
	p	0.9399	6.4959	7.3793	88.7662
200	θ	2.4685	84.6359	90.7293	57.1014
	ω	0.0007	0.0009	0.0009	11.4027
	p	0.4850	2.9058	3.1411	50.6630
500	θ	0.6439	3.6796	4.0943	22.8275
	ω	−0.0003	0.0003	0.0003	7.0875
	p	0.1288	0.1732	0.1898	22.0107
			Scenario 4		
50	θ	5.6309	174.5981	206.3046	116.8171
	ω	0.0018	0.0049	0.0049	15.8665
	p	1.9876	24.0226	27.9730	105.9069
100	θ	2.4184	69.2322	75.0807	58.5291
	ω	0.0021	0.0026	0.0026	11.3873
	p	0.8590	9.4644	10.2023	53.7601
200	θ	1.1560	13.2861	14.6224	33.6414
	ω	0.0004	0.0012	0.0012	7.7537
	p	0.4053	1.5395	1.7037	31.2498
500	θ	0.4167	1.7525	1.9261	16.9115
	ω	0.0000	0.0005	0.0005	4.8717
	p	0.1482	0.2586	0.2806	16.3072

Source: Elaborated by the author.

Table 73 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-deflated case with $\theta = 2.50$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	3.1431	3.1551	1.6214	1.5440	6.3161
	ω	0.4368	0.4451	0.0721	0.3066	0.5719
	p	1.6704	1.6608	0.7336	0.8400	3.2247
100	θ	2.7778	2.7898	0.7914	1.7307	4.4493
	ω	0.4402	0.4485	0.0510	0.3457	0.5368
	p	1.5246	1.5150	0.3767	0.9612	2.3854
200	θ	2.6470	2.6590	0.4989	1.9081	3.6658
	ω	0.4397	0.4480	0.0346	0.3721	0.5084
	p	1.4685	1.4589	0.2548	1.0660	2.0069
500	θ	2.5630	2.5750	0.2818	2.0912	3.1394
	ω	0.4395	0.4478	0.0224	0.3965	0.4831
	p	1.4299	1.4203	0.1539	1.1718	1.7398
Scenario 2						
50	θ	2.8842	2.8962	1.0801	1.7100	4.8506
	ω	0.7472	0.7555	0.0632	0.6227	0.8536
	p	2.6752	2.6656	0.8622	1.6777	4.2827
100	θ	2.6623	2.6743	0.5951	1.8679	3.7900
	ω	0.7496	0.7579	0.0424	0.6619	0.8280
	p	2.5136	2.5040	0.4855	1.8282	3.4597
200	θ	2.5774	2.5894	0.3507	2.0153	3.2941
	ω	0.7515	0.7598	0.0316	0.6900	0.8084
	p	2.4557	2.4461	0.3020	1.9673	3.0660
500	θ	2.5353	2.5473	0.1995	2.1727	2.9569
	ω	0.7520	0.7603	0.0173	0.7134	0.7887
	p	2.4241	2.4145	0.1735	2.1088	2.7865

Source: Elaborated by the author.

Table 74 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_h$ model (zero-deflated case with $\theta = 6.00$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	11.2632	11.2640	10.1514	2.4288	38.3828
	ω	0.2072	0.2148	0.0600	0.1107	0.3253
	p	2.5039	2.4977	2.3422	0.5412	8.4682
100	θ	10.7599	10.7607	12.6678	3.0569	32.2883
	ω	0.2066	0.2142	0.0424	0.1345	0.2896
	p	2.3399	2.3337	2.5487	0.6935	6.8301
200	θ	8.4685	8.4693	9.1998	3.5072	19.8132
	ω	0.2054	0.2130	0.0300	0.1528	0.2634
	p	1.8850	1.8788	1.7046	0.8107	4.2752
500	θ	6.6439	6.6447	1.9182	4.0268	10.8969
	ω	0.2043	0.2119	0.0173	0.1703	0.2406
	p	1.5288	1.5226	0.4162	0.9415	2.4659
Scenario 2						
50	θ	11.6309	11.6317	13.2136	3.0316	36.4990
	ω	0.3526	0.3602	0.0700	0.2305	0.4856
	p	4.3876	4.3814	4.9013	1.1980	13.4623
100	θ	8.4184	8.4192	8.3206	3.3456	20.1552
	ω	0.3529	0.3605	0.0510	0.2637	0.4475
	p	3.2590	3.2528	3.0764	1.3496	7.5749
200	θ	7.1560	7.1568	3.6450	3.8282	13.2586
	ω	0.3512	0.3588	0.0346	0.2870	0.4181
	p	2.8053	2.7991	1.2408	1.5454	5.0653
500	θ	6.4167	6.4175	1.3238	4.3964	9.3340
	ω	0.3508	0.3584	0.0224	0.3098	0.3929
	p	2.5482	2.5420	0.5085	1.7765	3.6462

Source: Elaborated by the author.

Table 75 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 2.50$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	θ	3.80	92.60	3.60	4.20	92.60	3.20
	ω	4.20	91.60	4.20	3.40	93.00	3.60
	p	2.40	94.20	3.40	3.60	93.60	2.80
100	θ	2.60	95.00	2.40	3.60	93.60	2.80
	ω	3.20	94.60	2.20	2.80	95.20	2.00
	p	2.40	95.40	2.20	3.20	94.00	2.80
200	θ	3.80	93.00	3.20	3.20	94.40	2.40
	ω	1.40	96.20	2.40	3.60	94.00	2.40
	p	3.20	93.60	3.20	3.00	94.00	3.00
500	θ	3.40	94.60	2.00	2.80	94.40	2.80
	ω	1.60	96.20	2.20	2.40	95.20	2.40
	p	3.80	93.00	3.20	2.40	94.60	3.00

Source: Elaborated by the author.

Table 76 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_h$ model, $\theta = 6.00$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	θ	2.60	93.40	4.00	7.20	87.20	5.60
	ω	3.20	92.80	4.00	3.40	92.40	4.20
	p	3.80	91.40	4.80	6.20	89.20	4.60
100	θ	5.60	91.60	2.80	3.80	92.60	3.60
	ω	3.40	93.40	3.20	4.40	92.80	2.80
	p	4.80	92.20	3.00	3.60	92.80	3.60
200	θ	3.40	94.00	2.60	4.20	91.60	4.20
	ω	3.00	94.20	2.80	3.00	93.80	3.20
	p	4.20	93.20	2.60	4.00	91.80	4.20
500	θ	2.20	95.40	2.40	3.00	93.80	3.20
	ω	3.40	93.60	3.00	2.60	94.40	3.00
	p	2.00	95.80	2.20	2.80	93.20	4.00

Source: Elaborated by the author.

Table 77 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 2.50$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 1		
50	θ	0.3166	0.7945	0.8947	24.9405
	ω	−0.0030	0.0055	0.0055	10.4541
	p	0.1658	0.2344	0.2619	25.0004
100	θ	0.1468	0.2801	0.3016	15.7823
	ω	0.0007	0.0025	0.0025	7.0281
	p	0.0823	0.0836	0.0903	15.5576
200	θ	0.0784	0.1070	0.1131	10.2906
	ω	0.0008	0.0011	0.0011	4.8188
	p	0.0457	0.0337	0.0358	10.5361
500	θ	0.0392	0.0365	0.0380	6.0358
	ω	0.0006	0.0005	0.0005	3.0695
	p	0.0229	0.0129	0.0135	6.4015
			Scenario 2		
50	θ	0.1774	0.3151	0.3466	17.2343
	ω	−0.0068	0.0009	0.0009	2.4583
	p	0.1512	0.2399	0.2627	15.4492
100	θ	0.0781	0.1335	0.1396	11.0977
	ω	−0.0030	0.0004	0.0004	1.6877
	p	0.0675	0.0997	0.1043	9.9036
200	θ	0.0440	0.0520	0.0539	7.0624
	ω	−0.0014	0.0002	0.0002	1.2590
	p	0.0382	0.0405	0.0419	6.4454
500	θ	0.0209	0.0189	0.0193	4.4106
	ω	−0.0002	0.0001	0.0001	0.7784
	p	0.0189	0.0147	0.0150	4.0805

Source: Elaborated by the author.

Table 78 – Empirical properties of the Bayesian estimators using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 6.00$).

n	Parameter	Bias	Var	MSE	MAPE (%)
			Scenario 3		
50	θ	4.4422	88.5931	108.3263	96.1490
	ω	0.0026	0.0039	0.0039	21.1509
	p	1.0729	5.7678	6.9189	101.9241
100	θ	2.9044	76.8044	85.2397	66.2762
	ω	0.0017	0.0020	0.0020	14.8816
	p	0.6840	3.8959	4.3638	69.7217
200	θ	1.2096	11.5184	12.9814	34.0224
	ω	0.0011	0.0009	0.0009	10.2884
	p	0.2959	0.6304	0.7180	36.7814
500	θ	0.4511	2.0118	2.2154	17.3364
	ω	−0.0001	0.0004	0.0004	6.4957
	p	0.1097	0.1329	0.1449	19.4601
			Scenario 4		
50	θ	3.6216	95.9068	109.0225	79.9508
	ω	−0.0003	0.0052	0.0052	14.0394
	p	1.4694	17.2267	19.3858	82.0812
100	θ	1.4939	48.8965	51.1283	40.4640
	ω	0.0017	0.0028	0.0028	10.2507
	p	0.6362	10.4698	10.8746	42.5884
200	θ	0.6921	4.7636	5.2426	23.4989
	ω	0.0002	0.0012	0.0012	6.7406
	p	0.2859	0.7846	0.8663	25.2916
500	θ	0.2740	0.9269	1.0020	12.6203
	ω	−0.0002	0.0005	0.0005	4.2638
	p	0.1139	0.1801	0.1931	13.8640

Source: Elaborated by the author.

Table 79 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-deflated case with $\theta = 2.50$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	2.8166	2.8131	0.8913	1.7253	4.5418
	ω	0.5546	0.5560	0.0742	0.4198	0.6854
	p	1.5658	1.5543	0.4841	0.9405	2.5715
100	θ	2.6468	2.6433	0.5292	1.8972	3.6660
	ω	0.5583	0.5597	0.0500	0.4617	0.6527
	p	1.4823	1.4708	0.2891	1.0479	2.0817
200	θ	2.5784	2.5749	0.3271	2.0453	3.2381
	ω	0.5584	0.5598	0.0332	0.4897	0.6260
	p	1.4457	1.4342	0.1836	1.1359	1.8331
500	θ	2.5392	2.5357	0.1910	2.1965	2.9302
	ω	0.5582	0.5596	0.0224	0.5145	0.6013
	p	1.4229	1.4114	0.1136	1.2232	1.6527
Scenario 2						
50	θ	2.6774	2.6739	0.5613	1.8661	3.8085
	ω	0.9490	0.9504	0.0300	0.8785	0.9882
	p	2.5512	2.5397	0.4898	1.8554	3.5666
100	θ	2.5781	2.5746	0.3654	2.0070	3.2954
	ω	0.9529	0.9543	0.0200	0.9054	0.9838
	p	2.4675	2.4560	0.3158	1.9772	3.1053
200	θ	2.5440	2.5405	0.2280	2.1346	3.0244
	ω	0.9545	0.9559	0.0141	0.9222	0.9782
	p	2.4382	2.4267	0.2012	2.0858	2.8626
500	θ	2.5209	2.5174	0.1375	2.2581	2.8118
	ω	0.9556	0.9570	0.0100	0.9360	0.9717
	p	2.4189	2.4074	0.1212	2.1917	2.6744

Source: Elaborated by the author.

Table 80 – *Posterior* estimates of θ, ω, and p - $\mathcal{ZMPS}_u$ model (zero-deflated case with $\theta = 6.00$).

n	Parameter	Mean	Median	Std. Dev.	95% BCI	
					Lower	Upper
Scenario 1						
50	θ	10.4422	10.4448	9.4124	2.8623	32.5257
	ω	0.2404	0.2512	0.0624	0.1365	0.3635
	p	2.4729	2.4853	2.4016	0.5952	8.0047
100	θ	8.9044	8.9070	8.7638	3.4017	22.6765
	ω	0.2396	0.2504	0.0447	0.1626	0.3264
	p	2.0840	2.0964	1.9738	0.7321	5.4541
200	θ	7.2096	7.2122	3.3939	3.8561	13.6095
	ω	0.2389	0.2497	0.0300	0.1829	0.2999
	p	1.6959	1.7083	0.7940	0.8501	3.3083
500	θ	6.4511	6.4537	1.4184	4.4318	9.4385
	ω	0.2377	0.2485	0.0200	0.2016	0.2757
	p	1.5097	1.5221	0.3646	0.9950	2.2739
Scenario 2						
50	θ	9.6216	9.6242	9.7932	3.3481	26.1363
	ω	0.4074	0.4182	0.0721	0.2795	0.5422
	p	3.8694	3.8818	4.1505	1.2454	10.7738
100	θ	7.4939	7.4965	6.9926	3.7456	15.2235
	ω	0.4094	0.4202	0.0529	0.3166	0.5058
	p	3.0362	3.0486	3.2357	1.4294	6.3587
200	θ	6.6921	6.6947	2.1826	4.2119	10.7335
	ω	0.4079	0.4187	0.0346	0.3414	0.4762
	p	2.6859	2.6983	0.8858	1.6246	4.4158
500	θ	6.2740	6.2766	0.9628	4.7273	8.3450
	ω	0.4076	0.4184	0.0224	0.3651	0.4508
	p	2.5139	2.5263	0.4244	1.8472	3.4110

Source: Elaborated by the author.

Table 81 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 2.50$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	θ	4.00	92.80	3.20	4.00	92.60	3.40
	ω	3.40	93.20	3.40	0.00	97.60	2.40
	p	3.60	92.80	3.60	3.80	93.20	3.00
100	θ	2.80	95.00	2.20	3.20	93.80	3.00
	ω	3.80	94.80	1.40	3.00	94.00	3.00
	p	2.60	95.60	1.80	3.20	94.00	2.80
200	θ	3.40	93.40	3.20	3.60	94.00	2.40
	ω	1.80	96.80	1.40	2.60	94.80	2.60
	p	3.00	93.20	3.80	3.20	94.60	2.20
500	θ	3.80	92.60	3.60	2.20	95.40	2.40
	ω	3.20	94.00	2.80	3.40	94.40	2.20
	p	3.60	93.20	3.20	2.80	95.00	2.20

Source: Elaborated by the author.

Table 82 – Coverage probabilities (%) of the BCIs using zero-deflated samples ($\mathcal{ZMPS}_u$ model, $\theta = 6.00$).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	θ	5.00	90.00	5.00	4.00	92.80	3.20
	ω	3.60	92.80	3.60	3.00	93.80	3.20
	p	3.80	91.80	4.40	4.80	91.40	3.80
100	θ	4.40	91.60	4.00	2.40	94.20	3.40
	ω	3.20	92.80	4.00	3.60	93.20	3.20
	p	4.00	91.60	4.40	2.40	93.80	3.80
200	θ	3.60	93.80	2.60	2.80	94.00	3.20
	ω	2.80	94.60	2.60	2.20	95.60	2.20
	p	3.80	92.40	3.80	3.60	92.80	3.60
500	θ	2.80	95.40	1.80	2.60	95.00	2.40
	ω	3.40	93.80	2.80	1.60	94.80	3.60
	p	3.00	94.20	2.80	2.00	95.00	3.00

Source: Elaborated by the author.

A.3.2 Study 2

The following tables illustrate the results obtained in the second simulation study, considering scenarios 1-4 of zero-inflated and zero-deflated cases as described in Section 2.6 (Table 6).

Table 83 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPL}$ is the true model (zero inflation).

Scenario	n	Fitted	% DIC	% EAIC	% EBIC	% LMPL
1	50	$\mathcal{ZMPL}$	64.20	64.40	64.40	70.60
		$\mathcal{ZMP}$	35.80	35.60	35.60	29.40
	100	$\mathcal{ZMPL}$	82.60	82.60	82.60	85.80
		$\mathcal{ZMP}$	17.40	17.40	17.40	14.20
	200	$\mathcal{ZMPL}$	94.80	94.80	94.80	95.80
		$\mathcal{ZMP}$	5.20	5.20	5.20	4.20
	500	$\mathcal{ZMPL}$	99.80	99.80	99.80	100.00
		$\mathcal{ZMP}$	0.20	0.20	0.20	0.00
2	50	$\mathcal{ZMPL}$	95.80	95.80	95.80	96.20
		$\mathcal{ZMP}$	4.20	4.20	4.20	3.80
	100	$\mathcal{ZMPL}$	99.20	99.20	99.20	99.20
		$\mathcal{ZMP}$	0.80	0.80	0.80	0.80
	200	$\mathcal{ZMPL}$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
	500	$\mathcal{ZMPL}$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
3	50	$\mathcal{ZMPL}$	72.00	70.00	70.00	86.20
		$\mathcal{ZMP}$	28.00	30.00	30.00	13.80
	100	$\mathcal{ZMPL}$	60.40	54.00	54.00	74.20
		$\mathcal{ZMP}$	39.60	46.00	46.00	25.80
	200	$\mathcal{ZMPL}$	50.40	39.80	39.80	57.20
		$\mathcal{ZMP}$	49.60	60.20	60.20	42.80
	500	$\mathcal{ZMPL}$	37.40	34.00	34.00	43.00
		$\mathcal{ZMP}$	62.60	66.00	66.00	57.00
4	50	$\mathcal{ZMPL}$	47.00	38.00	38.00	57.40
		$\mathcal{ZMP}$	53.00	62.00	62.00	42.60
	100	$\mathcal{ZMPL}$	37.20	33.20	33.20	43.40
		$\mathcal{ZMP}$	62.80	66.80	66.80	56.60
	200	$\mathcal{ZMPL}$	42.20	40.80	40.80	49.00
		$\mathcal{ZMP}$	57.80	59.20	59.20	51.00
	500	$\mathcal{ZMPL}$	51.80	52.00	52.00	57.00
		$\mathcal{ZMP}$	48.20	48.00	48.00	43.00

Source: Elaborated by the author.

Table 84 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPL}$ is the true model (zero deflation).

Scenario	n	Fitted	% DIC	% EAIC	% EBIC	% LMPL
1	50	$\mathcal{ZMPL}$	49.40	49.20	49.20	56.60
		$\mathcal{ZMP}$	50.60	50.80	50.80	43.40
	100	$\mathcal{ZMPL}$	62.20	62.00	62.00	65.80
		$\mathcal{ZMP}$	37.80	38.00	38.00	34.20
	200	$\mathcal{ZMPL}$	71.60	71.40	71.40	74.00
		$\mathcal{ZMP}$	28.40	28.40	28.40	26.00
	500	$\mathcal{ZMPL}$	85.80	86.20	86.20	86.80
		$\mathcal{ZMP}$	14.20	13.80	13.80	13.20
2	50	$\mathcal{ZMPL}$	58.60	58.80	58.80	63.00
		$\mathcal{ZMP}$	41.40	41.20	41.20	37.00
	100	$\mathcal{ZMPL}$	68.00	68.00	68.00	70.60
		$\mathcal{ZMP}$	32.00	32.00	32.00	29.40
	200	$\mathcal{ZMPL}$	80.20	80.20	80.20	81.80
		$\mathcal{ZMP}$	19.80	19.80	19.80	18.20
	500	$\mathcal{ZMPL}$	92.40	92.40	92.40	92.80
		$\mathcal{ZMP}$	7.60	7.60	7.60	7.20
3	50	$\mathcal{ZMPL}$	40.80	33.80	33.80	47.40
		$\mathcal{ZMP}$	59.20	66.20	66.20	52.60
	100	$\mathcal{ZMPL}$	34.80	32.00	32.00	39.60
		$\mathcal{ZMP}$	65.20	68.00	68.00	60.40
	200	$\mathcal{ZMPL}$	45.20	44.20	44.20	49.80
		$\mathcal{ZMP}$	54.80	55.80	55.80	50.20
	500	$\mathcal{ZMPL}$	51.60	52.40	52.40	55.40
		$\mathcal{ZMP}$	48.40	47.60	47.60	44.60
4	50	$\mathcal{ZMPL}$	37.00	33.20	33.20	43.60
		$\mathcal{ZMP}$	63.00	66.80	66.80	56.40
	100	$\mathcal{ZMPL}$	40.60	40.00	40.00	45.80
		$\mathcal{ZMP}$	59.40	60.00	60.00	54.20
	200	$\mathcal{ZMPL}$	48.20	47.80	47.80	53.00
		$\mathcal{ZMP}$	51.80	52.20	52.20	47.00
	500	$\mathcal{ZMPL}$	59.60	59.60	59.60	62.00
		$\mathcal{ZMP}$	40.40	40.40	40.40	38.00

Source: Elaborated by the author.

Table 85 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_h$ is the true model (zero inflation).

Scenario	n	Fitted	% DIC	% EAIC	% EBIC	% LMPL
1	50	$\mathcal{ZMPS}_h$	67.60	67.20	67.20	75.00
		$\mathcal{ZMP}$	32.40	32.80	32.80	25.00
	100	$\mathcal{ZMPS}_h$	85.00	84.80	84.80	87.60
		$\mathcal{ZMP}$	15.00	15.20	15.20	12.40
	200	$\mathcal{ZMPS}_h$	94.60	94.60	94.60	95.60
		$\mathcal{ZMP}$	5.40	5.40	5.40	4.40
	500	$\mathcal{ZMPS}_h$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
2	50	$\mathcal{ZMPS}_h$	94.80	94.80	94.80	96.20
		$\mathcal{ZMP}$	5.20	5.20	5.20	3.80
	100	$\mathcal{ZMPS}_h$	99.20	99.20	99.20	99.40
		$\mathcal{ZMP}$	0.80	0.80	0.80	0.60
	200	$\mathcal{ZMPS}_h$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
	500	$\mathcal{ZMPS}_h$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
3	50	$\mathcal{ZMPS}_h$	73.60	71.40	71.40	83.00
		$\mathcal{ZMP}$	26.40	28.60	28.60	17.00
	100	$\mathcal{ZMPS}_h$	61.60	56.20	56.20	75.40
		$\mathcal{ZMP}$	38.40	43.80	43.80	24.60
	200	$\mathcal{ZMPS}_h$	45.80	40.20	40.20	55.40
		$\mathcal{ZMP}$	54.20	59.80	59.80	44.60
	500	$\mathcal{ZMPS}_h$	35.00	31.40	31.40	41.60
		$\mathcal{ZMP}$	65.00	68.60	68.60	58.40
4	50	$\mathcal{ZMPS}_h$	42.40	36.60	36.60	54.80
		$\mathcal{ZMP}$	57.60	63.40	63.40	45.20
	100	$\mathcal{ZMPS}_h$	34.00	31.60	31.60	44.80
		$\mathcal{ZMP}$	66.00	68.40	68.40	55.20
	200	$\mathcal{ZMPS}_h$	37.80	36.80	36.80	45.60
		$\mathcal{ZMP}$	62.20	63.20	63.20	54.40
	500	$\mathcal{ZMPS}_h$	52.20	52.40	52.40	56.60
		$\mathcal{ZMP}$	47.80	47.60	47.60	43.40

Source: Elaborated by the author.

Table 86 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_h$ is the true model (zero deflation).

Scenario	n	Fitted	% DIC	% EAIC	% EBIC	% LMPL
1	50	$\mathcal{ZMPS}_h$	45.80	45.40	45.40	52.80
		$\mathcal{ZMP}$	54.20	54.60	54.60	47.20
	100	$\mathcal{ZMPS}_h$	58.80	59.20	59.20	64.00
		$\mathcal{ZMP}$	41.20	40.80	40.80	36.00
	200	$\mathcal{ZMPS}_h$	71.00	71.00	71.00	73.40
		$\mathcal{ZMP}$	29.00	29.00	29.00	26.60
	500	$\mathcal{ZMPS}_h$	86.20	86.40	86.40	87.40
		$\mathcal{ZMP}$	13.80	13.60	13.60	12.60
2	50	$\mathcal{ZMPS}_h$	55.00	55.00	55.00	60.20
		$\mathcal{ZMP}$	45.00	45.00	45.00	39.80
	100	$\mathcal{ZMPS}_h$	66.60	66.60	66.60	70.00
		$\mathcal{ZMP}$	33.40	33.40	33.40	30.00
	200	$\mathcal{ZMPS}_h$	77.20	77.60	77.60	79.20
		$\mathcal{ZMP}$	22.80	22.40	22.40	20.80
	500	$\mathcal{ZMPS}_h$	90.20	90.00	90.00	91.00
		$\mathcal{ZMP}$	9.80	10.00	10.00	9.00
3	50	$\mathcal{ZMPS}_h$	36.40	31.40	31.40	45.60
		$\mathcal{ZMP}$	63.60	68.60	68.60	54.40
	100	$\mathcal{ZMPS}_h$	33.20	30.80	30.80	40.80
		$\mathcal{ZMP}$	66.80	69.20	69.20	59.20
	200	$\mathcal{ZMPS}_h$	40.00	39.40	39.40	44.80
		$\mathcal{ZMP}$	60.00	60.60	60.60	55.20
	500	$\mathcal{ZMPS}_h$	51.20	51.00	51.00	54.00
		$\mathcal{ZMP}$	48.80	49.00	49.00	46.00
4	50	$\mathcal{ZMPS}_h$	34.00	29.80	29.80	40.60
		$\mathcal{ZMP}$	66.00	70.20	70.20	59.40
	100	$\mathcal{ZMPS}_h$	38.00	36.60	36.60	43.60
		$\mathcal{ZMP}$	62.00	63.40	63.40	56.40
	200	$\mathcal{ZMPS}_h$	46.40	46.60	46.60	49.80
		$\mathcal{ZMP}$	53.60	53.40	53.40	50.20
	500	$\mathcal{ZMPS}_h$	56.20	57.20	57.20	60.40
		$\mathcal{ZMP}$	43.80	42.80	42.80	39.60

Source: Elaborated by the author.

Table 87 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_u$ is the true model (zero inflation).

Scenario	n	Fitted	% DIC	% EAIC	% EBIC	% LMPL
1	50	$\mathcal{ZMPS}_u$	74.80	74.80	74.80	81.40
		$\mathcal{ZMP}$	25.20	25.20	25.20	18.60
	100	$\mathcal{ZMPS}_u$	89.40	89.40	89.40	91.80
		$\mathcal{ZMP}$	10.60	10.60	10.60	8.20
	200	$\mathcal{ZMPS}_u$	98.00	98.00	98.00	98.00
		$\mathcal{ZMP}$	2.00	2.00	2.00	2.00
	500	$\mathcal{ZMPS}_u$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
2	50	$\mathcal{ZMPS}_u$	96.40	96.40	96.40	97.00
		$\mathcal{ZMP}$	3.60	3.60	3.60	3.00
	100	$\mathcal{ZMPS}_u$	99.80	99.80	99.80	99.80
		$\mathcal{ZMP}$	0.20	0.20	0.20	0.20
	200	$\mathcal{ZMPS}_u$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
	500	$\mathcal{ZMPS}_u$	100.00	100.00	100.00	100.00
		$\mathcal{ZMP}$	0.00	0.00	0.00	0.00
3	50	$\mathcal{ZMPS}_u$	64.20	64.00	64.00	75.80
		$\mathcal{ZMP}$	35.80	36.00	36.00	24.20
	100	$\mathcal{ZMPS}_u$	48.00	45.60	45.60	61.20
		$\mathcal{ZMP}$	52.00	54.40	54.40	38.80
	200	$\mathcal{ZMPS}_u$	33.20	30.80	30.80	43.40
		$\mathcal{ZMP}$	66.80	69.20	69.20	56.60
	500	$\mathcal{ZMPS}_u$	37.40	35.20	35.20	45.60
		$\mathcal{ZMP}$	62.60	64.80	64.80	54.40
4	50	$\mathcal{ZMPS}_u$	32.80	28.20	28.20	42.00
		$\mathcal{ZMP}$	67.20	71.80	71.80	58.00
	100	$\mathcal{ZMPS}_u$	33.40	32.20	32.20	41.00
		$\mathcal{ZMP}$	66.60	67.80	67.80	59.00
	200	$\mathcal{ZMPS}_u$	44.20	43.00	43.00	48.40
		$\mathcal{ZMP}$	55.80	57.00	57.00	51.60
	500	$\mathcal{ZMPS}_u$	53.60	53.60	53.60	55.80
		$\mathcal{ZMP}$	46.40	46.40	46.40	44.20

Source: Elaborated by the author.

Table 88 – *Posterior* estimates of correct selection probability when $\mathcal{ZMPS}_u$ is the true model (zero deflation).

Scenario	n	Fitted	% DIC	% EAIC	% EBIC	% LMPL
1	50	$\mathcal{ZMPS}_u$	56.00	56.00	56.00	62.40
		$\mathcal{ZMP}$	44.00	44.00	44.00	37.60
	100	$\mathcal{ZMPS}_u$	68.60	68.40	68.40	71.40
		$\mathcal{ZMP}$	31.40	31.60	31.60	28.60
	200	$\mathcal{ZMPS}_u$	77.80	77.80	77.80	79.80
		$\mathcal{ZMP}$	22.20	22.20	22.20	20.20
	500	$\mathcal{ZMPS}_u$	92.60	92.60	92.60	93.20
		$\mathcal{ZMP}$	7.40	7.40	7.40	6.80
2	50	$\mathcal{ZMPS}_u$	63.20	63.60	63.60	67.00
		$\mathcal{ZMP}$	36.80	36.40	36.40	33.00
	100	$\mathcal{ZMPS}_u$	72.60	72.80	72.80	75.60
		$\mathcal{ZMP}$	27.40	27.20	27.20	24.40
	200	$\mathcal{ZMPS}_u$	83.40	83.60	83.60	85.40
		$\mathcal{ZMP}$	16.60	16.40	16.40	14.60
	500	$\mathcal{ZMPS}_u$	96.80	96.80	96.80	96.80
		$\mathcal{ZMP}$	3.20	3.20	3.20	3.20
3	50	$\mathcal{ZMPS}_u$	30.20	25.80	25.80	38.40
		$\mathcal{ZMP}$	69.80	74.20	74.20	61.60
	100	$\mathcal{ZMPS}_u$	36.20	35.20	35.20	40.60
		$\mathcal{ZMP}$	63.80	64.80	64.80	59.40
	200	$\mathcal{ZMPS}_u$	43.80	44.00	44.00	49.40
		$\mathcal{ZMP}$	56.20	56.00	56.00	50.60
	500	$\mathcal{ZMPS}_u$	53.20	53.40	53.40	55.60
		$\mathcal{ZMP}$	46.80	46.60	46.60	44.40
4	50	$\mathcal{ZMPS}_u$	35.40	33.00	33.00	42.00
		$\mathcal{ZMP}$	64.60	67.00	67.00	58.00
	100	$\mathcal{ZMPS}_u$	44.20	43.80	43.80	50.80
		$\mathcal{ZMP}$	55.80	56.20	56.20	49.20
	200	$\mathcal{ZMPS}_u$	51.40	51.40	51.40	54.40
		$\mathcal{ZMP}$	48.60	48.60	48.60	45.60
	500	$\mathcal{ZMPS}_u$	61.20	61.60	61.60	64.40
		$\mathcal{ZMP}$	38.80	38.40	38.40	35.60

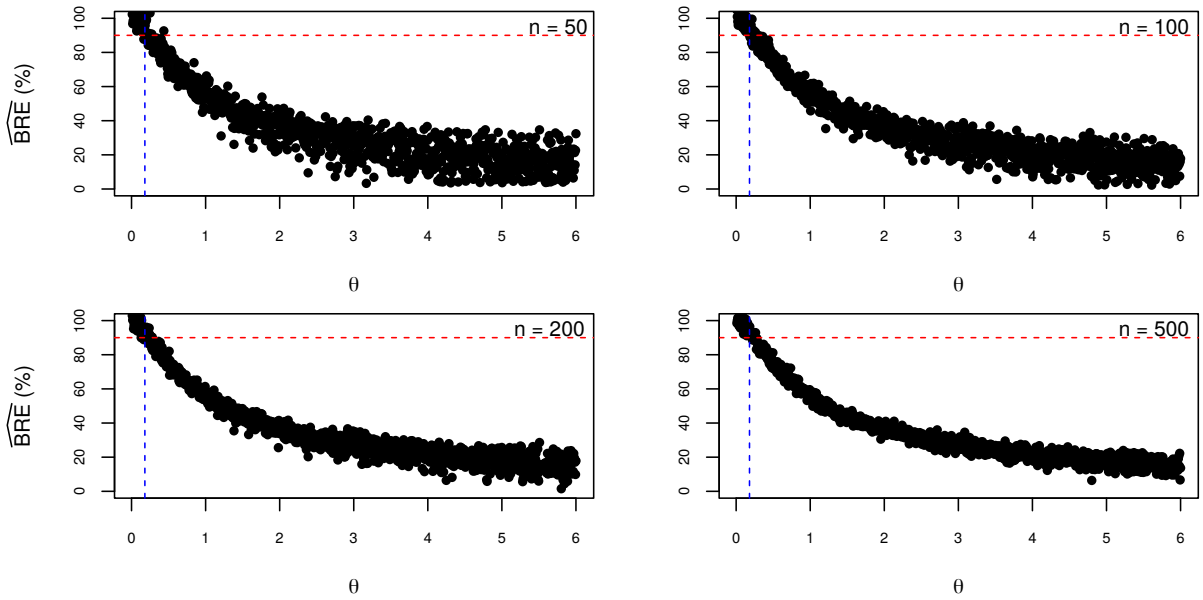
Source: Elaborated by the author.

A.3.3 Study 3

The following figures illustrate the results obtained in the third simulation study, described in Subsection 2.6.3. In each figure, the horizontal dashed lines (red) were placed at 90% BRE.

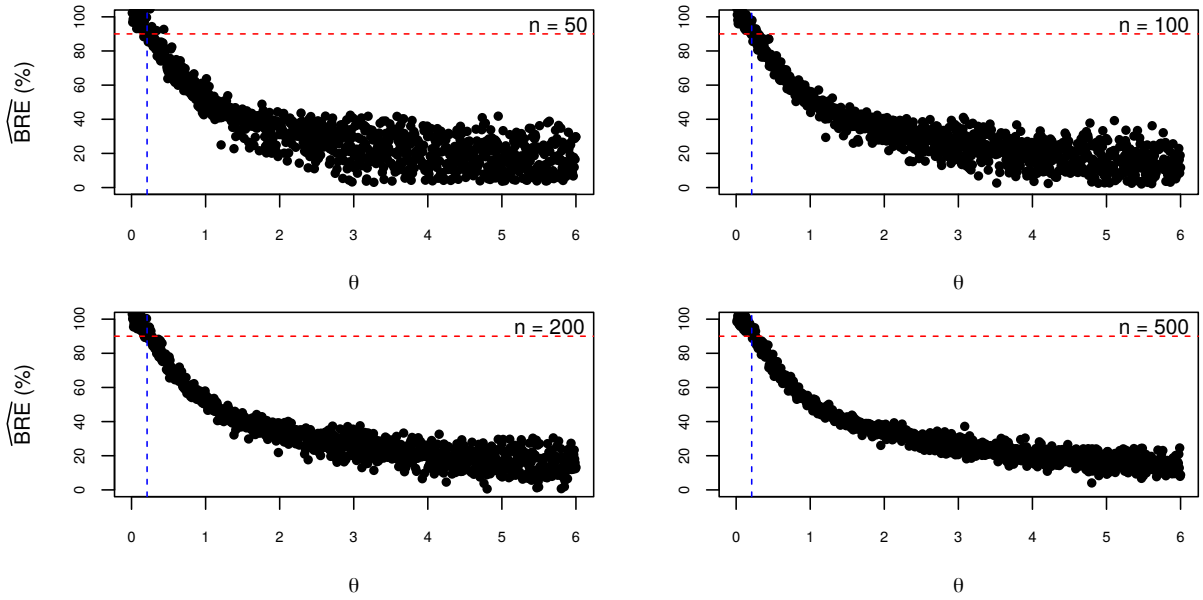
The vertical dashed lines (blue) were placed at $\theta = 0.18$ in Figure 34, at $\theta = 0.21$ in Figure 35 and at $\theta = 0.31$ in Figure 36.

Figure 34 – BRE of estimated $\mathcal{ZMPL}$ distribution when $\mathcal{PL}$ is the true model.



Source: Elaborated by the author.

Figure 35 – BRE of estimated $\mathcal{ZMPS}_h$ distribution when $\mathcal{PS}_h$ is the true model.



Source: Elaborated by the author.

A.4 Datasets

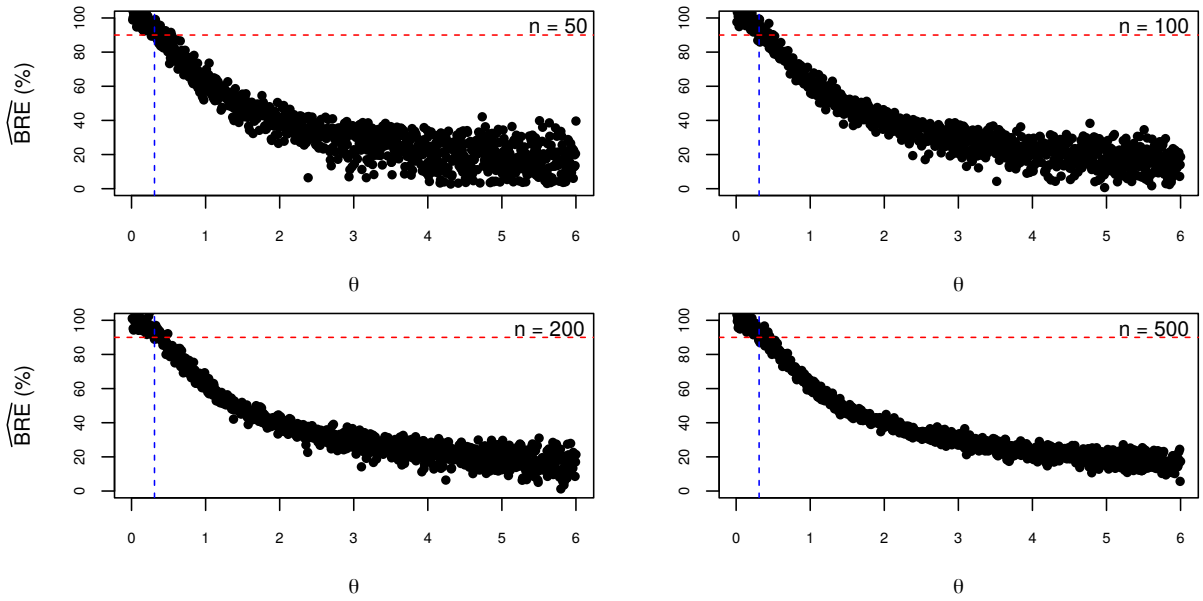
The three real datasets used in the paper to illustrate the usefulness of the proposed class of models are provided in Table 89. A brief description of each dataset is presented in Section 2.7 of the manuscript.

Table 89 – Real datasets used in the paper.

Dataset 1																			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3
4	4	4	4	5	5	5	6	6	6	6	6	6	7	8	17				
Dataset 2																			
0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4
4	5	6	6	7	10														
Dataset 3																			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	5

Source: Elaborated by the author.

Figure 36 – BRE of estimated $\mathcal{ZMPS}_u$ distribution when $\mathcal{PS}_u$ is the true model.



Source: Elaborated by the author.

APPENDIX

B

SUPPLEMENT FOR CHAPTER 4

B.1 Influential points

The identification of influential observations is one of the essential steps in any statistical analysis. Usually, the presence of influential points impacts the inferential procedures and the subsequent conclusions considerably. In this way, this subsection is dedicated to presenting some case deletion Bayesian diagnostic measures that can be useful to quantify the influence of each observation in a given dataset.

The computation of divergence measures between *posterior* distributions is a very useful way to quantify influence. According Csiszár (1967), the φ-divergence measure between two densities f and g for $\boldsymbol{\theta} \in \mathcal{D}$ is defined by

$$d_{\varphi} = \int_{\mathcal{D}} g(\boldsymbol{\theta})\, \varphi\left[\frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\right] \mathrm{d}\boldsymbol{\theta}, \tag{B.1}$$

where φ is a smooth convex, lower semicontinuous function such that $\varphi(1) = 0$. Some popular divergence measures can be obtained by choosing specific functions for φ (PENG; DEY, 1995). We are interested in the well-known Kullback-Leibler divergence that can be obtained by setting $\varphi(t) = -\log(t)$.

Suppose that we are studying a discrete random variable Y whose distribution is indexed by a parametric vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Aiming to estimate such vector, we have observed n independent values of Y, hence obtaining the full observed vector $\boldsymbol{y}$. Now, let $\boldsymbol{y}_{(-i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ be a vector obtained after removal of the i-th observation from $\boldsymbol{y}$. Given a *prior* distribution $\pi(\boldsymbol{\theta})$, the full *posterior* density of $\boldsymbol{\theta}$ can be expressed as

$$\pi(\boldsymbol{\theta}; \boldsymbol{y}) = \frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})\, \pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})\, \pi(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}},$$

where $\mathcal{L}$ stands for the likelihood function of $\boldsymbol{\theta}$. Conversely, using the vector without the i-th

observation, the *posterior* distribution of $\boldsymbol{\theta}$ can be written as

$$\pi\left(\boldsymbol{\theta};\boldsymbol{y}_{(-i)}\right)=\frac{\mathcal{L}\left(\boldsymbol{\theta};\boldsymbol{y}_{(-i)}\right)\pi\left(\boldsymbol{\theta}\right)}{\int_{\boldsymbol{\Theta}}\mathcal{L}\left(\boldsymbol{\theta};\boldsymbol{y}_{(-i)}\right)\pi\left(\boldsymbol{\theta}\right)\mathrm{d}\boldsymbol{\theta}}.$$

Now, taking $f(\boldsymbol{\theta})=\pi(\boldsymbol{\theta};\boldsymbol{y}_{(-i)})$ and $g(\boldsymbol{\theta})=\pi(\boldsymbol{\theta};\boldsymbol{y})$, Equation (B.1) becomes

$$d_{\varphi}=\mathbb{E}_{\boldsymbol{\theta}}\left\{\varphi\left[\frac{\mathbb{E}_{\boldsymbol{\theta}}^{-1}\left[\mathrm{P}^{-1}(Y_i=y_i;\boldsymbol{\theta});\boldsymbol{y}\right]}{\mathrm{P}(Y_i=y_i;\boldsymbol{\theta})}\right];\boldsymbol{y}\right\}, \tag{B.2}$$

where $\mathbb{E}_{\boldsymbol{\theta}}^{-1}[\mathrm{P}^{-1}(Y_i=y_i;\boldsymbol{\theta});\boldsymbol{y}]$ is the conditional predictive ordinate (CPO) statistic (GEISSER, 1993) for the i-th observation. Given a sample $\{\boldsymbol{\theta}^1,\ldots,\boldsymbol{\theta}^M\}$ from the *posterior* distribution $\pi(\boldsymbol{\theta};\boldsymbol{y})$, a Monte Carlo (MC) estimator for the CPO_i is given by

$$\widehat{\mathrm{CPO}}_i=\left[\frac{1}{M}\sum_{k=1}^{M}\mathrm{P}^{-1}\left(Y_i=y_i;\boldsymbol{\theta}^{(k)}\right)\right]^{-1}, \tag{B.3}$$

and hence, one can estimate the local influence of a particular observation y_i on the *posterior* distribution of $\boldsymbol{\theta}$ as

$$\hat{d}_{\varphi}=\frac{1}{M}\sum_{k=1}^{M}\varphi\left[\frac{\widehat{\mathrm{CPO}}_i}{\mathrm{P}\left(Y_i=y_i;\boldsymbol{\theta}^{(k)}\right)}\right].$$

From Equation (B.2), one can notice that, if $\pi(\boldsymbol{\theta};\boldsymbol{y}_{(-i)})=\pi(\boldsymbol{\theta};\boldsymbol{y})$, then there is no divergence caused by the observation y_i. In practice, however, it may be quite difficult to define a threshold value for the divergence measure in order to decide about the magnitude of the influence. A measure of calibration for the Kullback-Leibler divergence was proposed by McCulloch (1989). The idea is based on the typical toy binary experiment of tossing a coin once and observing its upper face is observed. This experiment can be described by $\mathrm{P}(Y=y;\rho)=\rho^{y}(1-\rho)^{1-y}$, where $\rho\in[0,1]$ is the probability of success. Regardless what success means, if the coin is unbiased, then $\mathrm{P}(Y=y;\rho)=0.50$, $y\in\{0,1\}$. Thus, the φ-divergence between a (possibly) biased and an unbiased coin is given by

$$d_{\varphi}(\rho)=\frac{\varphi\left(2\rho\right)+\varphi\left[2\left(1-\rho\right)\right]}{2},$$

from which one may conclude that the divergence between two *posteriors* distributions can be associated with the biasedness of a coin (PENG; DEY, 1995). By analogy, this implies that predict unobserved responses through $\pi(\boldsymbol{\theta};\boldsymbol{y}_{(-i)})$ instead of $\pi(\boldsymbol{\theta};\boldsymbol{y})$ is equivalent to describe a not observed event as having probability ρ_i, when the correct probability is 0.50. For the Kullback-Leibler divergence, we have

$$d_{\varphi}(\rho_i)=-\frac{1}{2}\log\left[4\rho_i\left(1-\rho_i\right)\right].$$

The function $d_{\varphi}(\rho)$ is symmetric about $\rho=0.50$ and increases as ρ moves away from 0.50. Also, $\inf_{\rho\in(0,1)}d_{\varphi}(\rho)=0$, which is attained at $\rho=0.50$ since $d_{\varphi}(0.50)=\varphi(1)=0$.

Therefore, a general measure of calibration based on a φ-divergence can be obtained by solving equation $2d_{\varphi}(\rho) - \varphi(2\rho) - \varphi[2(1-\rho)] = 0$. An MC estimator for the calibration measure (ρ) associated with the Kullback-Leibler divergence is given by

$$\hat{\rho}_i = \frac{1}{2}\left[1 + \sqrt{1 - e^{-2\hat{d}_i}}\right],$$

where the local influence of each y_i can be estimated by

$$\hat{d}_i = \frac{1}{M}\sum_{k=1}^{M} \log\left[\mathrm{P}\Big(Y_i = y_i; \boldsymbol{\theta}^{(k)}\Big)\right] - \log\left(\widehat{\mathrm{CPO}}_i\right).$$

One can notice that $\rho_i \in [0.50, 1]$ and therefore, for $\rho_i \gg 0.50$, the i-th observation can be considered as an influential point. For example, if $\rho_i \geqslant 0.80$ is considered a significative bias, then the i-th observation will be classified as being influential if $\hat{d}_i > 0.223$ $\big(d_{\varphi}(0.80) \approx 0.223\big)$.

B.2 Model comparison

There are several methods for Bayesian model selection that are useful to compare competing models fitted to the same dataset. One of the most used criteria is the deviance information criterion (DIC), which was proposed to work simultaneously as a measure of fit and complexity of the model. To define such a measure, suppose again that we are studying a discrete random variable Y whose distribution is indexed by a parametric vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and let $\boldsymbol{y}$ as a vector of n independent observations obtained from Y. The DIC criterion is given by

$$\mathrm{DIC} = \bar{\mathrm{D}}(\boldsymbol{\theta}) + \rho_{\mathrm{D}} = 2\bar{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\tilde{\boldsymbol{\theta}}),$$

where $\bar{\mathrm{D}}(\boldsymbol{\theta}) = -2\mathbb{E}[\ell(\boldsymbol{\theta}; \boldsymbol{y})]$ is the *posterior* expectation of the *deviance* and ℓ is the log-likelihood function of $\boldsymbol{\theta}$. In this case, the *deviance* is evaluated at some estimate $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ (e.g., the *posterior* conditional mean). In connection with a measure of model complexity, the criterion considers the measure $\rho_{\mathrm{D}} = \bar{\mathrm{D}}(\boldsymbol{\theta}) - \mathrm{D}(\tilde{\boldsymbol{\theta}})$, which correspond to the effective number of parameters in the model. One can notice that the computation of $\bar{\mathrm{D}}(\boldsymbol{\theta})$ is a complex numerical problem. In this case, a Monte Carlo estimator for such a measure is given by

$$\bar{\mathrm{D}} = -\frac{2}{M}\sum_{k=1}^{M} \ell\left(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}\right),$$

and hence, the DIC can be approximated by

$$\widehat{\mathrm{DIC}} = 2\bar{\mathrm{D}} - \mathrm{D}(\tilde{\boldsymbol{\theta}}).$$

Using the estimate $\bar{\mathrm{D}}$, one can define other measures that can be considered when comparing models. The expected Akaike information criterion (EAIC) and the expected Bayesian information criterion (EBIC) can be estimated as

$$\widehat{\mathrm{EAIC}} = \bar{\mathrm{D}} + 2q \quad \text{and} \quad \widehat{\mathrm{EBIC}} = \bar{\mathrm{D}} + q\log(n),$$

where q is the total number of estimated model parameters. See Carlin and Louis (2001) and Brooks (2002) for further details on these comparison criteria.

Another widely used criterion is derived from the CPO measure, which is based on the cross-validation criterion to compare models. For the i-th individual, the CPO can be estimated using Equation (B.3). A summary statistic of the estimated CPO's is the log-marginal pseudo-likelihood (LMPL) given by the sum of the logarithms of $\widehat{\text{CPO}}_i$'s. Regarding model comparison, we have that the lower the value of DIC, EAIC, and EBIC, the better the fit. On the other hand, for the latter criterion, we have that the larger the LMPL, the better the fit. One can notice that, for all the presented criteria, the computation of the likelihood function is a crucial step to estimate such measures.

B.3 Residual analysis

The residual analysis plays an essential role in the task of validating the results obtained from a regression model. In general, residual metrics are responsible for indicating departures from the underlying model assumptions by quantifying the portion of data variability that was not explained by the fitted model. Assessing the adequacy of a regression model using residual metrics is a common practice nowadays due to the availability of statistical packages providing diagnostic tools for well-established models. However, deriving appropriate residuals is not always an easy task for non-normal models that accommodate overdispersion and mixed-effects. In this way, we will consider here a popular residual metric proposed by Dunn and Smyth (1996), the RQRs, which can be straightforwardly used in our context to assess the appropriateness of the proposed model when fitted to real data.

For obvious reasons, we focus on the definition of RQRs for discrete random variables. In this case, the RQR associated to the i-th observation is defined as $r_i = \Phi^{-1}(u_i)$, where Φ denotes the cdf of the standard Normal distribution and u_i is a Uniform random variable defined on $(a_i, b_i]$, with $a_i = \lim_{y \uparrow y_i} \mathrm{F}(y_i)$ and $b_i = \mathrm{F}(y_i)$, where $\mathrm{F}(y_i)$ is the cdf of the current model. In our case, we may obtain an MC estimator for the RQR as $\hat{r}_i = \Phi^{-1}(u_i)$, with $u_i \sim \mathcal{U}(\lim_{y \uparrow y_i} \hat{\mathrm{F}}_*(y_i), \hat{\mathrm{F}}_*(y_i)]$. Here, $\hat{\mathrm{F}}_*(y_i) \equiv \mathrm{F}_*(y_i; \hat{\mu}_i, \hat{\omega}_i)$ is an estimate for the probability of $Y_i \leqslant y_i$ using cdf (4.11), where $\hat{\mu}_i$ and $\hat{\omega}_i$ depend on the fitted model as $\hat{\mu}_i = \log(\boldsymbol{x}_{1i}^\top \hat{\boldsymbol{\beta}}_1)$ and $\hat{\omega}_i = g_2^{-1}(\boldsymbol{x}_{2i}^\top \hat{\boldsymbol{\beta}}_2)$.

The main assumption for this metric is that $\hat{r}_i \sim \mathcal{N}(0,1)$ must hold, whichever the variability degree of $\hat{\mu}_i$ and $\hat{\omega}_i$. In this case, after model fitting, one has to evaluate if these residuals are normally distributed around zero, which can be made through adherence tests and by using graphical techniques as histograms and Half-Normal probability plots. An excellent alternative for checking whether RQRs are consistent with the fitted model is the inclusion of simulated envelopes in their Half-Normal plot. Thus, if a significative subset of estimated residuals falls outside the envelope bands, then the adequacy of the fitted model must be questioned, and further investigation on the corresponding observations are necessary. An

algorithm for obtaining simulated envelopes for a Half-Normal plot is provided by Moral, Hinde and Demétrio (2017).

APPENDIX

C

SUPPLEMENT FOR CHAPTER 5

C.1 Algorithms

C.1.1 Random-walk Metropolis

Algorithm 7 – Random-walk Metropolis

procedure RWM($N, \boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}$)
  Set $t \leftarrow 1$ and $\nu \leftarrow n(n+1)^{-1}$
  while $t \leqslant N$ **do**
    Generate $\boldsymbol{\psi}_1 \sim \mathcal{N}_{\bar{q}_1}[\nu\boldsymbol{\beta}_1^{(t-1)}, \nu\boldsymbol{\mathcal{S}}_1^{(t-1)}]$ and $\boldsymbol{\psi}_2 \sim \mathcal{N}_{\bar{q}_2}[\nu\boldsymbol{\beta}_2^{(t-1)}, \nu\boldsymbol{\mathcal{S}}_2^{(t-1)}]$
    Set $\alpha_1 \leftarrow \exp\{\pi_1(\boldsymbol{\psi}_1;\boldsymbol{y}) - \pi_1(\boldsymbol{\beta}_1^{(t-1)};\boldsymbol{y})\}$
    Set $\alpha_2 \leftarrow \exp\{\pi_2(\boldsymbol{\psi}_2;\boldsymbol{y}) - \pi_2(\boldsymbol{\beta}_2^{(t-1)};\boldsymbol{y})\}$
    Set $\boldsymbol{\beta}_1^{(t)} \leftarrow \boldsymbol{\beta}_1^{(t-1)}$ and $\boldsymbol{\beta}_2^{(t)} \leftarrow \boldsymbol{\beta}_2^{(t-1)}$
    Generate $u_1, u_2 \sim \mathcal{U}(0,1)$
    if $u_1 \leqslant \min\{1, \alpha_1\}$ and $u_2 \leqslant \min\{1, \alpha_2\}$ **then**
      Set $\boldsymbol{\beta}_1^{(t)} \leftarrow \boldsymbol{\psi}_1$ and $\boldsymbol{\beta}_2^{(t)} \leftarrow \boldsymbol{\psi}_2$
    end if
    Set $t \leftarrow t+1$
  end while
  return $\{\boldsymbol{\beta}^t\}_{t=1}^N$
end procedure

C.1.2 Sequential-Search

Algorithm 8 – Sequential-Search

procedure SEQSEA($\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$)
  Generate $x, u \sim \mathcal{U}(0,1)$
  Set $\mu \leftarrow \exp\{\beta_{10} + \beta_{11}x\}$ and $\omega \leftarrow g_2^{-1}(\beta_{20} + \beta_{21}x)$
  Set $k \leftarrow (1-\omega)$ and $y \leftarrow 0$
  while $u > k$ **do**
    Set $y \leftarrow y+1$ and $k \leftarrow k + \omega\,\mathrm{P}_*(Y = y;\mu)$
  end while
  return y
end procedure

C.2 Divergence measures

The acronymous used in Table 90 refers to following divergence measures: Kullback-Leibler (KL), Jeffrey (J), Variational (L^1), Chi-Square (CS), and Hellinger (H).

Table 90 – MC estimators for some φ-divergence measures and its calibration.

φ	$\hat{d}_\varphi$	$d_\varphi(\rho)$	$\hat{\rho}_\varphi$
KL_i	$\frac{1}{M}\sum_{t=1}^{M}\log\left(f_i^t\right)-\log\left(\widehat{CPO}_i\right)$	$-\frac{1}{2}\log\left[4\rho_i\left(1-\rho_i\right)\right]$	$\frac{1}{2}\left[1+\sqrt{1-e^{-2\hat{d}_i}}\right]$
J_i	$\frac{1}{M}\sum_{t=1}^{M}\left(\frac{\widehat{CPO}_i}{f_i^t}-1\right)\log\left(\frac{\widehat{CPO}_i}{f_i^t}\right)$	$-\frac{(1-2\rho_i)}{2}\log\left(\frac{\rho_i}{1-\rho_i}\right)$	no closed-form
L_i^1	$\frac{1}{2M}\sum_{t=1}^{M}\frac{\left\lvert\widehat{CPO}_i-f_i^t\right\rvert}{f_i^t}$	$\frac{1}{2}\left\lvert 1-2\rho_i\right\rvert$	$\frac{1}{2}+\hat{d}_i$
CS_i	$\frac{1}{M}\sum_{t=1}^{M}\frac{\left(\widehat{CPO}_i-f_i^t\right)^2}{(f_i^t)^2}$	$(1-2\rho_i)^2$	$\frac{1}{2}\left[1+\sqrt{\hat{d}_i}\right]$
H_i	$\frac{1}{2M}\sum_{t=1}^{M}\left(\sqrt{\frac{\widehat{CPO}_i}{f_i^t}}-1\right)^2$	$1-\frac{1}{\sqrt{2}}\left(\sqrt{\rho_i}+\sqrt{1-\rho_i}\right)$	$\frac{1}{2}+\left[\sqrt{\hat{d}_i}-\sqrt{\hat{d}_i^3}\right]\sqrt{2-\hat{d}_i}$

Source: Elaborated by the author.

C.3 Results from the simulation study

C.3.1 Zero-inflated artificial data

C.3.1.1 Using logit link function

Table 91 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 1		
50	β_{10}	−0.116	0.336	1.020	29.127
	β_{11}	−0.527	1.784	1.088	30.652
	β_{20}	0.119	0.329	1.022	45.714
	β_{21}	0.116	0.990	1.007	78.781
100	β_{10}	−0.097	0.161	1.031	20.605
	β_{11}	−0.138	0.480	1.020	18.120
	β_{20}	0.084	0.247	1.014	40.288
	β_{21}	0.053	0.774	1.002	70.962
200	β_{10}	−0.033	0.058	1.010	12.471
	β_{11}	−0.098	0.193	1.026	11.590
	β_{20}	0.012	0.110	1.001	26.005
	β_{21}	0.061	0.412	1.004	50.919
500	β_{10}	−0.032	0.022	1.024	7.870
	β_{11}	−0.011	0.064	1.001	6.794
	β_{20}	0.009	0.049	1.001	17.633
	β_{21}	0.023	0.169	1.002	33.136
			Scenario 2		
50	β_{10}	−0.164	0.180	1.085	21.689
	β_{11}	−0.136	0.382	1.025	16.319
	β_{20}	0.128	0.328	1.026	45.883
	β_{21}	−0.081	0.896	1.004	151.990
100	β_{10}	−0.078	0.094	1.034	15.908
	β_{11}	−0.075	0.204	1.014	11.916
	β_{20}	0.049	0.217	1.006	36.781
	β_{21}	−0.016	0.574	1.000	117.390
200	β_{10}	−0.051	0.044	1.031	11.079
	β_{11}	−0.010	0.104	1.000	8.443
	β_{20}	0.010	0.101	1.000	25.223
	β_{21}	0.013	0.322	1.000	90.972
500	β_{10}	−0.024	0.015	1.019	6.533
	β_{11}	−0.005	0.033	1.000	4.778
	β_{20}	0.005	0.039	1.000	15.802
	β_{21}	0.002	0.104	1.000	52.292

Source: Elaborated by the author.

Table 92 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 3					
50	β_{10}	−0.338	0.584	1.115	37.493
	β_{11}	0.281	1.758	1.023	69.019
	β_{20}	0.109	0.388	1.016	49.700
	β_{21}	0.186	1.102	1.016	84.008
100	β_{10}	−0.120	0.241	1.032	24.811
	β_{11}	−0.036	1.045	1.001	52.507
	β_{20}	0.056	0.222	1.007	38.480
	β_{21}	0.104	0.681	1.008	67.186
200	β_{10}	−0.071	0.105	1.025	16.547
	β_{11}	−0.005	0.526	1.000	38.617
	β_{20}	0.006	0.106	1.000	25.932
	β_{21}	0.066	0.405	1.005	50.259
500	β_{10}	−0.024	0.039	1.008	10.202
	β_{11}	−0.002	0.171	1.000	21.413
	β_{20}	0.005	0.051	1.000	17.858
	β_{21}	0.043	0.166	1.006	32.488
Scenario 4					
50	β_{10}	−0.221	0.386	1.070	31.170
	β_{11}	0.138	1.024	1.009	54.128
	β_{20}	0.086	0.380	1.010	48.477
	β_{21}	−0.063	1.108	1.002	165.519
100	β_{10}	−0.114	0.192	1.035	23.082
	β_{11}	0.062	0.584	1.003	41.147
	β_{20}	0.068	0.218	1.011	37.185
	β_{21}	−0.076	0.700	1.004	133.345
200	β_{10}	−0.052	0.067	1.021	13.693
	β_{11}	0.011	0.236	1.000	26.593
	β_{20}	0.040	0.080	1.010	22.414
	β_{21}	−0.033	0.269	1.002	81.162
500	β_{10}	−0.025	0.028	1.012	8.884
	β_{11}	0.022	0.093	1.002	16.306
	β_{20}	−0.001	0.042	1.000	16.202
	β_{21}	0.010	0.118	1.000	54.116

Source: Elaborated by the author.

Table 93 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
50	β_{10}	1.384	1.386	0.568	0.416	2.348
	β_{11}	2.473	2.461	1.228	0.736	4.230
	β_{20}	−0.881	−0.865	0.561	−2.184	0.395
	β_{21}	−0.884	−0.873	0.988	−3.221	1.432
100	β_{10}	1.403	1.403	0.390	0.701	2.105
	β_{11}	2.862	2.855	0.679	1.631	4.105
	β_{20}	−0.916	−0.907	0.490	−1.932	0.080
	β_{21}	−0.947	−0.942	0.878	−2.784	0.878
200	β_{10}	1.467	1.467	0.238	1.031	1.903
	β_{11}	2.902	2.897	0.428	2.089	3.721
	β_{20}	−0.988	−0.983	0.332	−1.635	−0.346
	β_{21}	−0.939	−0.933	0.639	−2.219	0.332
500	β_{10}	1.468	1.468	0.147	1.188	1.748
	β_{11}	2.989	2.987	0.253	2.505	3.476
	β_{20}	−0.991	−0.988	0.220	−1.414	−0.571
	β_{21}	−0.977	−0.975	0.410	−1.758	−0.200
Scenario 2						
50	β_{10}	1.336	1.336	0.391	0.588	2.087
	β_{11}	2.864	2.858	0.603	1.725	4.011
	β_{20}	−0.872	−0.860	0.558	−2.071	0.308
	β_{21}	0.419	0.416	0.943	−1.571	2.415
100	β_{10}	1.422	1.422	0.296	0.849	1.997
	β_{11}	2.925	2.922	0.445	2.023	3.828
	β_{20}	−0.951	−0.943	0.464	−1.880	−0.032
	β_{21}	0.484	0.479	0.758	−1.071	2.052
200	β_{10}	1.449	1.449	0.204	1.082	1.814
	β_{11}	2.990	2.989	0.322	2.395	3.590
	β_{20}	−0.990	−0.986	0.318	−1.582	−0.404
	β_{21}	0.513	0.513	0.567	−0.546	1.569
500	β_{10}	1.476	1.476	0.121	1.239	1.714
	β_{11}	2.995	2.994	0.181	2.631	3.359
	β_{20}	−0.995	−0.993	0.198	−1.380	−0.611
	β_{21}	0.502	0.502	0.322	−0.139	1.147

Source: Elaborated by the author.

Table 94 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
50	β_{10}	1.162	1.175	0.685	−0.184	2.482
	β_{11}	−1.219	−1.208	1.296	−3.982	1.521
	β_{20}	−0.891	−0.875	0.613	−2.200	0.382
	β_{21}	−0.814	−0.802	1.033	−3.143	1.484
100	β_{10}	1.380	1.386	0.476	0.414	2.335
	β_{11}	−1.536	−1.529	1.021	−3.569	0.479
	β_{20}	−0.944	−0.933	0.468	−1.960	0.057
	β_{21}	−0.896	−0.891	0.819	−2.729	0.928
200	β_{10}	1.429	1.432	0.317	0.846	2.009
	β_{11}	−1.505	−1.499	0.725	−2.882	−0.138
	β_{20}	−0.994	−0.989	0.325	−1.644	−0.353
	β_{21}	−0.934	−0.929	0.633	−2.218	0.335
500	β_{10}	1.476	1.477	0.196	1.108	1.843
	β_{11}	−1.502	−1.500	0.413	−2.317	−0.690
	β_{20}	−0.995	−0.993	0.225	−1.418	−0.575
	β_{21}	−0.957	−0.956	0.406	−1.736	−0.182
Scenario 4						
50	β_{10}	1.279	1.287	0.581	0.141	2.402
	β_{11}	−1.363	−1.356	1.003	−3.430	0.697
	β_{20}	−0.914	−0.901	0.610	−2.122	0.275
	β_{21}	0.437	0.434	1.051	−1.570	2.445
100	β_{10}	1.386	1.391	0.424	0.567	2.198
	β_{11}	−1.438	−1.437	0.762	−2.947	0.066
	β_{20}	−0.932	−0.925	0.462	−1.860	−0.016
	β_{21}	0.424	0.421	0.834	−1.137	1.991
200	β_{10}	1.448	1.450	0.253	0.945	1.949
	β_{11}	−1.489	−1.487	0.486	−2.502	−0.477
	β_{20}	−0.960	−0.957	0.280	−1.549	−0.376
	β_{21}	0.467	0.466	0.518	−0.587	1.522
500	β_{10}	1.475	1.476	0.166	1.150	1.799
	β_{11}	−1.478	−1.478	0.304	−2.083	−0.873
	β_{20}	−1.001	−1.000	0.204	−1.388	−0.616
	β_{21}	0.510	0.509	0.343	−0.133	1.155

Source: Elaborated by the author.

Table 95 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	3.00	92.80	4.20	1.00	93.80	5.20
	β_{11}	1.00	88.00	11.00	1.80	93.40	4.80
	β_{20}	1.80	98.20	0.00	3.60	95.80	0.60
	β_{21}	1.60	98.00	0.40	1.00	97.60	1.40
100	β_{10}	2.20	93.60	4.20	1.60	94.00	4.40
	β_{11}	1.80	92.60	5.60	1.20	95.20	3.60
	β_{20}	1.80	97.20	1.00	4.00	94.40	1.60
	β_{21}	2.00	97.20	0.80	2.00	95.00	3.00
200	β_{10}	1.60	93.00	5.40	2.40	92.60	5.00
	β_{11}	1.60	93.20	5.20	2.20	94.00	3.80
	β_{20}	3.20	95.00	1.80	3.80	94.00	2.20
	β_{21}	3.00	95.60	1.40	3.20	93.80	3.00
500	β_{10}	1.20	94.00	4.80	1.00	95.40	3.60
	β_{11}	2.60	94.60	2.80	1.40	96.00	2.60
	β_{20}	2.40	95.20	2.40	3.20	94.80	2.00
	β_{21}	2.60	95.20	2.20	2.40	95.40	2.20

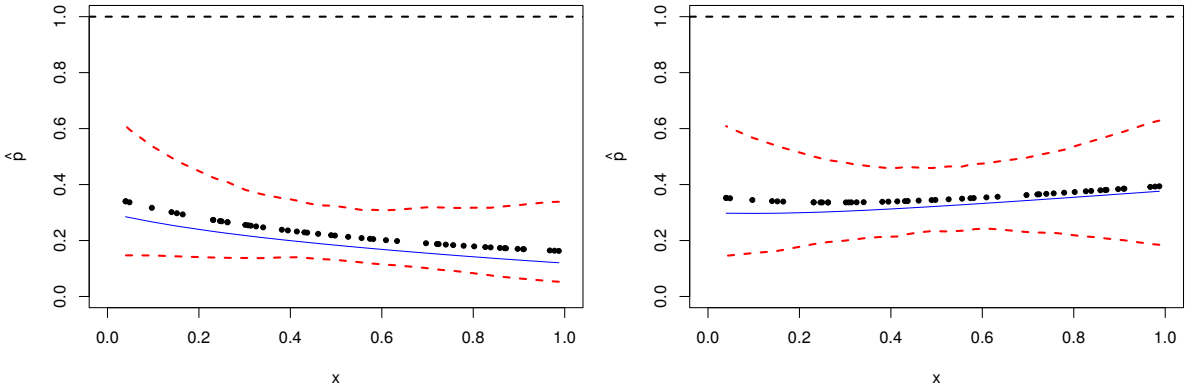
Source: Elaborated by the author.

Table 96 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	β_{10}	0.40	95.40	4.20	0.20	96.80	3.00
	β_{11}	3.20	96.40	0.40	2.40	97.00	0.60
	β_{20}	2.20	97.40	0.40	4.60	94.40	1.00
	β_{21}	1.80	97.60	0.60	1.80	94.80	3.40
100	β_{10}	1.40	95.80	2.80	1.20	95.00	3.80
	β_{11}	2.20	96.20	1.60	2.60	95.00	2.40
	β_{20}	2.40	97.40	0.20	4.20	94.60	1.20
	β_{21}	1.80	97.60	0.60	1.80	93.60	4.60
200	β_{10}	2.00	94.40	3.60	1.20	96.60	2.20
	β_{11}	3.20	95.00	1.80	1.80	97.00	1.20
	β_{20}	2.20	96.20	1.60	3.20	95.80	1.00
	β_{21}	2.80	95.60	1.60	2.40	94.40	3.20
500	β_{10}	2.20	94.40	3.40	2.20	95.80	2.00
	β_{11}	2.40	94.80	2.80	2.40	95.40	2.20
	β_{20}	3.80	94.40	1.80	2.20	95.00	2.80
	β_{21}	2.80	95.20	2.00	3.40	94.60	2.00

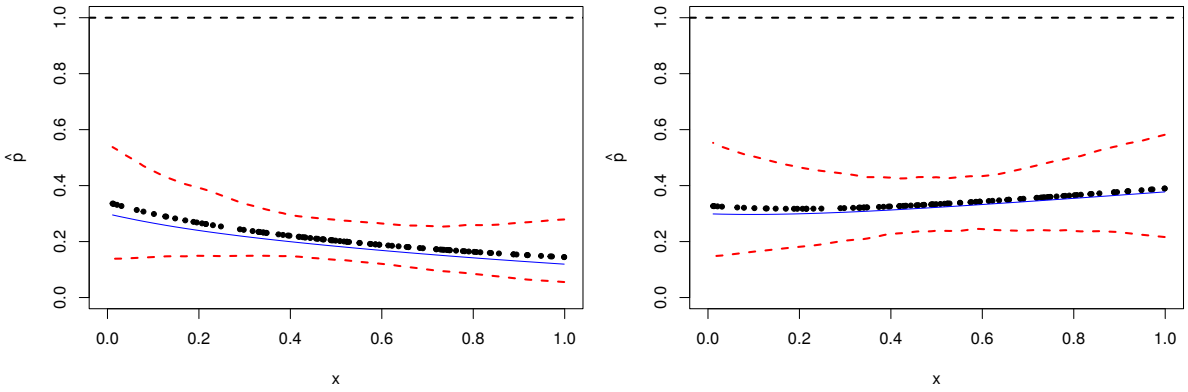
Source: Elaborated by the author.

Figure 37 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 1 and 2).



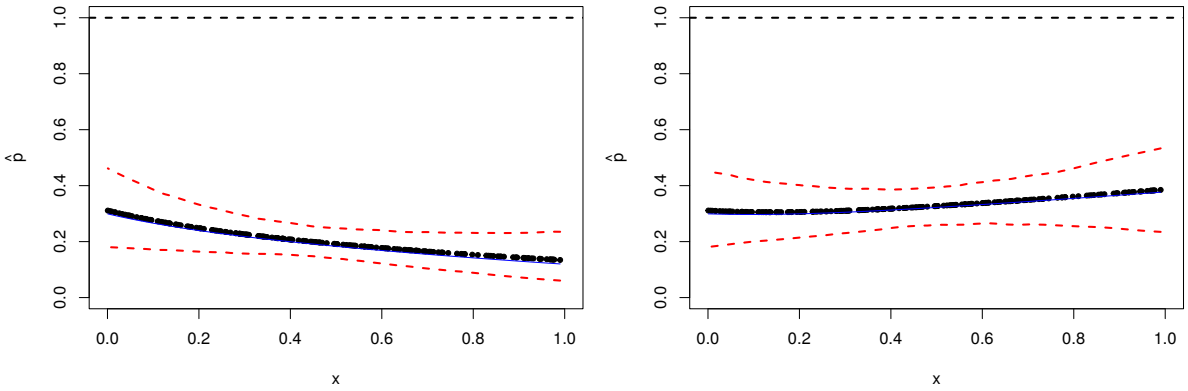
(a) Scenario 1 ($n = 50$)

(b) Scenario 2 ($n = 50$)



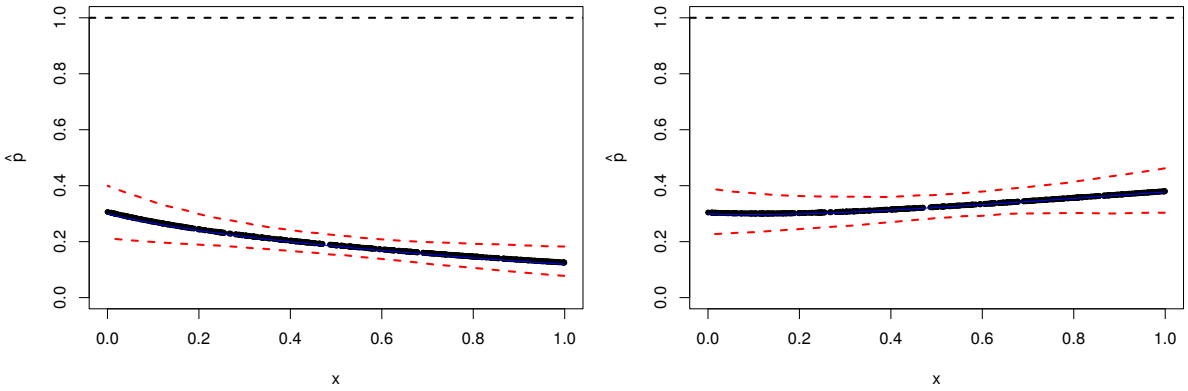
(c) Scenario 1 ($n = 100$)

(d) Scenario 2 ($n = 100$)



(e) Scenario 1 ($n = 200$)

(f) Scenario 2 ($n = 200$)

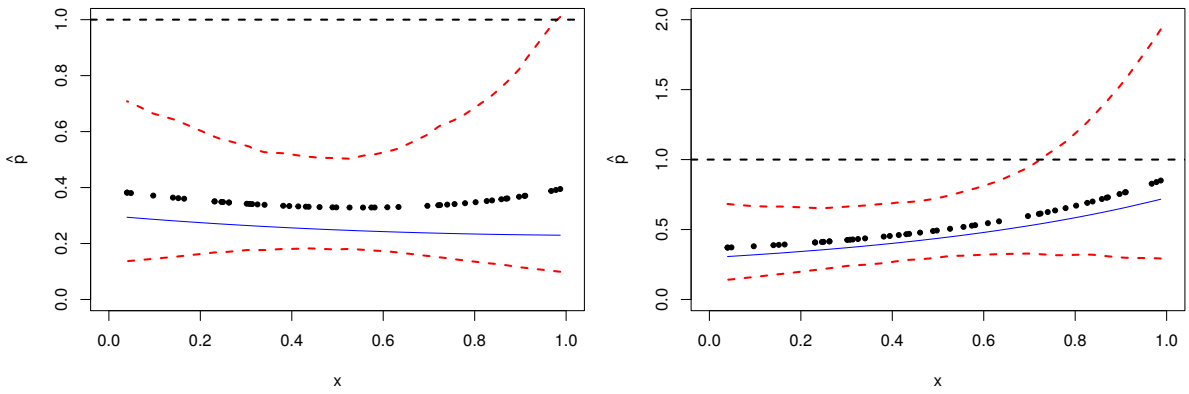


(g) Scenario 1 ($n = 500$)

(h) Scenario 2 ($n = 500$)

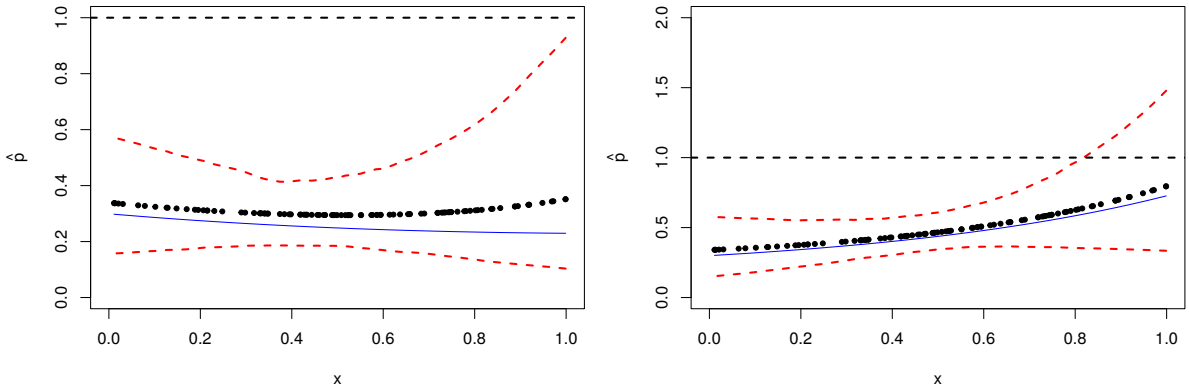
Source: Elaborated by the author.

Figure 38 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 3 and 4).



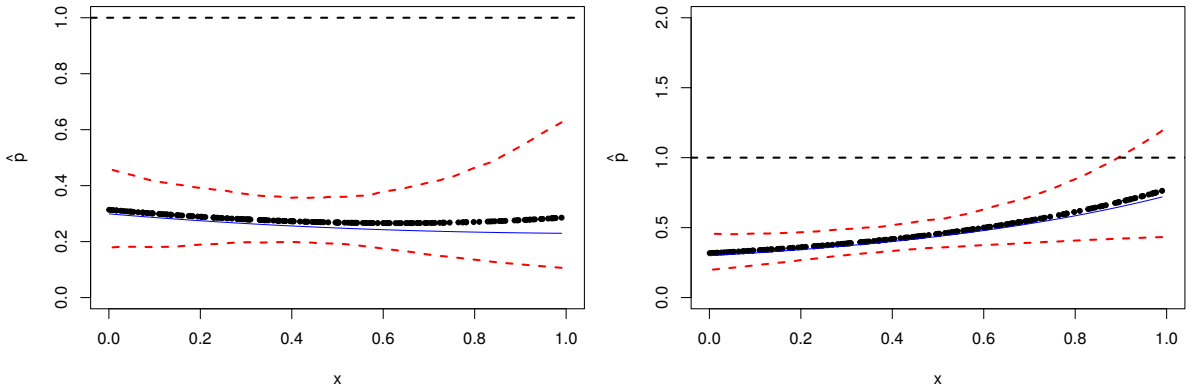
(a) Scenario 3 ($n = 50$)

(b) Scenario 4 ($n = 50$)



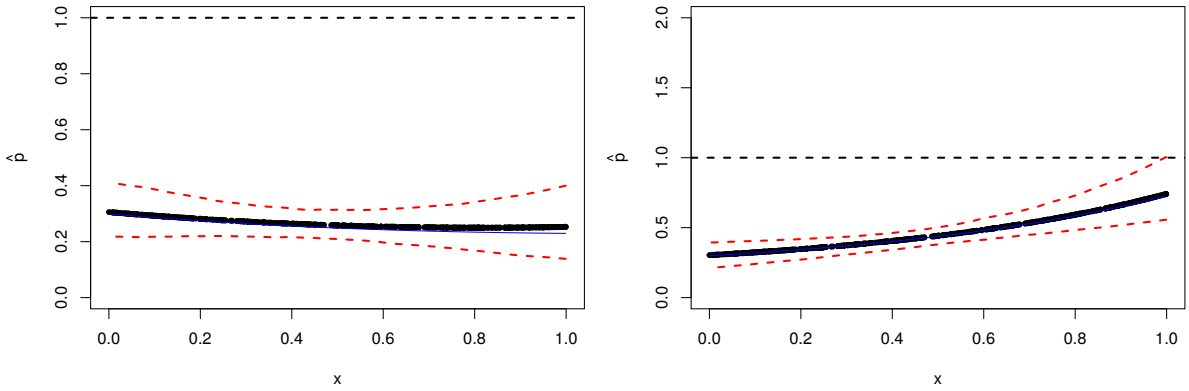
(c) Scenario 3 ($n = 100$)

(d) Scenario 4 ($n = 100$)



(e) Scenario 3 ($n = 200$)

(f) Scenario 4 ($n = 200$)



(g) Scenario 3 ($n = 500$)

(h) Scenario 4 ($n = 500$)

Source: Elaborated by the author.

C.3.1.2 Using probit link function

Table 97 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 1					
50	β_{10}	−0.013	0.992	1.000	49.716
	β_{11}	−1.969	9.844	1.284	81.167
	β_{20}	0.020	0.214	1.001	36.875
	β_{21}	0.191	0.754	1.025	69.121
100	β_{10}	−0.007	0.415	1.000	32.842
	β_{11}	−0.847	3.051	1.144	42.982
	β_{20}	0.003	0.148	1.000	30.516
	β_{21}	0.063	0.564	1.004	59.161
200	β_{10}	−0.010	0.134	1.000	18.566
	β_{11}	−0.405	1.003	1.093	24.666
	β_{20}	0.009	0.054	1.001	18.170
	β_{21}	0.010	0.266	1.000	40.134
500	β_{10}	−0.005	0.052	1.000	12.110
	β_{11}	−0.138	0.236	1.043	12.837
	β_{20}	0.001	0.025	1.000	12.704
	β_{21}	0.004	0.096	1.000	24.369
Scenario 2					
50	β_{10}	−0.213	0.299	1.086	27.133
	β_{11}	−0.117	0.586	1.012	19.488
	β_{20}	0.052	0.180	1.008	33.672
	β_{21}	−0.026	0.497	1.001	108.882
100	β_{10}	−0.131	0.143	1.066	19.972
	β_{11}	−0.026	0.292	1.001	14.063
	β_{20}	0.003	0.108	1.000	26.370
	β_{21}	−0.002	0.311	1.000	89.775
200	β_{10}	−0.056	0.055	1.030	12.321
	β_{11}	−0.018	0.121	1.001	9.102
	β_{20}	−0.002	0.035	1.000	15.016
	β_{21}	0.006	0.114	1.000	54.306
500	β_{10}	−0.030	0.024	1.020	8.184
	β_{11}	0.004	0.052	1.000	6.078
	β_{20}	0.003	0.016	1.000	10.023
	β_{21}	−0.004	0.043	1.000	32.685

Source: Elaborated by the author.

Table 98 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 3					
50	β_{10}	−0.558	1.168	1.167	54.835
	β_{11}	0.371	2.334	1.031	81.328
	β_{20}	0.064	0.217	1.009	37.291
	β_{21}	0.090	0.739	1.006	69.044
100	β_{10}	−0.335	0.687	1.093	41.996
	β_{11}	0.157	2.126	1.006	79.353
	β_{20}	0.000	0.142	1.000	29.447
	β_{21}	0.085	0.498	1.007	54.999
200	β_{10}	−0.131	0.215	1.042	23.877
	β_{11}	−0.013	1.186	1.000	57.991
	β_{20}	0.008	0.057	1.001	18.644
	β_{21}	0.024	0.260	1.001	39.806
500	β_{10}	−0.043	0.078	1.012	14.353
	β_{11}	−0.058	0.534	1.003	37.506
	β_{20}	0.007	0.022	1.001	11.511
	β_{21}	−0.001	0.092	1.000	23.509
Scenario 4					
50	β_{10}	−0.300	0.621	1.081	39.539
	β_{11}	0.201	1.467	1.014	64.091
	β_{20}	0.101	0.173	1.031	33.391
	β_{21}	−0.119	0.464	1.016	107.901
100	β_{10}	−0.116	0.306	1.023	28.224
	β_{11}	0.066	0.849	1.003	47.504
	β_{20}	−0.017	0.099	1.002	24.656
	β_{21}	0.045	0.269	1.004	82.445
200	β_{10}	−0.089	0.115	1.036	17.815
	β_{11}	0.058	0.372	1.005	32.020
	β_{20}	0.001	0.042	1.000	16.520
	β_{21}	0.016	0.132	1.001	57.849
500	β_{10}	−0.026	0.041	1.008	10.693
	β_{11}	0.001	0.136	1.000	19.688
	β_{20}	0.000	0.015	1.000	9.807
	β_{21}	0.014	0.042	1.002	33.297

Source: Elaborated by the author.

Table 99 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
50	β_{10}	1.487	1.490	0.996	−0.136	3.104
	β_{11}	1.031	1.022	2.442	−2.290	4.386
	β_{20}	−0.980	−0.964	0.462	−1.971	−0.013
	β_{21}	−0.809	−0.784	0.847	−2.737	1.072
100	β_{10}	1.493	1.495	0.644	0.330	2.654
	β_{11}	2.153	2.141	1.527	−0.182	4.495
	β_{20}	−0.997	−0.988	0.384	−1.754	−0.256
	β_{21}	−0.937	−0.926	0.748	−2.419	0.521
200	β_{10}	1.490	1.491	0.367	0.818	2.158
	β_{11}	2.595	2.585	0.916	1.128	4.080
	β_{20}	−0.991	−0.988	0.233	−1.458	−0.528
	β_{21}	−0.990	−0.980	0.516	−2.004	0.015
500	β_{10}	1.495	1.495	0.228	1.078	1.911
	β_{11}	2.862	2.857	0.465	2.022	3.707
	β_{20}	−0.999	−0.998	0.158	−1.300	−0.701
	β_{21}	−0.996	−0.993	0.310	−1.602	−0.392
Scenario 2						
50	β_{10}	1.287	1.286	0.504	0.339	2.232
	β_{11}	2.883	2.876	0.757	1.478	4.296
	β_{20}	−0.948	−0.939	0.421	−1.767	−0.139
	β_{21}	0.474	0.470	0.704	−0.863	1.819
100	β_{10}	1.369	1.369	0.355	0.656	2.079
	β_{11}	2.974	2.970	0.540	1.887	4.067
	β_{20}	−0.997	−0.992	0.328	−1.623	−0.380
	β_{21}	0.498	0.495	0.558	−0.536	1.537
200	β_{10}	1.444	1.443	0.227	0.991	1.897
	β_{11}	2.982	2.979	0.347	2.265	3.701
	β_{20}	−1.002	−1.000	0.188	−1.395	−0.612
	β_{21}	0.506	0.505	0.338	−0.186	1.199
500	β_{10}	1.470	1.470	0.152	1.176	1.763
	β_{11}	3.004	3.003	0.228	2.568	3.443
	β_{20}	−0.997	−0.996	0.124	−1.252	−0.742
	β_{21}	0.496	0.496	0.207	0.077	0.917

Source: Elaborated by the author.

Table 100 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
50	β_{10}	0.942	0.961	0.926	−1.084	2.925
	β_{11}	−1.129	−1.127	1.482	−5.479	3.199
	β_{20}	−0.936	−0.922	0.462	−1.922	0.025
	β_{21}	−0.910	−0.880	0.855	−2.863	0.986
100	β_{10}	1.165	1.180	0.758	−0.395	2.695
	β_{11}	−1.343	−1.338	1.450	−4.798	2.099
	β_{20}	−1.000	−0.991	0.377	−1.757	−0.260
	β_{21}	−0.915	−0.904	0.701	−2.390	0.539
200	β_{10}	1.369	1.375	0.445	0.502	2.224
	β_{11}	−1.513	−1.501	1.089	−3.844	0.809
	β_{20}	−0.992	−0.989	0.238	−1.458	−0.532
	β_{21}	−0.976	−0.966	0.510	−1.984	0.019
500	β_{10}	1.457	1.459	0.275	0.921	1.986
	β_{11}	−1.558	−1.551	0.728	−2.976	−0.152
	β_{20}	−0.993	−0.991	0.149	−1.293	−0.694
	β_{21}	−1.001	−0.998	0.304	−1.605	−0.399
Scenario 4						
50	β_{10}	1.200	1.211	0.729	−0.199	2.576
	β_{11}	−1.299	−1.292	1.194	−3.770	1.162
	β_{20}	−0.899	−0.891	0.404	−1.710	−0.099
	β_{21}	0.381	0.379	0.670	−0.951	1.720
100	β_{10}	1.384	1.390	0.541	0.325	2.433
	β_{11}	−1.434	−1.431	0.919	−3.288	0.421
	β_{20}	−1.017	−1.013	0.314	−1.644	−0.396
	β_{21}	0.545	0.542	0.517	−0.490	1.587
200	β_{10}	1.411	1.414	0.327	0.772	2.041
	β_{11}	−1.442	−1.440	0.607	−2.652	−0.232
	β_{20}	−0.999	−0.997	0.205	−1.392	−0.610
	β_{21}	0.516	0.516	0.363	−0.175	1.208
500	β_{10}	1.474	1.476	0.200	1.069	1.878
	β_{11}	−1.499	−1.498	0.369	−2.223	−0.776
	β_{20}	−1.000	−0.999	0.122	−1.254	−0.745
	β_{21}	0.514	0.513	0.204	0.095	0.933

Source: Elaborated by the author.

Table 101 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	2.60	94.60	2.80	1.20	93.20	5.60
	β_{11}	0.80	73.80	25.40	1.60	93.60	4.80
	β_{20}	2.40	97.40	0.20	5.20	93.80	1.00
	β_{21}	2.40	97.60	0.00	1.60	95.20	3.20
100	β_{10}	2.00	96.20	1.80	1.20	95.40	3.40
	β_{11}	0.40	87.00	12.60	1.20	95.20	3.60
	β_{20}	4.20	94.60	1.20	3.40	95.00	1.60
	β_{21}	3.00	95.80	1.20	3.00	93.80	3.20
200	β_{10}	4.20	92.40	3.40	1.00	95.20	3.80
	β_{11}	0.60	90.00	9.40	1.60	95.60	2.80
	β_{20}	3.20	94.80	2.00	2.00	96.40	1.60
	β_{21}	3.40	95.40	1.20	1.80	96.60	1.60
500	β_{10}	3.40	95.20	1.40	2.80	93.60	3.60
	β_{11}	1.00	92.60	6.40	1.60	95.00	3.40
	β_{20}	2.80	95.20	2.00	2.20	96.40	1.40
	β_{21}	2.60	95.40	2.00	2.20	95.60	2.20

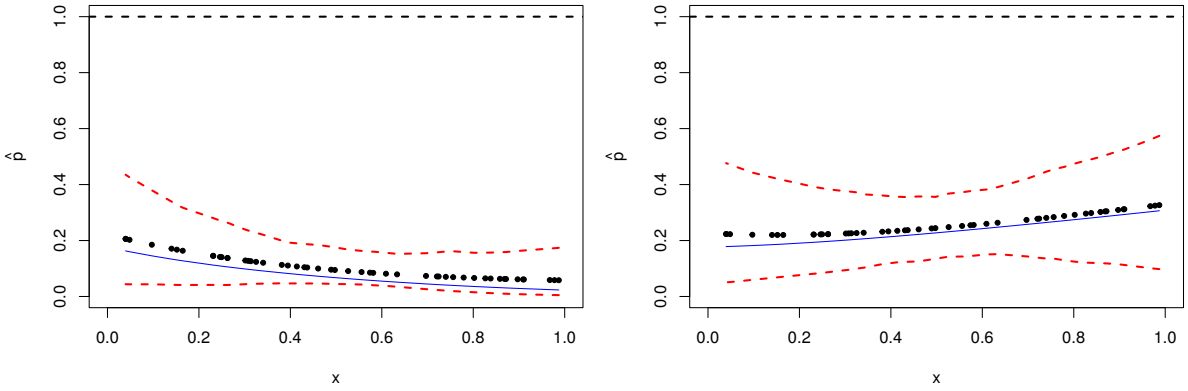
Source: Elaborated by the author.

Table 102 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	β_{10}	0.00	98.60	1.40	1.00	96.40	2.60
	β_{11}	0.60	99.40	0.00	2.80	96.40	0.80
	β_{20}	3.20	96.80	0.00	4.40	94.80	0.80
	β_{21}	2.40	97.00	0.60	1.20	94.40	4.40
100	β_{10}	0.40	98.00	1.60	2.60	96.00	1.40
	β_{11}	1.00	98.60	0.40	1.00	97.40	1.60
	β_{20}	2.60	96.60	0.80	1.60	97.20	1.20
	β_{21}	3.80	95.40	0.80	2.00	96.60	1.40
200	β_{10}	0.40	96.60	3.00	1.80	95.60	2.60
	β_{11}	1.20	98.00	0.80	2.20	96.20	1.60
	β_{20}	3.00	95.00	2.00	2.60	96.20	1.20
	β_{21}	3.80	95.80	0.40	1.60	95.40	3.00
500	β_{10}	1.80	96.20	2.00	2.20	96.00	1.80
	β_{11}	2.00	95.60	2.40	2.00	95.60	2.40
	β_{20}	3.20	94.80	2.00	2.80	95.60	1.60
	β_{21}	1.80	96.40	1.80	2.00	96.20	1.80

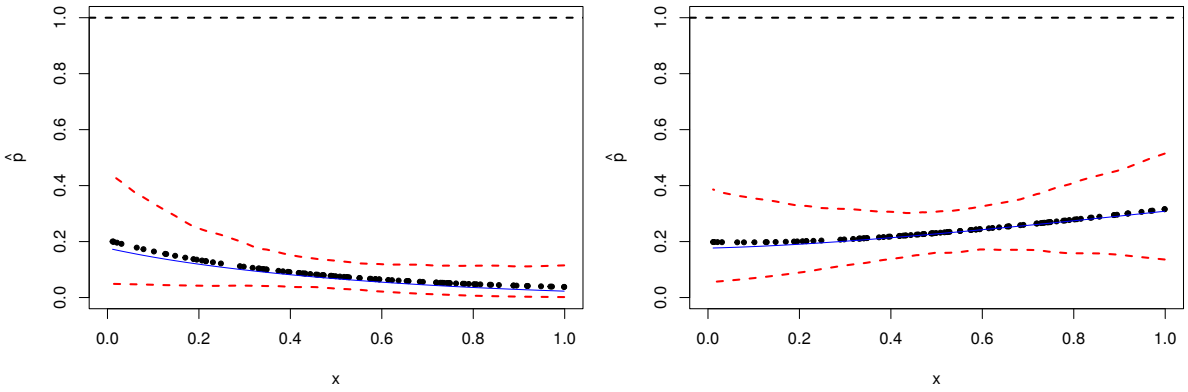
Source: Elaborated by the author.

Figure 39 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 1 and 2).



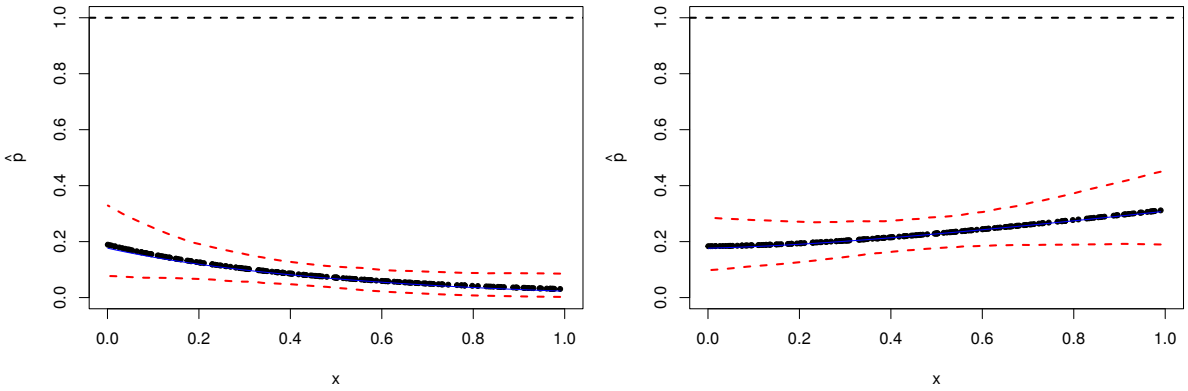
(a) Scenario 1 ($n = 50$)

(b) Scenario 2 ($n = 50$)



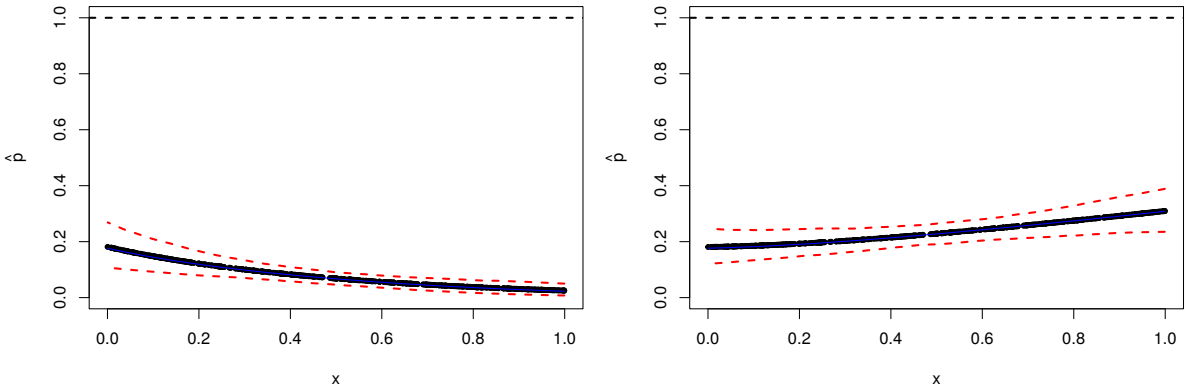
(c) Scenario 1 ($n = 100$)

(d) Scenario 2 ($n = 100$)



(e) Scenario 1 ($n = 200$)

(f) Scenario 2 ($n = 200$)

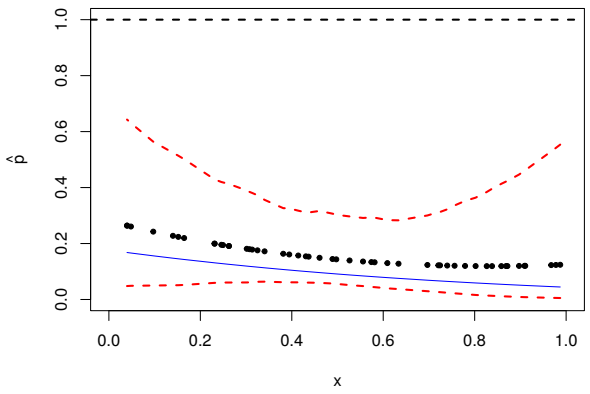


(g) Scenario 1 ($n = 500$)

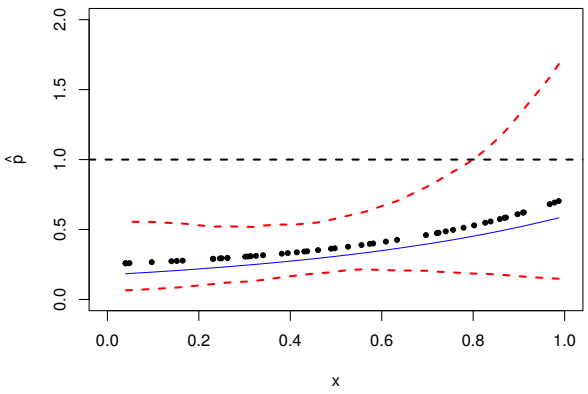
(h) Scenario 2 ($n = 500$)

Source: Elaborated by the author.

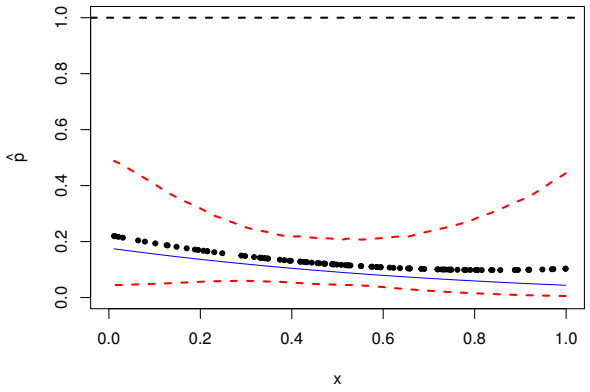
Figure 40 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 3 and 4).



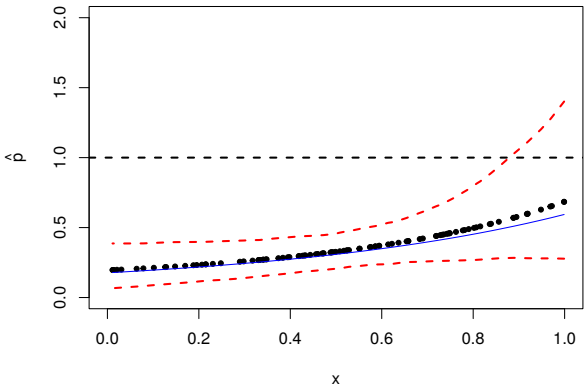
(a) Scenario 3 ($n = 50$)



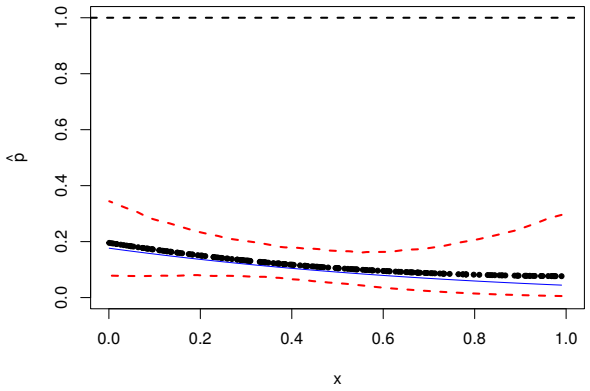
(b) Scenario 4 ($n = 50$)



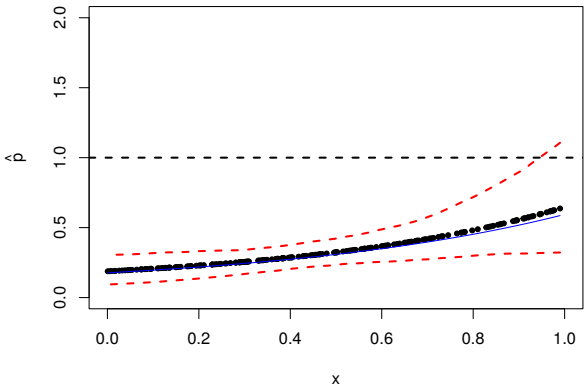
(c) Scenario 3 ($n = 100$)



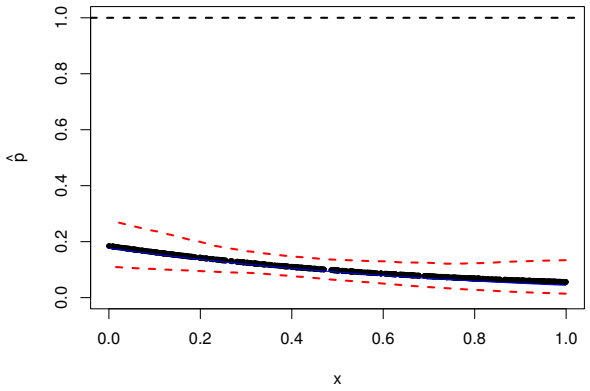
(d) Scenario 4 ($n = 100$)



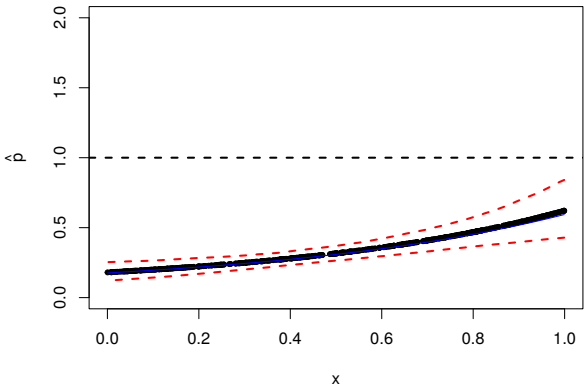
(e) Scenario 3 ($n = 200$)



(f) Scenario 4 ($n = 200$)



(g) Scenario 3 ($n = 500$)



(h) Scenario 4 ($n = 500$)

Source: Elaborated by the author.

C.3.1.3 Using complementary log-log link function

Table 103 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 1		
50	β_{10}	−0.145	0.290	1.038	27.436
	β_{11}	−0.369	1.254	1.059	26.086
	β_{20}	0.028	0.260	1.002	40.885
	β_{21}	0.222	0.816	1.032	72.841
100	β_{10}	−0.090	0.132	1.032	19.110
	β_{11}	−0.148	0.415	1.028	16.902
	β_{20}	0.015	0.183	1.001	35.014
	β_{21}	0.122	0.604	1.012	61.264
200	β_{10}	−0.022	0.047	1.005	11.434
	β_{11}	−0.107	0.185	1.032	11.402
	β_{20}	0.018	0.074	1.002	21.593
	β_{21}	0.060	0.296	1.006	42.673
500	β_{10}	−0.011	0.018	1.003	7.083
	β_{11}	−0.038	0.061	1.012	6.449
	β_{20}	0.005	0.031	1.000	13.982
	β_{21}	0.025	0.116	1.003	26.966
			Scenario 2		
50	β_{10}	−0.116	0.164	1.044	20.395
	β_{11}	−0.179	0.375	1.046	15.972
	β_{20}	0.098	0.243	1.020	39.207
	β_{21}	−0.083	0.657	1.005	127.997
100	β_{10}	−0.060	0.074	1.026	14.508
	β_{11}	−0.106	0.192	1.030	11.570
	β_{20}	0.039	0.129	1.006	28.591
	β_{21}	−0.018	0.357	1.000	96.053
200	β_{10}	−0.052	0.034	1.042	9.499
	β_{11}	−0.015	0.082	1.001	7.587
	β_{20}	0.022	0.054	1.004	18.028
	β_{21}	−0.021	0.154	1.001	61.469
500	β_{10}	−0.017	0.012	1.012	5.755
	β_{11}	−0.004	0.028	1.000	4.473
	β_{20}	0.006	0.023	1.001	12.253
	β_{21}	0.000	0.062	1.000	40.099

Source: Elaborated by the author.

Table 104 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 3					
50	β_{10}	−0.286	0.547	1.085	36.060
	β_{11}	0.242	1.647	1.018	67.210
	β_{20}	0.026	0.281	1.001	41.631
	β_{21}	0.248	0.961	1.034	78.137
100	β_{10}	−0.108	0.245	1.025	24.112
	β_{11}	−0.014	0.921	1.000	49.679
	β_{20}	0.010	0.174	1.000	33.560
	β_{21}	0.120	0.569	1.013	60.454
200	β_{10}	−0.055	0.083	1.019	14.855
	β_{11}	−0.023	0.442	1.001	34.941
	β_{20}	0.012	0.072	1.001	21.081
	β_{21}	0.036	0.278	1.002	41.432
500	β_{10}	−0.024	0.033	1.008	9.631
	β_{11}	−0.024	0.166	1.002	21.483
	β_{20}	0.002	0.032	1.000	14.379
	β_{21}	0.014	0.110	1.001	26.533
Scenario 4					
50	β_{10}	−0.232	0.352	1.087	29.729
	β_{11}	0.165	0.972	1.014	52.225
	β_{20}	0.048	0.209	1.006	37.053
	β_{21}	−0.007	0.550	1.000	117.533
100	β_{10}	−0.116	0.147	1.049	20.012
	β_{11}	0.078	0.434	1.007	34.804
	β_{20}	0.037	0.125	1.006	28.175
	β_{21}	−0.017	0.331	1.000	92.817
200	β_{10}	−0.043	0.066	1.014	13.459
	β_{11}	0.017	0.265	1.000	26.876
	β_{20}	0.029	0.054	1.008	18.318
	β_{21}	−0.031	0.174	1.003	65.728
500	β_{10}	−0.006	0.022	1.001	7.707
	β_{11}	−0.015	0.078	1.001	14.695
	β_{20}	0.005	0.022	1.001	11.691
	β_{21}	−0.012	0.059	1.001	38.117

Source: Elaborated by the author.

Table 105 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
50	β_{10}	1.355	1.355	0.519	0.438	2.270
	β_{11}	2.631	2.620	1.057	1.022	4.252
	β_{20}	−0.972	−0.945	0.510	−2.072	0.088
	β_{21}	−0.778	−0.768	0.876	−2.783	1.212
100	β_{10}	1.410	1.411	0.352	0.740	2.079
	β_{11}	2.852	2.844	0.627	1.679	4.037
	β_{20}	−0.985	−0.969	0.427	−1.837	−0.159
	β_{21}	−0.878	−0.877	0.767	−2.449	0.681
200	β_{10}	1.478	1.478	0.216	1.073	1.886
	β_{11}	2.893	2.889	0.417	2.131	3.662
	β_{20}	−0.982	−0.975	0.271	−1.518	−0.454
	β_{21}	−0.940	−0.935	0.541	−2.034	0.147
500	β_{10}	1.489	1.489	0.135	1.226	1.752
	β_{11}	2.962	2.961	0.243	2.500	3.423
	β_{20}	−0.995	−0.992	0.176	−1.348	−0.648
	β_{21}	−0.975	−0.974	0.339	−1.645	−0.307
Scenario 2						
50	β_{10}	1.384	1.384	0.388	0.688	2.080
	β_{11}	2.821	2.817	0.586	1.763	3.881
	β_{20}	−0.902	−0.882	0.484	−1.844	0.004
	β_{21}	0.417	0.414	0.806	−1.104	1.944
100	β_{10}	1.440	1.440	0.265	0.917	1.964
	β_{11}	2.894	2.892	0.425	2.079	3.714
	β_{20}	−0.961	−0.950	0.357	−1.685	−0.253
	β_{21}	0.482	0.478	0.597	−0.709	1.674
200	β_{10}	1.448	1.448	0.176	1.109	1.786
	β_{11}	2.985	2.983	0.286	2.434	3.537
	β_{20}	−0.978	−0.973	0.232	−1.441	−0.523
	β_{21}	0.479	0.480	0.392	−0.329	1.286
500	β_{10}	1.483	1.483	0.108	1.263	1.704
	β_{11}	2.996	2.995	0.167	2.660	3.333
	β_{20}	−0.994	−0.991	0.152	−1.297	−0.694
	β_{21}	0.500	0.499	0.248	0.009	0.994

Source: Elaborated by the author.

Table 106 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
50	β_{10}	1.214	1.223	0.682	−0.069	2.474
	β_{11}	−1.258	−1.248	1.260	−3.918	1.391
	β_{20}	−0.974	−0.949	0.530	−2.073	0.081
	β_{21}	−0.752	−0.743	0.949	−2.749	1.224
100	β_{10}	1.392	1.398	0.483	0.477	2.293
	β_{11}	−1.514	−1.510	0.960	−3.446	0.395
	β_{20}	−0.990	−0.974	0.418	−1.845	−0.162
	β_{21}	−0.880	−0.878	0.745	−2.450	0.691
200	β_{10}	1.445	1.447	0.283	0.903	1.986
	β_{11}	−1.523	−1.518	0.664	−2.827	−0.225
	β_{20}	−0.988	−0.981	0.268	−1.530	−0.457
	β_{21}	−0.964	−0.959	0.526	−2.069	0.134
500	β_{10}	1.476	1.478	0.180	1.129	1.821
	β_{11}	−1.524	−1.522	0.407	−2.307	−0.745
	β_{20}	−0.998	−0.995	0.180	−1.350	−0.648
	β_{21}	−0.986	−0.985	0.331	−1.659	−0.315
Scenario 4						
50	β_{10}	1.268	1.275	0.546	0.231	2.295
	β_{11}	−1.335	−1.330	0.972	−3.202	0.524
	β_{20}	−0.952	−0.933	0.455	−1.901	−0.037
	β_{21}	0.493	0.490	0.742	−1.036	2.025
100	β_{10}	1.384	1.389	0.365	0.624	2.141
	β_{11}	−1.422	−1.421	0.654	−2.805	−0.033
	β_{20}	−0.963	−0.951	0.352	−1.690	−0.256
	β_{21}	0.483	0.480	0.575	−0.706	1.677
200	β_{10}	1.457	1.458	0.254	0.990	1.921
	β_{11}	−1.483	−1.482	0.515	−2.415	−0.551
	β_{20}	−0.971	−0.967	0.231	−1.433	−0.517
	β_{21}	0.469	0.469	0.416	−0.337	1.275
500	β_{10}	1.494	1.494	0.146	1.193	1.792
	β_{11}	−1.515	−1.514	0.278	−2.074	−0.957
	β_{20}	−0.995	−0.993	0.146	−1.299	−0.694
	β_{21}	0.488	0.488	0.242	−0.004	0.983

Source: Elaborated by the author.

Table 107 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	1.60	92.40	6.00	2.00	91.80	6.20
	β_{11}	2.00	89.40	8.60	1.80	91.20	7.00
	β_{20}	4.60	95.20	0.20	6.40	92.60	1.00
	β_{21}	1.80	97.80	0.40	2.20	94.40	3.40
100	β_{10}	1.40	94.60	4.00	2.60	94.60	2.80
	β_{11}	0.40	94.00	5.60	0.80	93.40	5.80
	β_{20}	4.80	94.80	0.40	3.40	95.00	1.60
	β_{21}	3.40	95.40	1.20	1.40	95.00	3.60
200	β_{10}	2.20	94.00	3.80	1.40	94.80	3.80
	β_{11}	1.60	93.40	5.00	1.80	94.80	3.40
	β_{20}	4.60	94.40	1.00	4.40	94.40	1.20
	β_{21}	2.60	96.00	1.40	1.40	95.60	3.00
500	β_{10}	1.80	95.00	3.20	1.00	96.60	2.40
	β_{11}	2.00	93.40	4.60	1.40	96.80	1.80
	β_{20}	2.80	95.20	2.00	2.20	96.80	1.00
	β_{21}	2.80	95.20	2.00	2.20	95.20	2.60

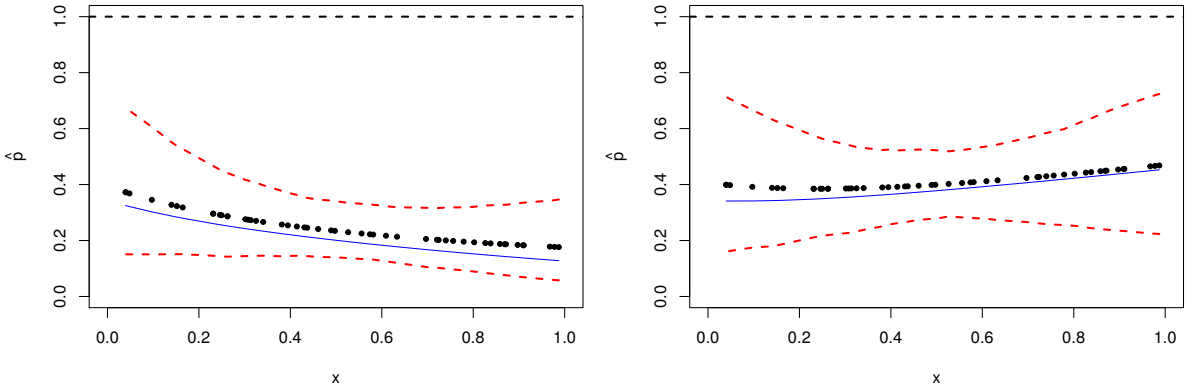
Source: Elaborated by the author.

Table 108 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	β_{10}	1.40	93.80	4.80	0.60	95.60	3.80
	β_{11}	2.20	97.00	0.80	3.40	95.60	1.00
	β_{20}	2.20	96.80	1.00	4.00	95.80	0.20
	β_{21}	3.60	96.40	0.00	1.60	96.20	2.20
100	β_{10}	1.40	95.60	3.00	1.20	96.60	2.20
	β_{11}	2.60	95.80	1.60	2.00	96.00	2.00
	β_{20}	4.20	95.00	0.80	3.60	95.40	1.00
	β_{21}	1.80	97.40	0.80	1.00	97.00	2.00
200	β_{10}	2.40	95.20	2.40	1.80	94.00	4.20
	β_{11}	1.80	96.20	2.00	3.40	93.40	3.20
	β_{20}	2.60	95.80	1.60	2.60	96.20	1.20
	β_{21}	2.80	96.20	1.00	2.40	94.60	3.00
500	β_{10}	2.20	94.40	3.40	3.00	94.80	2.20
	β_{11}	2.00	95.20	2.80	1.40	95.40	3.20
	β_{20}	3.40	94.80	1.80	2.80	96.20	1.00
	β_{21}	1.60	96.60	1.80	1.80	95.40	2.80

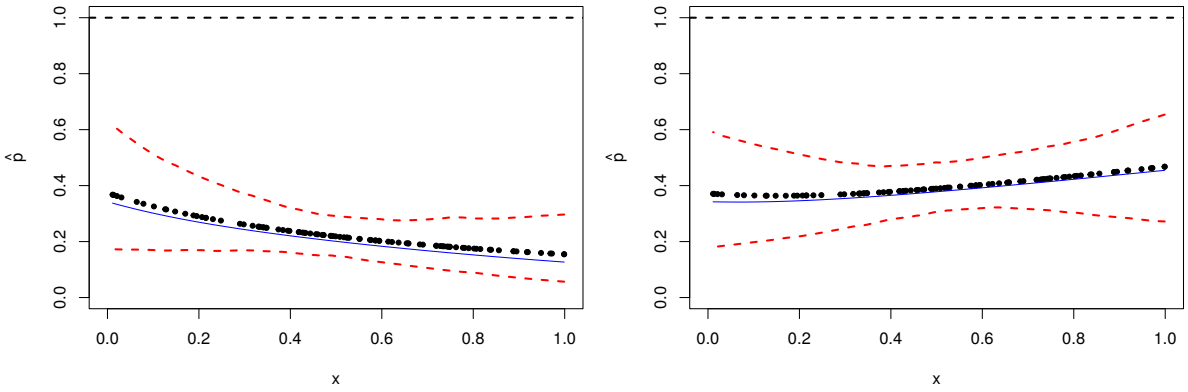
Source: Elaborated by the author.

Figure 41 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 1 and 2).



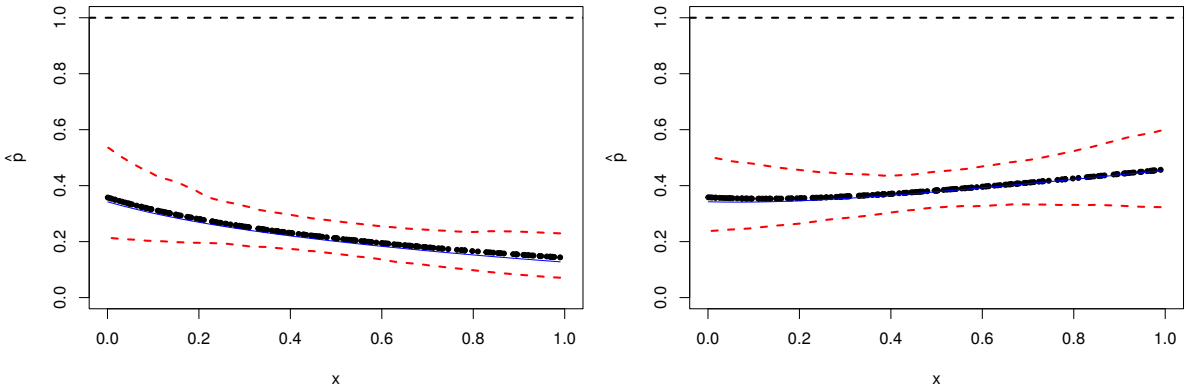
(a) Scenario 1 ($n = 50$)

(b) Scenario 2 ($n = 50$)



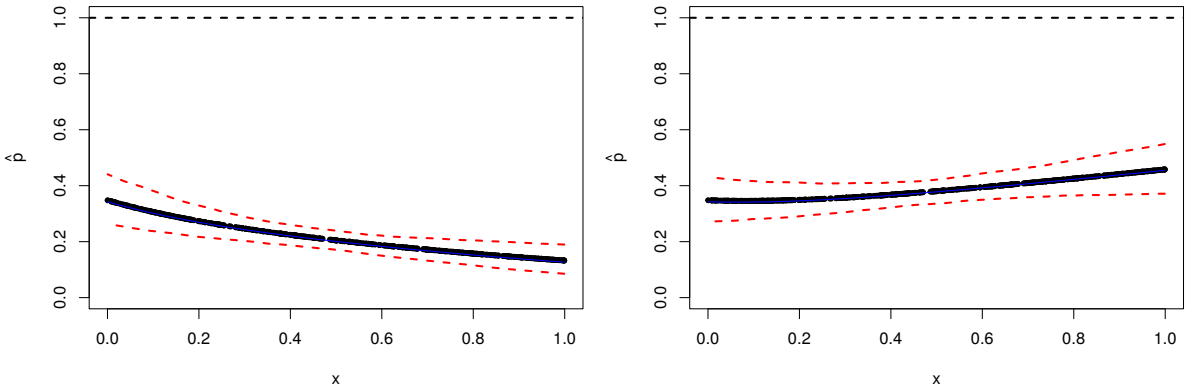
(c) Scenario 1 ($n = 100$)

(d) Scenario 2 ($n = 100$)



(e) Scenario 1 ($n = 200$)

(f) Scenario 2 ($n = 200$)

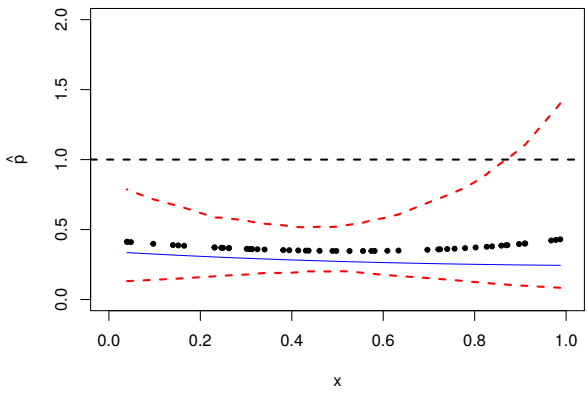


(g) Scenario 1 ($n = 500$)

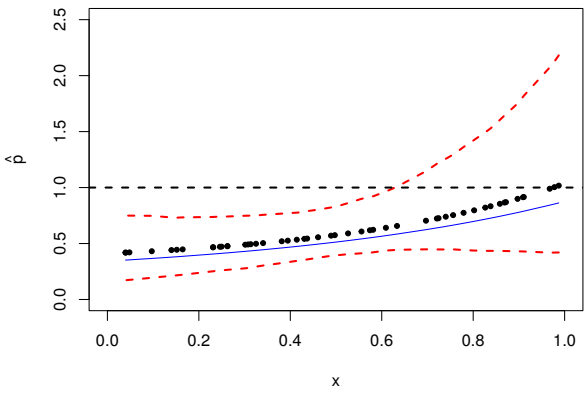
(h) Scenario 2 ($n = 500$)

Source: Elaborated by the author.

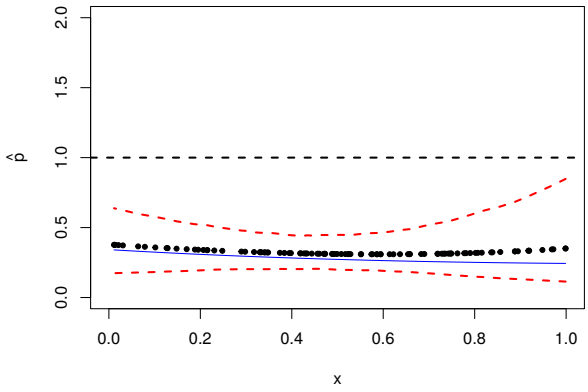
Figure 42 – *Posterior* estimates of parameter p using zero-inflated samples (Scenarios 3 and 4).



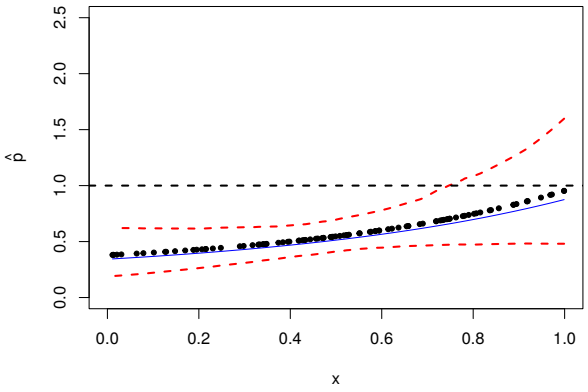
(a) Scenario 3 ($n = 50$)



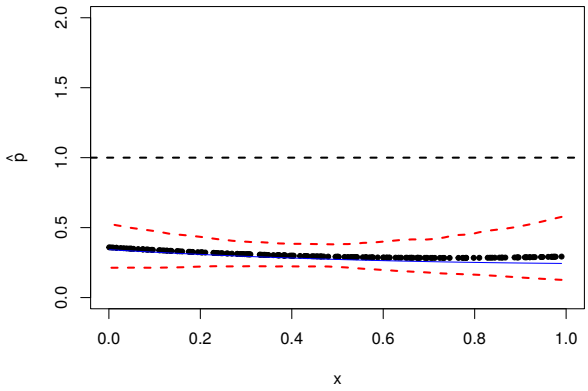
(b) Scenario 4 ($n = 50$)



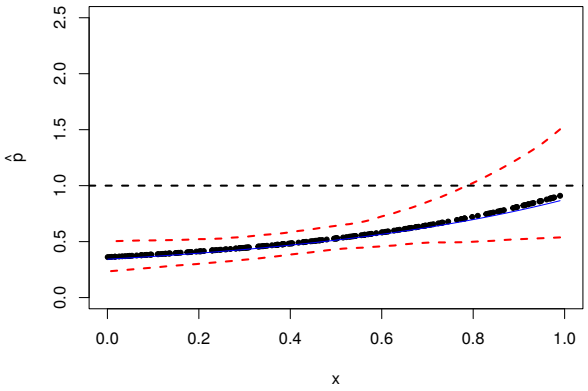
(c) Scenario 3 ($n = 100$)



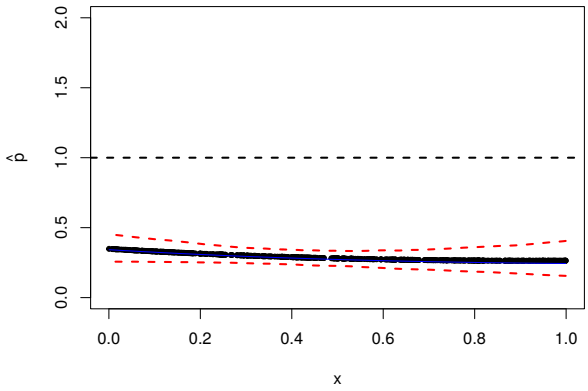
(d) Scenario 4 ($n = 100$)



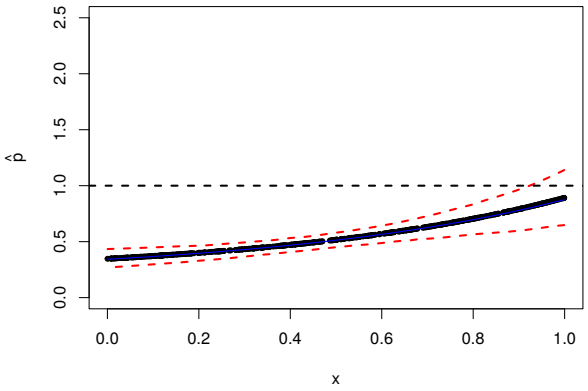
(e) Scenario 3 ($n = 200$)



(f) Scenario 4 ($n = 200$)



(g) Scenario 3 ($n = 500$)



(h) Scenario 4 ($n = 500$)

Source: Elaborated by the author.

C.3.2 Zero-deflated artificial data

C.3.2.1 Using logit link function

Table 109 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 1		
50	β_{10}	0.116	0.342	1.020	46.269
	β_{11}	−0.209	0.905	1.025	75.544
	β_{20}	−0.071	0.298	1.008	85.289
	β_{21}	−0.017	0.904	1.000	148.588
100	β_{10}	0.040	0.221	1.004	36.505
	β_{11}	−0.069	0.581	1.004	60.373
	β_{20}	−0.010	0.189	1.000	68.498
	β_{21}	−0.061	0.566	1.003	118.380
200	β_{10}	0.043	0.093	1.010	25.193
	β_{11}	−0.067	0.268	1.008	41.512
	β_{20}	−0.016	0.079	1.002	44.834
	β_{21}	−0.005	0.277	1.000	85.763
500	β_{10}	−0.004	0.041	1.000	16.515
	β_{11}	0.004	0.105	1.000	25.966
	β_{20}	0.005	0.035	1.000	30.158
	β_{21}	−0.022	0.107	1.002	52.544
			Scenario 2		
50	β_{10}	0.070	0.329	1.007	44.744
	β_{11}	−0.121	0.928	1.008	75.648
	β_{20}	−0.188	0.383	1.050	33.581
	β_{21}	0.144	0.939	1.011	78.153
100	β_{10}	0.040	0.185	1.004	33.702
	β_{11}	−0.075	0.508	1.006	56.059
	β_{20}	−0.069	0.251	1.010	27.134
	β_{21}	0.028	0.648	1.001	64.126
200	β_{10}	0.004	0.082	1.000	22.578
	β_{11}	−0.007	0.269	1.000	41.316
	β_{20}	−0.066	0.099	1.023	16.943
	β_{21}	0.074	0.297	1.010	43.689
500	β_{10}	0.013	0.038	1.002	15.487
	β_{11}	−0.021	0.100	1.002	25.244
	β_{20}	−0.022	0.045	1.006	11.579
	β_{21}	0.018	0.121	1.001	27.868

Source: Elaborated by the author.

Table 110 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 3					
50	β_{10}	0.007	0.416	1.000	51.077
	β_{11}	0.578	1.484	1.136	63.871
	β_{20}	−0.046	0.298	1.004	88.854
	β_{21}	−0.032	0.908	1.000	156.233
100	β_{10}	0.004	0.320	1.000	45.682
	β_{11}	0.249	1.085	1.030	55.080
	β_{20}	−0.039	0.198	1.004	70.640
	β_{21}	−0.012	0.563	1.000	120.764
200	β_{10}	0.001	0.155	1.000	31.560
	β_{11}	0.142	0.607	1.017	41.578
	β_{20}	−0.023	0.080	1.003	45.237
	β_{21}	−0.004	0.265	1.000	81.871
500	β_{10}	−0.006	0.064	1.000	20.189
	β_{11}	0.077	0.237	1.013	25.468
	β_{20}	−0.002	0.033	1.000	28.544
	β_{21}	−0.009	0.098	1.000	49.648
Scenario 4					
50	β_{10}	−0.009	0.356	1.000	47.455
	β_{11}	0.562	1.376	1.139	62.652
	β_{20}	−0.126	0.372	1.022	32.045
	β_{21}	0.058	1.017	1.002	80.557
100	β_{10}	−0.053	0.294	1.005	42.976
	β_{11}	0.366	1.155	1.064	56.965
	β_{20}	−0.103	0.239	1.023	26.300
	β_{21}	0.113	0.650	1.010	65.186
200	β_{10}	0.012	0.113	1.001	27.492
	β_{11}	0.154	0.530	1.023	39.172
	β_{20}	−0.059	0.108	1.016	17.372
	β_{21}	0.070	0.335	1.007	46.286
500	β_{10}	−0.013	0.052	1.002	17.895
	β_{11}	0.086	0.207	1.018	24.165
	β_{20}	−0.013	0.045	1.002	11.191
	β_{21}	0.017	0.122	1.001	27.718

Source: Elaborated by the author.

Table 111 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
50	β_{10}	−0.884	−0.871	0.574	−2.129	0.336
	β_{11}	0.791	0.787	0.928	−1.156	2.737
	β_{20}	0.429	0.423	0.541	−0.722	1.591
	β_{21}	0.483	0.478	0.951	−1.517	2.486
100	β_{10}	−0.960	−0.952	0.469	−1.917	−0.013
	β_{11}	0.931	0.928	0.759	−0.590	2.455
	β_{20}	0.490	0.485	0.435	−0.399	1.387
	β_{21}	0.439	0.437	0.750	−1.119	1.995
200	β_{10}	−0.957	−0.953	0.301	−1.568	−0.351
	β_{11}	0.933	0.933	0.513	−0.086	1.956
	β_{20}	0.484	0.482	0.280	−0.078	1.051
	β_{21}	0.495	0.493	0.526	−0.564	1.554
500	β_{10}	−1.004	−1.003	0.203	−1.408	−0.602
	β_{11}	1.004	1.003	0.324	0.376	1.633
	β_{20}	0.505	0.503	0.187	0.137	0.874
	β_{21}	0.478	0.478	0.327	−0.164	1.121
Scenario 2						
50	β_{10}	−0.930	−0.917	0.569	−2.095	0.212
	β_{11}	0.879	0.878	0.956	−1.013	2.782
	β_{20}	1.312	1.294	0.589	0.063	2.598
	β_{21}	−0.856	−0.848	0.958	−2.945	1.212
100	β_{10}	−0.960	−0.953	0.429	−1.852	−0.080
	β_{11}	0.925	0.923	0.709	−0.548	2.405
	β_{20}	1.431	1.420	0.496	0.445	2.442
	β_{21}	−0.972	−0.964	0.804	−2.626	0.673
200	β_{10}	−0.996	−0.993	0.287	−1.564	−0.431
	β_{11}	0.993	0.993	0.519	−0.001	1.986
	β_{20}	1.434	1.428	0.308	0.800	2.077
	β_{21}	−0.926	−0.923	0.540	−2.041	0.189
500	β_{10}	−0.987	−0.986	0.194	−1.359	−0.618
	β_{11}	0.979	0.979	0.315	0.372	1.587
	β_{20}	1.478	1.475	0.211	1.058	1.902
	β_{21}	−0.982	−0.980	0.347	−1.668	−0.300

Source: Elaborated by the author.

Table 112 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
50	β_{10}	−0.993	−0.970	0.645	−2.515	0.496
	β_{11}	−0.922	−0.910	1.073	−3.654	1.794
	β_{20}	0.454	0.447	0.544	−0.698	1.624
	β_{21}	0.468	0.463	0.952	−1.546	2.478
100	β_{10}	−0.996	−0.980	0.566	−2.206	0.185
	β_{11}	−1.251	−1.244	1.012	−3.489	0.970
	β_{20}	0.461	0.457	0.443	−0.427	1.356
	β_{21}	0.488	0.486	0.750	−1.067	2.045
200	β_{10}	−0.999	−0.993	0.394	−1.764	−0.242
	β_{11}	−1.359	−1.349	0.766	−2.936	0.204
	β_{20}	0.477	0.476	0.282	−0.086	1.041
	β_{21}	0.496	0.494	0.515	−0.555	1.557
500	β_{10}	−1.006	−1.003	0.252	−1.505	−0.512
	β_{11}	−1.423	−1.419	0.480	−2.388	−0.466
	β_{20}	0.498	0.497	0.182	0.130	0.868
	β_{21}	0.491	0.491	0.312	−0.152	1.134
Scenario 4						
50	β_{10}	−1.009	−0.989	0.596	−2.426	0.373
	β_{11}	−0.938	−0.921	1.030	−3.628	1.734
	β_{20}	1.374	1.355	0.597	0.109	2.667
	β_{21}	−0.942	−0.934	1.007	−3.039	1.143
100	β_{10}	−1.053	−1.040	0.539	−2.169	0.040
	β_{11}	−1.134	−1.127	1.011	−3.277	1.003
	β_{20}	1.397	1.386	0.478	0.412	2.401
	β_{21}	−0.887	−0.880	0.798	−2.538	0.761
200	β_{10}	−0.988	−0.982	0.335	−1.682	−0.302
	β_{11}	−1.346	−1.335	0.711	−2.861	0.154
	β_{20}	1.441	1.436	0.322	0.805	2.084
	β_{21}	−0.930	−0.928	0.574	−2.045	0.189
500	β_{10}	−1.013	−1.010	0.227	−1.468	−0.563
	β_{11}	−1.414	−1.410	0.447	−2.350	−0.485
	β_{20}	1.487	1.484	0.213	1.066	1.912
	β_{21}	−0.983	−0.981	0.348	−1.670	−0.299

Source: Elaborated by the author.

Table 113 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	3.40	96.20	0.40	4.40	94.40	1.20
	β_{11}	0.80	96.60	2.60	2.20	95.20	2.60
	β_{20}	0.40	97.00	2.60	0.20	95.60	4.20
	β_{21}	1.40	97.60	1.00	2.40	97.40	0.20
100	β_{10}	3.80	94.60	1.60	2.20	96.80	1.00
	β_{11}	1.60	96.40	2.00	1.60	95.60	2.80
	β_{20}	1.60	96.00	2.40	0.80	95.20	4.00
	β_{21}	1.40	96.20	2.40	2.60	96.00	1.40
200	β_{10}	3.40	96.20	0.40	3.00	95.00	2.00
	β_{11}	1.20	96.60	2.20	3.80	93.60	2.60
	β_{20}	0.80	96.00	3.20	0.80	96.00	3.20
	β_{21}	1.20	97.00	1.80	2.40	96.40	1.20
500	β_{10}	2.20	95.00	2.80	2.60	95.20	2.20
	β_{11}	2.40	94.80	2.80	2.00	95.60	2.40
	β_{20}	0.80	95.80	3.40	1.40	95.20	3.40
	β_{21}	2.40	95.80	1.80	2.80	95.00	2.20

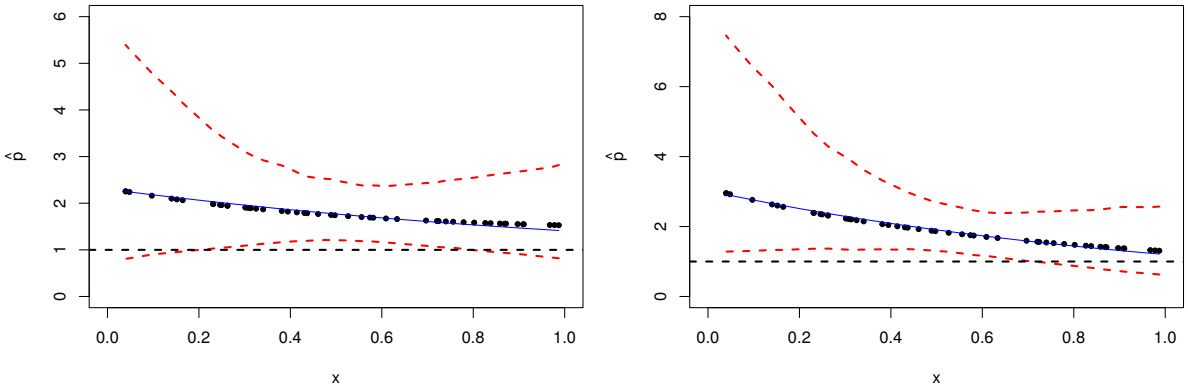
Source: Elaborated by the author.

Table 114 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	β_{10}	2.20	97.60	0.20	1.40	98.60	0.00
	β_{11}	3.20	96.80	0.00	2.60	97.40	0.00
	β_{20}	0.40	97.60	2.00	1.00	96.00	3.00
	β_{21}	0.80	97.60	1.60	2.00	96.60	1.40
100	β_{10}	2.40	97.00	0.60	1.20	97.60	1.20
	β_{11}	1.80	97.60	0.60	3.40	96.40	0.20
	β_{20}	2.00	95.00	3.00	0.80	94.60	4.60
	β_{21}	1.60	96.40	2.00	2.80	96.20	1.00
200	β_{10}	3.40	94.40	2.20	3.40	95.80	0.80
	β_{11}	4.00	95.20	0.80	2.80	96.20	1.00
	β_{20}	2.20	95.60	2.20	2.00	94.80	3.20
	β_{21}	1.20	96.80	2.00	3.60	94.40	2.00
500	β_{10}	3.40	94.60	2.00	2.20	94.80	3.00
	β_{11}	4.00	94.40	1.60	3.40	95.80	0.80
	β_{20}	2.20	94.60	3.20	1.20	95.20	3.60
	β_{21}	1.40	96.80	1.80	3.20	95.40	1.40

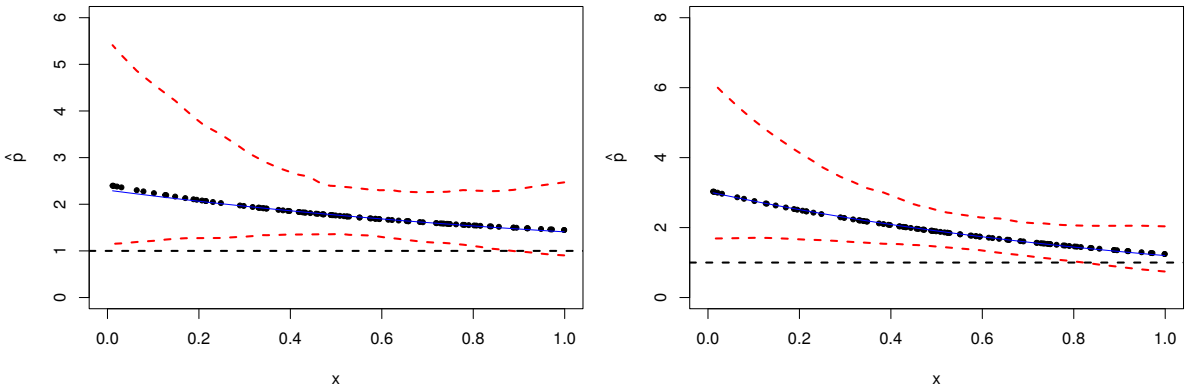
Source: Elaborated by the author.

Figure 43 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 1 and 2).



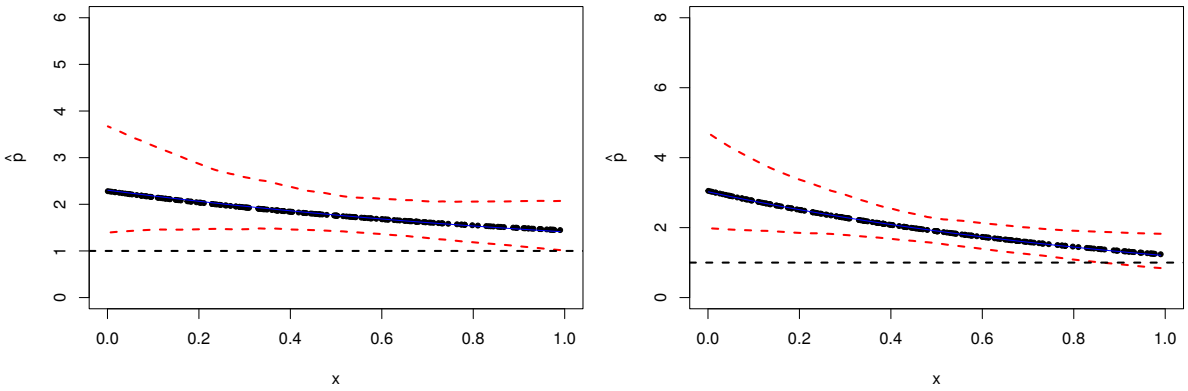
(a) Scenario 1 ($n = 50$)

(b) Scenario 2 ($n = 50$)



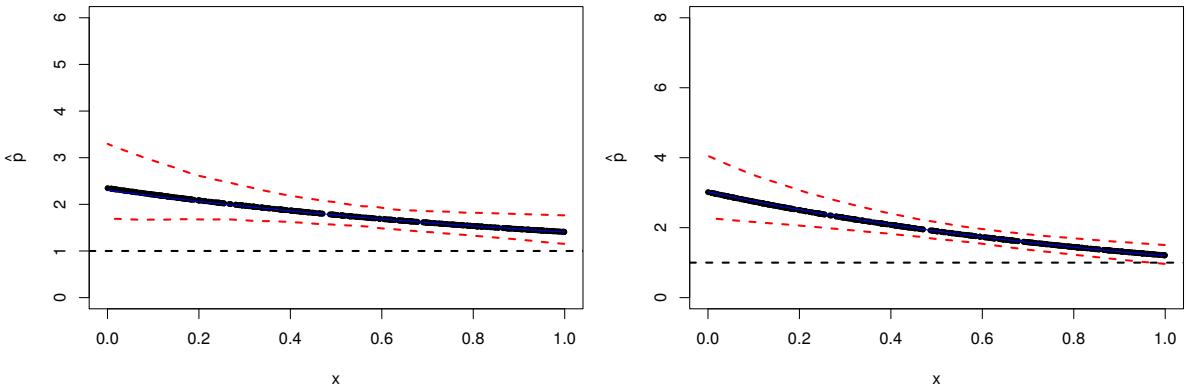
(c) Scenario 1 ($n = 100$)

(d) Scenario 2 ($n = 100$)



(e) Scenario 1 ($n = 200$)

(f) Scenario 2 ($n = 200$)

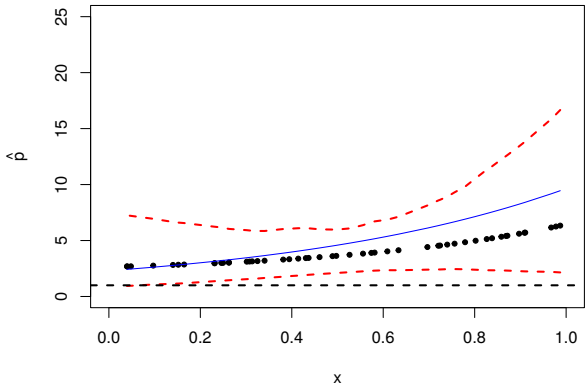


(g) Scenario 1 ($n = 500$)

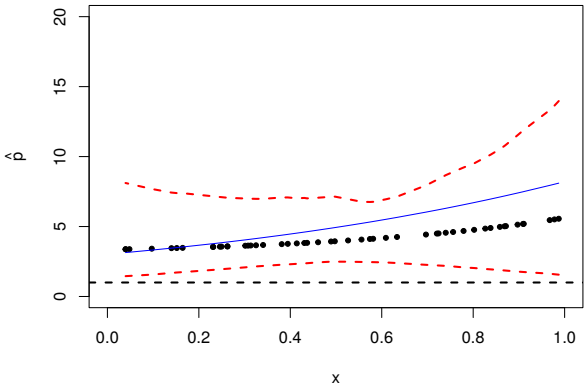
(h) Scenario 2 ($n = 500$)

Source: Elaborated by the author.

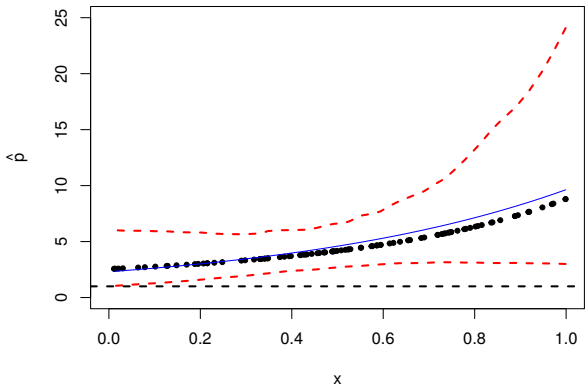
Figure 44 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 3 and 4).



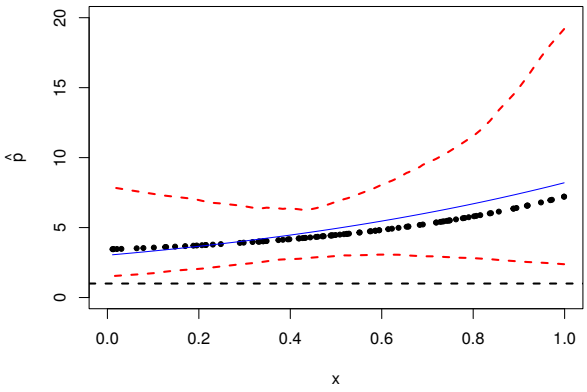
(a) Scenario 3 ($n = 50$)



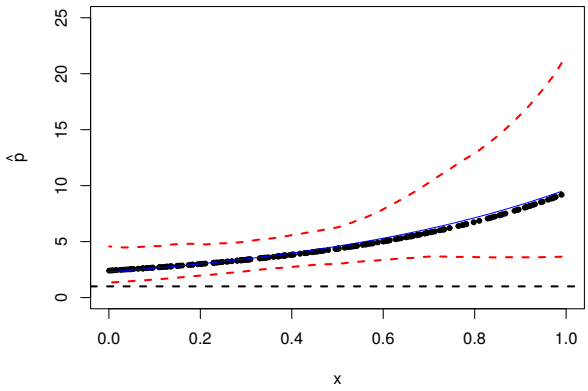
(b) Scenario 4 ($n = 50$)



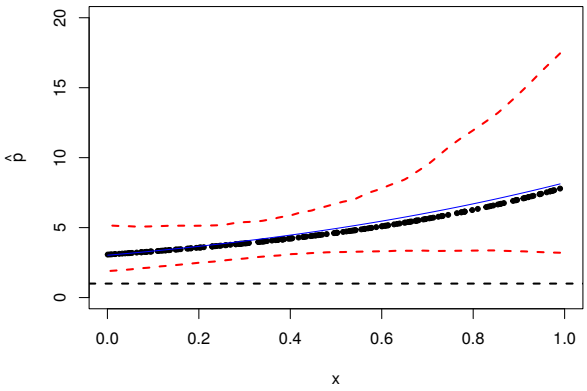
(c) Scenario 3 ($n = 100$)



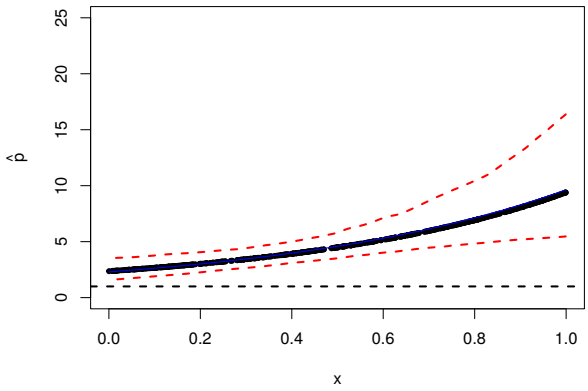
(d) Scenario 4 ($n = 100$)



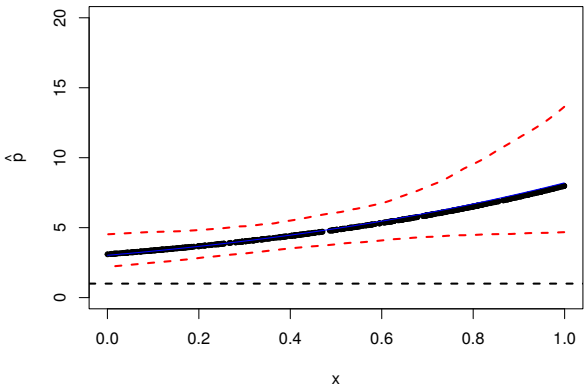
(e) Scenario 3 ($n = 200$)



(f) Scenario 4 ($n = 200$)



(g) Scenario 3 ($n = 500$)



(h) Scenario 4 ($n = 500$)

Source: Elaborated by the author.

C.3.2.2 Using probit link function

Table 115 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 1					
50	β_{10}	0.081	0.324	1.010	45.581
	β_{11}	−0.164	0.839	1.016	74.955
	β_{20}	−0.001	0.156	1.000	62.082
	β_{21}	−0.050	0.471	1.003	109.956
100	β_{10}	0.053	0.197	1.007	35.633
	β_{11}	−0.075	0.485	1.006	55.800
	β_{20}	−0.034	0.092	1.006	47.660
	β_{21}	0.030	0.301	1.002	86.294
200	β_{10}	0.012	0.078	1.001	22.714
	β_{11}	−0.017	0.231	1.001	38.300
	β_{20}	−0.022	0.035	1.007	29.615
	β_{21}	0.020	0.123	1.002	55.809
500	β_{10}	0.016	0.040	1.003	15.990
	β_{11}	−0.028	0.098	1.004	25.124
	β_{20}	0.006	0.016	1.001	19.987
	β_{21}	−0.011	0.047	1.001	35.119
Scenario 2					
50	β_{10}	0.050	0.266	1.005	41.381
	β_{11}	−0.084	0.744	1.005	67.351
	β_{20}	−0.072	0.205	1.013	24.018
	β_{21}	0.057	0.512	1.003	56.514
100	β_{10}	0.012	0.171	1.000	32.812
	β_{11}	−0.022	0.445	1.001	52.762
	β_{20}	−0.014	0.138	1.001	19.169
	β_{21}	0.002	0.380	1.000	48.545
200	β_{10}	0.010	0.072	1.001	21.244
	β_{11}	−0.020	0.214	1.001	37.192
	β_{20}	−0.031	0.043	1.012	11.013
	β_{21}	0.042	0.136	1.007	29.241
500	β_{10}	0.019	0.028	1.007	13.449
	β_{11}	−0.028	0.074	1.005	21.794
	β_{20}	0.002	0.026	1.000	8.697
	β_{21}	−0.004	0.061	1.000	19.548

Source: Elaborated by the author.

Table 116 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 3					
50	β_{10}	0.018	0.371	1.000	47.775
	β_{11}	0.463	1.251	1.099	58.669
	β_{20}	−0.019	0.146	1.001	60.599
	β_{21}	−0.040	0.446	1.002	104.663
100	β_{10}	−0.003	0.287	1.000	43.011
	β_{11}	0.253	0.969	1.035	51.804
	β_{20}	−0.020	0.079	1.002	44.592
	β_{21}	0.017	0.253	1.001	79.059
200	β_{10}	−0.028	0.125	1.003	27.594
	β_{11}	0.159	0.539	1.024	38.997
	β_{20}	−0.003	0.035	1.000	29.779
	β_{21}	0.006	0.136	1.000	56.819
500	β_{10}	−0.005	0.058	1.000	19.370
	β_{11}	0.054	0.223	1.007	25.061
	β_{20}	−0.002	0.014	1.000	19.111
	β_{21}	−0.006	0.046	1.000	34.241
Scenario 4					
50	β_{10}	−0.076	0.372	1.008	48.160
	β_{11}	0.624	1.415	1.174	62.839
	β_{20}	−0.069	0.215	1.011	25.209
	β_{21}	0.034	0.546	1.001	58.479
100	β_{10}	−0.051	0.231	1.006	38.240
	β_{11}	0.352	0.930	1.074	51.961
	β_{20}	0.013	0.119	1.001	17.847
	β_{21}	−0.050	0.320	1.004	44.541
200	β_{10}	0.005	0.111	1.000	25.939
	β_{11}	0.133	0.544	1.017	39.311
	β_{20}	−0.010	0.059	1.001	12.777
	β_{21}	0.001	0.163	1.000	32.126
500	β_{10}	−0.001	0.045	1.000	17.131
	β_{11}	0.079	0.185	1.017	23.264
	β_{20}	−0.011	0.022	1.003	8.025
	β_{21}	0.014	0.054	1.002	18.553

Source: Elaborated by the author.

Table 117 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
50	β_{10}	−0.919	−0.906	0.563	−2.094	0.239
	β_{11}	0.836	0.832	0.901	−0.993	2.668
	β_{20}	0.499	0.495	0.394	−0.267	1.276
	β_{21}	0.450	0.446	0.684	−0.905	1.811
100	β_{10}	−0.947	−0.940	0.440	−1.850	−0.054
	β_{11}	0.925	0.922	0.693	−0.500	2.357
	β_{20}	0.466	0.464	0.302	−0.114	1.047
	β_{21}	0.530	0.528	0.548	−0.501	1.560
200	β_{10}	−0.988	−0.985	0.280	−1.567	−0.412
	β_{11}	0.983	0.982	0.480	0.025	1.945
	β_{20}	0.478	0.477	0.186	0.114	0.844
	β_{21}	0.520	0.518	0.350	−0.176	1.214
500	β_{10}	−0.984	−0.983	0.199	−1.362	−0.608
	β_{11}	0.972	0.972	0.311	0.385	1.560
	β_{20}	0.506	0.506	0.125	0.269	0.744
	β_{21}	0.489	0.489	0.217	0.068	0.909
Scenario 2						
50	β_{10}	−0.950	−0.938	0.514	−2.034	0.121
	β_{11}	0.916	0.914	0.858	−0.855	2.694
	β_{20}	1.428	1.410	0.447	0.496	2.388
	β_{21}	−0.943	−0.931	0.713	−2.445	0.534
100	β_{10}	−0.988	−0.981	0.413	−1.826	−0.160
	β_{11}	0.978	0.976	0.667	−0.406	2.370
	β_{20}	1.486	1.477	0.371	0.771	2.218
	β_{21}	−0.998	−0.991	0.616	−2.162	0.154
200	β_{10}	−0.990	−0.987	0.268	−1.521	−0.462
	β_{11}	0.980	0.981	0.463	0.048	1.915
	β_{20}	1.469	1.465	0.205	1.017	1.924
	β_{21}	−0.958	−0.956	0.366	−1.726	−0.194
500	β_{10}	−0.981	−0.980	0.167	−1.327	−0.636
	β_{11}	0.972	0.972	0.270	0.403	1.540
	β_{20}	1.502	1.500	0.163	1.204	1.803
	β_{21}	−1.004	−1.003	0.246	−1.475	−0.536

Source: Elaborated by the author.

Table 118 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
50	β_{10}	−0.982	−0.960	0.609	−2.438	0.431
	β_{11}	−1.037	−1.023	1.018	−3.661	1.571
	β_{20}	0.481	0.477	0.381	−0.284	1.248
	β_{21}	0.460	0.456	0.666	−0.883	1.816
100	β_{10}	−1.003	−0.989	0.536	−2.138	0.113
	β_{11}	−1.247	−1.240	0.951	−3.347	0.841
	β_{20}	0.480	0.479	0.281	−0.100	1.063
	β_{21}	0.517	0.515	0.503	−0.514	1.551
200	β_{10}	−1.028	−1.022	0.352	−1.750	−0.312
	β_{11}	−1.341	−1.332	0.717	−2.829	0.126
	β_{20}	0.497	0.496	0.187	0.133	0.863
	β_{21}	0.506	0.504	0.368	−0.188	1.203
500	β_{10}	−1.005	−1.002	0.240	−1.475	−0.536
	β_{11}	−1.446	−1.443	0.470	−2.354	−0.540
	β_{20}	0.498	0.498	0.120	0.261	0.736
	β_{21}	0.494	0.494	0.214	0.074	0.913
Scenario 4						
50	β_{10}	−1.076	−1.056	0.605	−2.432	0.241
	β_{11}	−0.876	−0.862	1.013	−3.440	1.667
	β_{20}	1.431	1.414	0.459	0.503	2.389
	β_{21}	−0.966	−0.956	0.738	−2.459	0.509
100	β_{10}	−1.051	−1.039	0.478	−2.089	−0.028
	β_{11}	−1.147	−1.140	0.898	−3.168	0.856
	β_{20}	1.513	1.503	0.345	0.795	2.250
	β_{21}	−1.050	−1.043	0.563	−2.218	0.105
200	β_{10}	−0.995	−0.990	0.332	−1.645	−0.353
	β_{11}	−1.367	−1.356	0.725	−2.797	0.049
	β_{20}	1.490	1.486	0.242	1.034	1.952
	β_{21}	−0.999	−0.997	0.404	−1.771	−0.233
500	β_{10}	−1.001	−0.999	0.212	−1.425	−0.581
	β_{11}	−1.421	−1.418	0.423	−2.297	−0.552
	β_{20}	1.489	1.487	0.150	1.192	1.789
	β_{21}	−0.986	−0.984	0.233	−1.456	−0.518

Source: Elaborated by the author.

Table 119 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	4.60	94.80	0.60	3.40	96.20	0.40
	β_{11}	0.80	96.80	2.40	1.60	95.80	2.60
	β_{20}	1.60	95.80	2.60	0.00	97.00	3.00
	β_{21}	1.40	96.00	2.60	1.60	97.80	0.60
100	β_{10}	3.60	95.20	1.20	2.60	96.40	1.00
	β_{11}	0.80	96.00	3.20	1.60	96.60	1.80
	β_{20}	1.60	95.40	3.00	1.60	94.20	4.20
	β_{21}	2.80	94.40	2.80	3.40	94.60	2.00
200	β_{10}	2.60	96.20	1.20	2.80	95.40	1.80
	β_{11}	2.20	96.00	1.80	1.60	96.20	2.20
	β_{20}	1.80	94.80	3.40	0.80	95.60	3.60
	β_{21}	1.80	95.40	2.80	3.40	95.40	1.20
500	β_{10}	4.60	93.60	1.80	4.00	94.80	1.20
	β_{11}	2.00	94.40	3.60	1.00	96.60	2.40
	β_{20}	1.80	95.40	2.80	3.00	94.00	3.00
	β_{21}	3.20	94.60	2.20	2.60	94.20	3.20

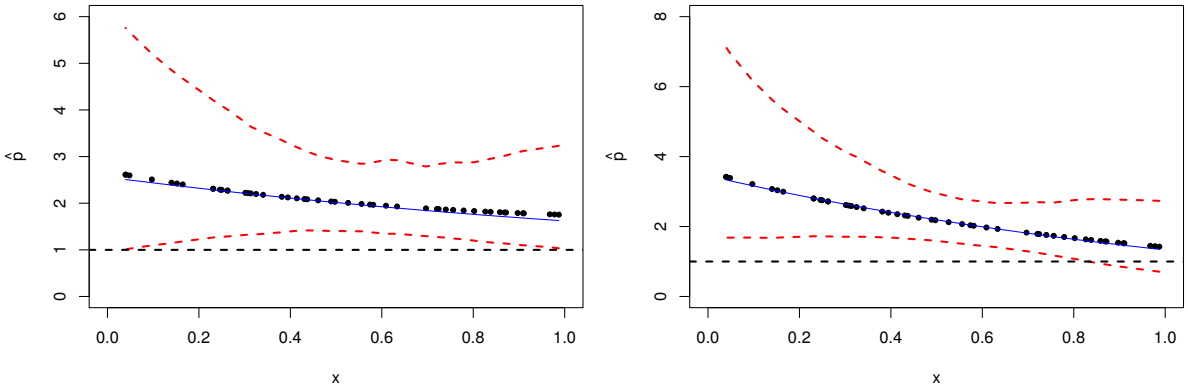
Source: Elaborated by the author.

Table 120 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	β_{10}	1.80	98.00	0.20	1.00	98.00	1.00
	β_{11}	1.80	98.00	0.20	3.80	96.20	0.00
	β_{20}	1.40	95.60	3.00	0.40	96.00	3.60
	β_{21}	1.40	96.00	2.60	3.20	95.60	1.20
100	β_{10}	1.40	98.20	0.40	2.00	97.20	0.80
	β_{11}	2.60	97.00	0.40	2.20	97.20	0.60
	β_{20}	1.80	96.40	1.80	0.60	97.00	2.40
	β_{21}	2.00	95.60	2.40	1.40	96.80	1.80
200	β_{10}	2.60	95.80	1.60	4.00	93.40	2.60
	β_{11}	3.20	96.20	0.60	5.60	93.40	1.00
	β_{20}	2.40	95.40	2.20	2.00	94.60	3.40
	β_{21}	2.40	94.40	3.20	2.60	94.60	2.80
500	β_{10}	2.20	95.60	2.20	3.20	95.60	1.20
	β_{11}	4.00	94.00	2.00	2.40	96.80	0.80
	β_{20}	3.60	94.80	1.60	1.40	94.40	4.20
	β_{21}	2.40	94.40	3.20	2.40	95.60	2.00

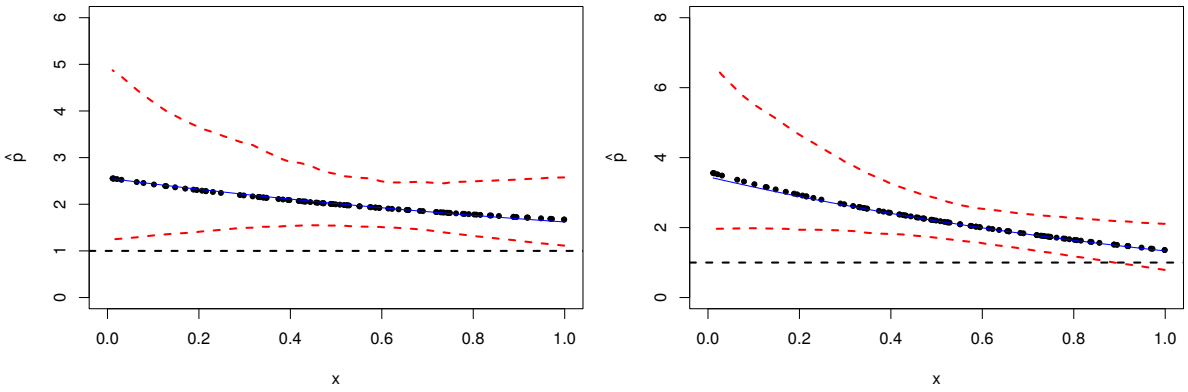
Source: Elaborated by the author.

Figure 45 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 1 and 2).



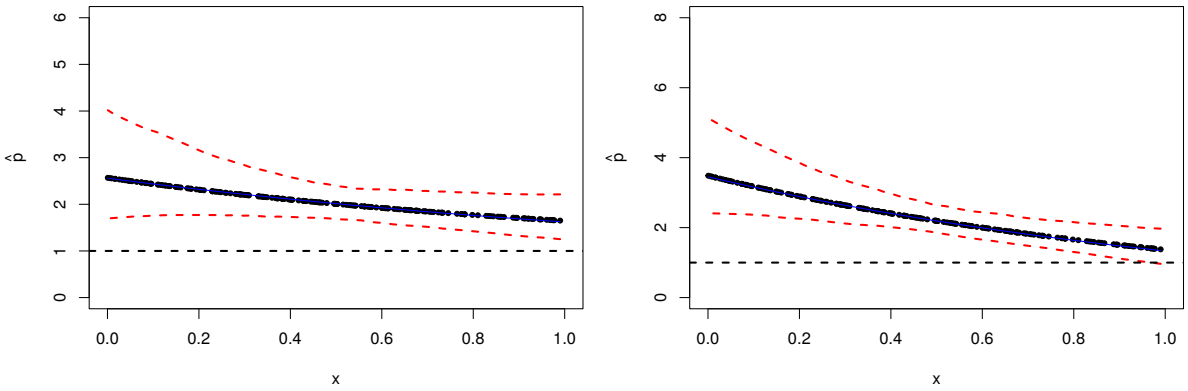
(a) Scenario 1 ($n = 50$)

(b) Scenario 2 ($n = 50$)



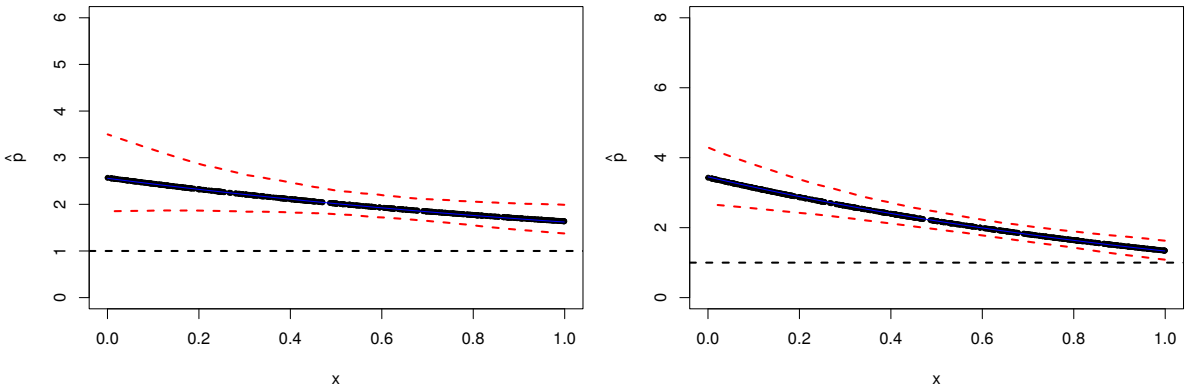
(c) Scenario 1 ($n = 100$)

(d) Scenario 2 ($n = 100$)



(e) Scenario 1 ($n = 200$)

(f) Scenario 2 ($n = 200$)

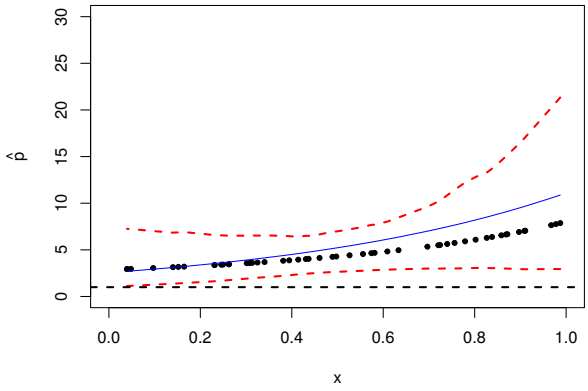


(g) Scenario 1 ($n = 500$)

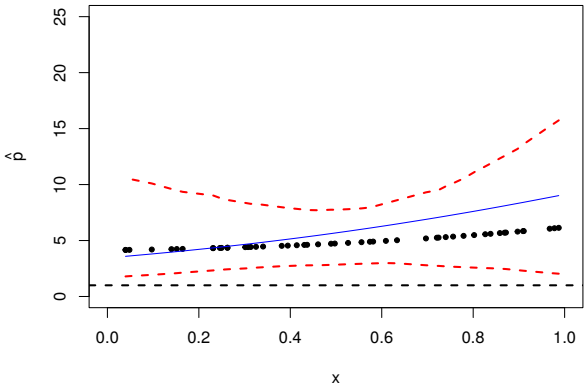
(h) Scenario 2 ($n = 500$)

Source: Elaborated by the author.

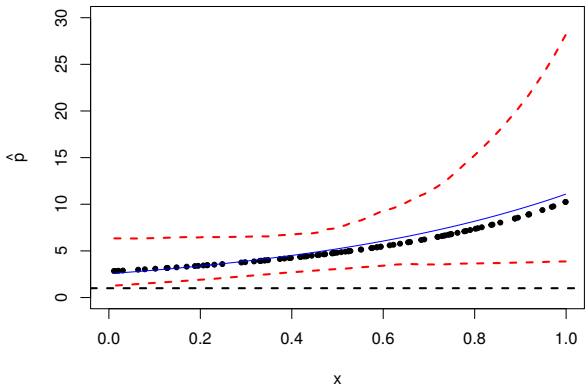
Figure 46 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 3 and 4).



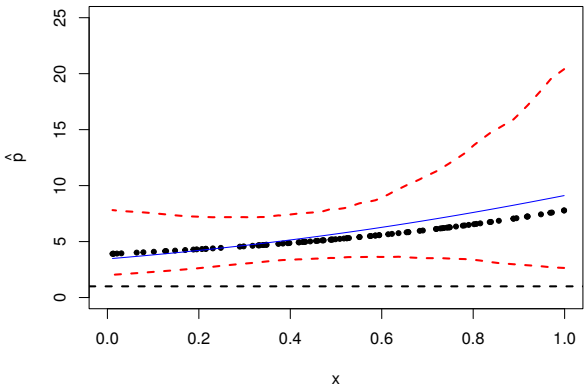
(a) Scenario 3 ($n = 50$)



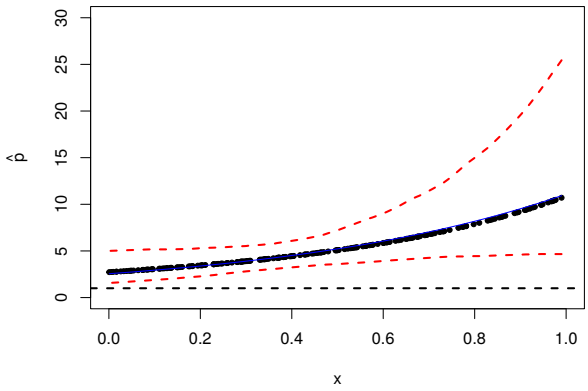
(b) Scenario 4 ($n = 50$)



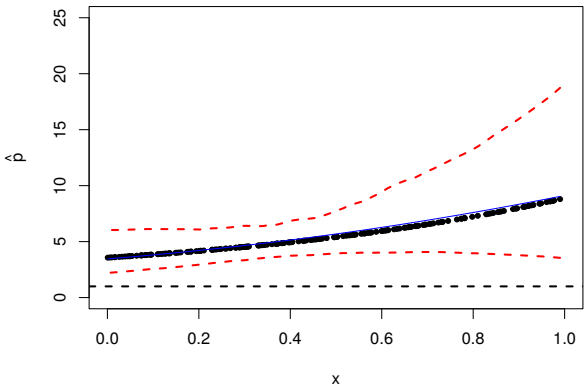
(c) Scenario 3 ($n = 100$)



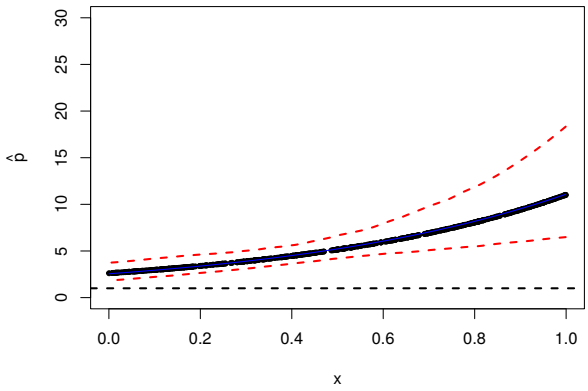
(d) Scenario 4 ($n = 100$)



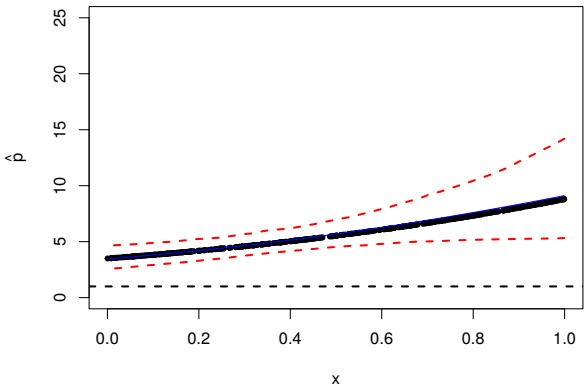
(e) Scenario 3 ($n = 200$)



(f) Scenario 4 ($n = 200$)



(g) Scenario 3 ($n = 500$)



(h) Scenario 4 ($n = 500$)

Source: Elaborated by the author.

C.3.2.3 Using complementary log-log link function

Table 121 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 1					
50	β_{10}	0.029	0.293	1.001	43.753
	β_{11}	−0.053	0.701	1.002	66.609
	β_{20}	0.003	0.145	1.000	58.741
	β_{21}	0.026	0.453	1.001	104.497
100	β_{10}	0.038	0.173	1.004	33.532
	β_{11}	−0.076	0.442	1.007	52.590
	β_{20}	0.008	0.084	1.000	45.671
	β_{21}	−0.018	0.265	1.001	81.344
200	β_{10}	−0.008	0.083	1.000	23.138
	β_{11}	−0.005	0.227	1.000	38.595
	β_{20}	−0.015	0.037	1.003	30.755
	β_{21}	0.024	0.128	1.002	57.261
500	β_{10}	0.016	0.030	1.004	14.056
	β_{11}	−0.024	0.073	1.004	21.857
	β_{20}	0.000	0.013	1.000	18.317
	β_{21}	0.001	0.038	1.000	31.269
Scenario 2					
50	β_{10}	0.051	0.227	1.006	37.842
	β_{11}	−0.087	0.653	1.006	63.138
	β_{20}	0.018	0.241	1.001	26.949
	β_{21}	−0.021	0.604	1.000	62.615
100	β_{10}	0.046	0.148	1.007	30.337
	β_{11}	−0.075	0.421	1.007	51.438
	β_{20}	−0.003	0.141	1.000	19.373
	β_{21}	−0.005	0.344	1.000	45.974
200	β_{10}	0.013	0.065	1.001	20.105
	β_{11}	−0.028	0.188	1.002	34.055
	β_{20}	0.012	0.051	1.001	11.641
	β_{21}	−0.010	0.131	1.000	28.292
500	β_{10}	0.008	0.028	1.001	13.446
	β_{11}	−0.013	0.072	1.001	21.606
	β_{20}	0.003	0.026	1.000	8.472
	β_{21}	−0.012	0.059	1.001	19.093

Source: Elaborated by the author.

Table 122 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
Scenario 3					
50	β_{10}	0.046	0.333	1.003	45.611
	β_{11}	0.370	1.070	1.071	55.202
	β_{20}	−0.008	0.149	1.000	59.638
	β_{21}	0.000	0.478	1.000	105.774
100	β_{10}	−0.013	0.246	1.000	39.508
	β_{11}	0.294	0.834	1.056	48.762
	β_{20}	−0.005	0.082	1.000	44.686
	β_{21}	−0.009	0.230	1.000	75.884
200	β_{10}	0.007	0.100	1.000	25.257
	β_{11}	0.129	0.482	1.018	36.353
	β_{20}	−0.006	0.029	1.001	26.834
	β_{21}	−0.004	0.111	1.000	53.146
500	β_{10}	−0.002	0.051	1.000	18.389
	β_{11}	0.048	0.179	1.006	22.600
	β_{20}	0.003	0.014	1.000	19.103
	β_{21}	−0.009	0.045	1.001	34.072
Scenario 4					
50	β_{10}	−0.038	0.322	1.002	45.114
	β_{11}	0.564	1.249	1.159	58.836
	β_{20}	−0.024	0.225	1.001	25.803
	β_{21}	−0.003	0.524	1.000	58.195
100	β_{10}	0.013	0.228	1.000	38.213
	β_{11}	0.188	0.898	1.020	49.778
	β_{20}	0.010	0.140	1.000	19.337
	β_{21}	−0.042	0.334	1.003	45.964
200	β_{10}	−0.015	0.088	1.001	23.347
	β_{11}	0.163	0.449	1.031	35.552
	β_{20}	−0.008	0.064	1.000	13.124
	β_{21}	0.021	0.169	1.001	32.044
500	β_{10}	0.007	0.041	1.000	16.224
	β_{11}	0.031	0.170	1.003	21.934
	β_{20}	0.011	0.027	1.002	8.345
	β_{21}	−0.013	0.058	1.002	18.718

Source: Elaborated by the author.

Table 123 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
50	β_{10}	−0.971	−0.960	0.540	−2.078	0.113
	β_{11}	0.947	0.943	0.836	−0.760	2.667
	β_{20}	0.503	0.506	0.381	−0.256	1.256
	β_{21}	0.526	0.513	0.672	−0.831	1.907
100	β_{10}	−0.962	−0.956	0.414	−1.805	−0.126
	β_{11}	0.924	0.922	0.661	−0.418	2.270
	β_{20}	0.508	0.511	0.290	−0.051	1.062
	β_{21}	0.482	0.478	0.514	−0.499	1.474
200	β_{10}	−1.008	−1.006	0.287	−1.550	−0.470
	β_{11}	0.995	0.995	0.476	0.095	1.897
	β_{20}	0.485	0.486	0.191	0.139	0.829
	β_{21}	0.524	0.522	0.358	−0.132	1.181
500	β_{10}	−0.984	−0.983	0.174	−1.338	−0.633
	β_{11}	0.976	0.975	0.270	0.424	1.528
	β_{20}	0.500	0.500	0.113	0.277	0.722
	β_{21}	0.501	0.500	0.194	0.110	0.892
Scenario 2						
50	β_{10}	−0.949	−0.939	0.474	−1.991	0.077
	β_{11}	0.913	0.911	0.803	−0.768	2.600
	β_{20}	1.518	1.477	0.491	0.488	2.633
	β_{21}	−1.021	−0.984	0.777	−2.690	0.579
100	β_{10}	−0.954	−0.948	0.382	−1.750	−0.168
	β_{11}	0.925	0.924	0.644	−0.384	2.233
	β_{20}	1.497	1.480	0.376	0.756	2.270
	β_{21}	−1.005	−0.986	0.586	−2.210	0.168
200	β_{10}	−0.987	−0.985	0.254	−1.496	−0.484
	β_{11}	0.972	0.972	0.432	0.090	1.855
	β_{20}	1.512	1.505	0.226	1.046	1.991
	β_{21}	−1.010	−1.002	0.362	−1.792	−0.242
500	β_{10}	−0.992	−0.991	0.169	−1.325	−0.661
	β_{11}	0.987	0.987	0.268	0.447	1.525
	β_{20}	1.503	1.500	0.162	1.206	1.804
	β_{21}	−1.012	−1.009	0.243	−1.478	−0.552

Source: Elaborated by the author.

Table 124 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
50	β_{10}	−0.954	−0.935	0.576	−2.304	0.363
	β_{11}	−1.131	−1.117	0.966	−3.610	1.332
	β_{20}	0.492	0.496	0.386	−0.264	1.243
	β_{21}	0.500	0.489	0.691	−0.845	1.866
100	β_{10}	−1.013	−1.001	0.496	−2.073	0.029
	β_{11}	−1.206	−1.200	0.864	−3.170	0.750
	β_{20}	0.495	0.498	0.286	−0.063	1.047
	β_{21}	0.491	0.488	0.479	−0.488	1.476
200	β_{10}	−0.993	−0.988	0.316	−1.664	−0.333
	β_{11}	−1.371	−1.363	0.682	−2.756	0.002
	β_{20}	0.494	0.495	0.169	0.149	0.836
	β_{21}	0.496	0.494	0.333	−0.155	1.150
500	β_{10}	−1.002	−1.000	0.226	−1.440	−0.568
	β_{11}	−1.452	−1.450	0.421	−2.304	−0.602
	β_{20}	0.503	0.504	0.120	0.280	0.725
	β_{21}	0.491	0.491	0.212	0.101	0.881
Scenario 4						
50	β_{10}	−1.038	−1.020	0.566	−2.329	0.224
	β_{11}	−0.936	−0.921	0.964	−3.372	1.476
	β_{20}	1.476	1.440	0.474	0.477	2.552
	β_{21}	−1.003	−0.967	0.724	−2.620	0.549
100	β_{10}	−0.987	−0.977	0.477	−1.984	−0.011
	β_{11}	−1.312	−1.303	0.929	−3.247	0.616
	β_{20}	1.510	1.493	0.375	0.766	2.287
	β_{21}	−1.042	−1.021	0.576	−2.251	0.134
200	β_{10}	−1.015	−1.011	0.296	−1.643	−0.394
	β_{11}	−1.337	−1.328	0.650	−2.696	0.008
	β_{20}	1.492	1.485	0.253	1.028	1.967
	β_{21}	−0.979	−0.971	0.410	−1.757	−0.211
500	β_{10}	−0.993	−0.992	0.203	−1.403	−0.589
	β_{11}	−1.469	−1.466	0.411	−2.304	−0.635
	β_{20}	1.511	1.508	0.163	1.212	1.813
	β_{21}	−1.013	−1.010	0.241	−1.479	−0.551

Source: Elaborated by the author.

Table 125 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
50	β_{10}	3.40	95.80	0.80	3.40	95.80	0.80
	β_{11}	1.20	96.60	2.20	1.00	96.00	3.00
	β_{20}	1.60	96.80	1.60	0.00	97.40	2.60
	β_{21}	0.80	97.40	1.80	2.00	98.00	0.00
100	β_{10}	3.20	96.20	0.60	3.80	94.60	1.60
	β_{11}	1.20	95.80	3.00	1.80	95.80	2.40
	β_{20}	2.00	96.80	1.20	0.40	96.80	2.80
	β_{21}	2.40	96.00	1.60	2.40	96.80	0.80
200	β_{10}	3.00	93.80	3.20	2.60	95.20	2.20
	β_{11}	2.20	95.00	2.80	2.00	96.00	2.00
	β_{20}	3.20	93.80	3.00	0.60	96.80	2.60
	β_{21}	3.40	94.40	2.20	2.20	97.40	0.40
500	β_{10}	2.20	96.60	1.20	3.40	94.40	2.20
	β_{11}	0.80	96.40	2.80	1.80	94.40	3.80
	β_{20}	2.20	96.00	1.80	2.40	92.80	4.80
	β_{21}	2.20	95.40	2.40	3.00	94.60	2.40

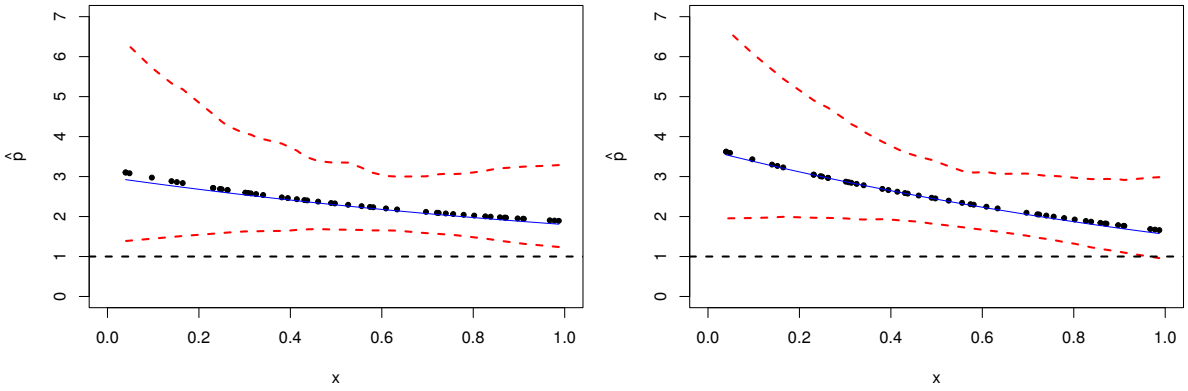
Source: Elaborated by the author.

Table 126 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4).

n	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
50	β_{10}	3.20	96.80	0.00	2.20	97.80	0.00
	β_{11}	1.00	98.60	0.40	3.80	96.20	0.00
	β_{20}	1.00	97.40	1.60	0.00	96.40	3.60
	β_{21}	0.80	97.00	2.20	1.80	98.00	0.20
100	β_{10}	2.60	96.60	0.80	3.00	96.20	0.80
	β_{11}	2.80	96.80	0.40	4.00	95.40	0.60
	β_{20}	2.00	95.80	2.20	0.20	96.80	3.00
	β_{21}	1.60	96.40	2.00	1.40	98.20	0.40
200	β_{10}	2.20	96.80	1.00	2.40	96.20	1.40
	β_{11}	3.40	95.40	1.20	3.80	95.60	0.60
	β_{20}	1.80	95.60	2.60	1.20	94.40	4.40
	β_{21}	2.80	95.00	2.20	4.20	94.00	1.80
500	β_{10}	1.60	96.20	2.20	3.00	95.20	1.80
	β_{11}	3.00	96.00	1.00	3.00	96.00	1.00
	β_{20}	2.40	94.00	3.60	2.80	94.80	2.40
	β_{21}	2.20	95.00	2.80	2.00	95.60	2.40

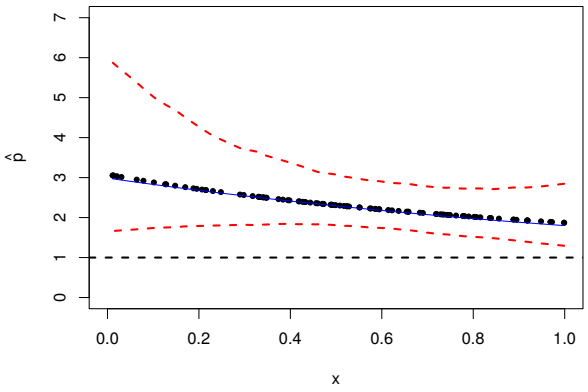
Source: Elaborated by the author.

Figure 47 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 1 and 2).

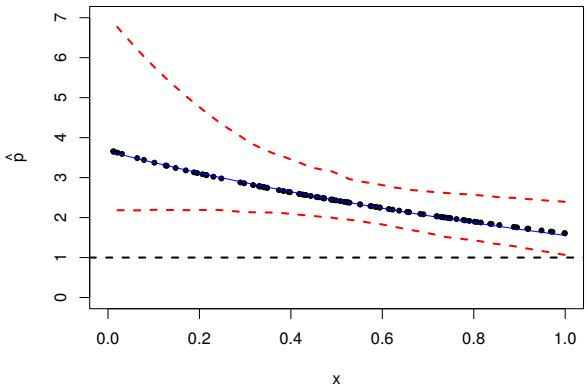


(a) Scenario 1 ($n = 50$)

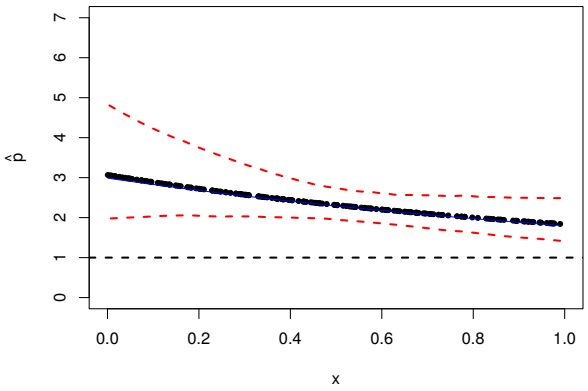
(b) Scenario 2 ($n = 50$)



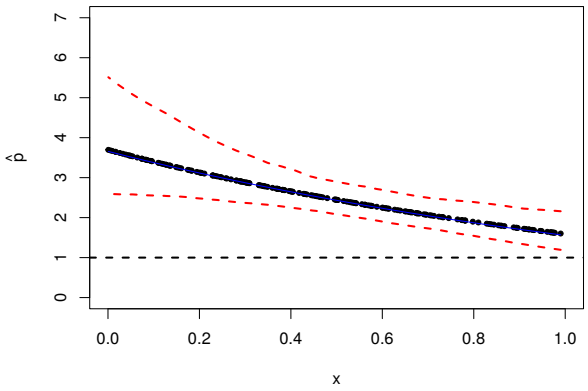
(c) Scenario 1 ($n = 100$)



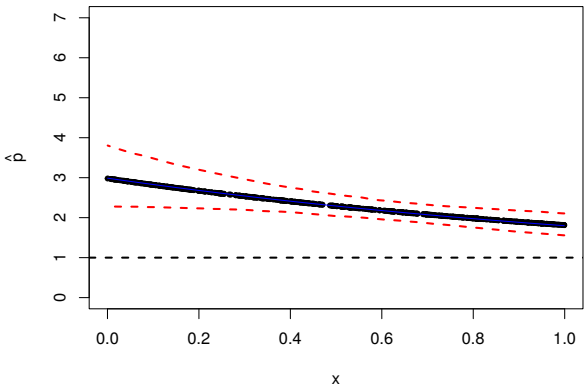
(d) Scenario 2 ($n = 100$)



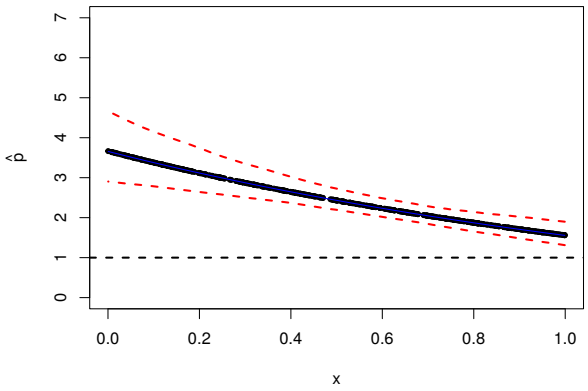
(e) Scenario 1 ($n = 200$)



(f) Scenario 2 ($n = 200$)



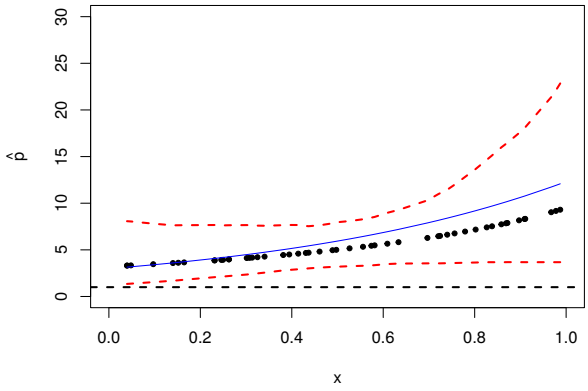
(g) Scenario 1 ($n = 500$)



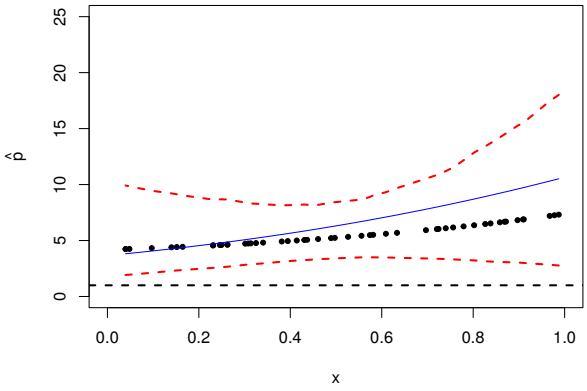
(h) Scenario 2 ($n = 500$)

Source: Elaborated by the author.

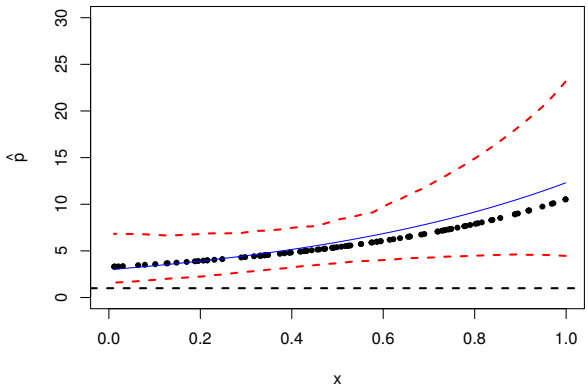
Figure 48 – *Posterior* estimates of parameter p using zero-deflated samples (Scenarios 3 and 4).



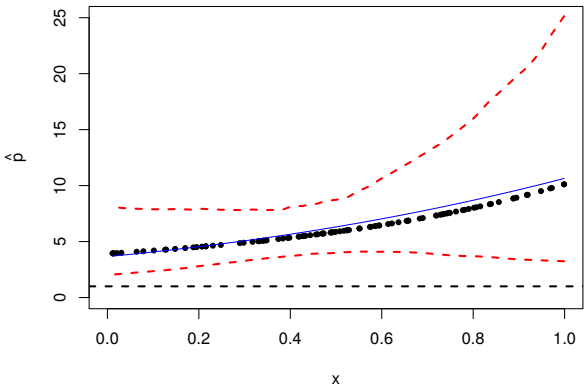
(a) Scenario 3 ($n = 50$)



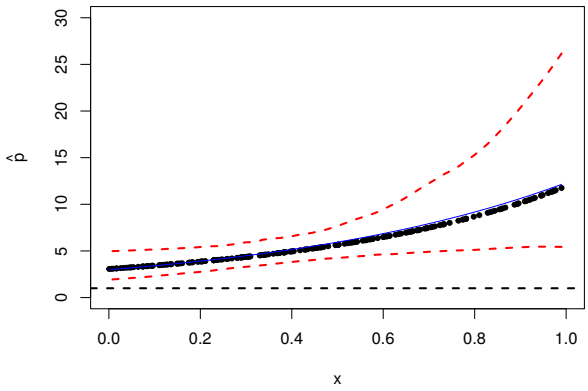
(b) Scenario 4 ($n = 50$)



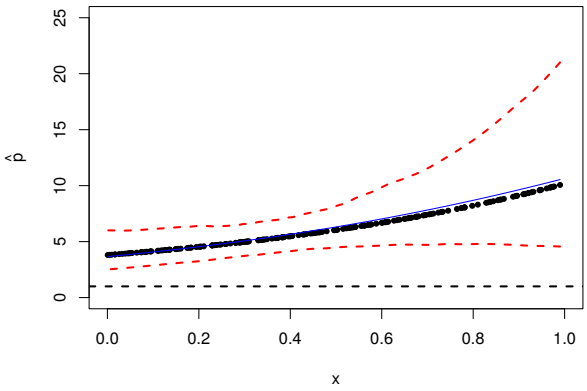
(c) Scenario 3 ($n = 100$)



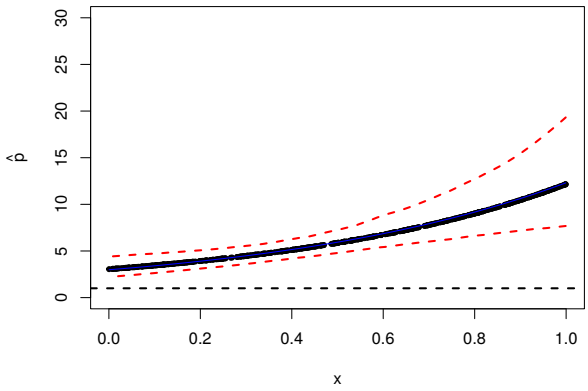
(d) Scenario 4 ($n = 100$)



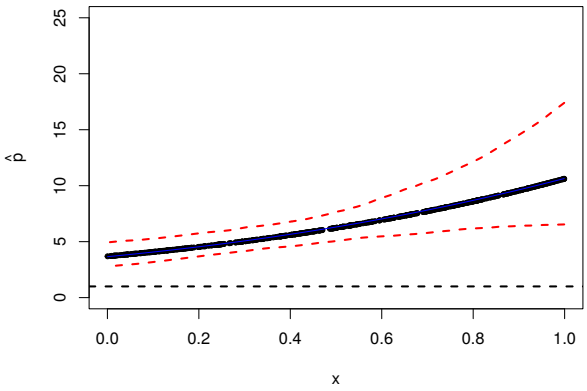
(e) Scenario 3 ($n = 200$)



(f) Scenario 4 ($n = 200$)



(g) Scenario 3 ($n = 500$)



(h) Scenario 4 ($n = 500$)

Source: Elaborated by the author.

APPENDIX

D

SUPPLEMENT FOR CHAPTER 6

D.1 Divergence measures

The acronymous used in Table 127 refers to following divergence measures: Kullback-Leibler (KL), Jeffrey (J), Variational (L^1), Chi-Square (CS), and Hellinger (H).

Table 127 – AMC estimators for some well-known φ-divergence measures and its calibration.

φ	$\hat{d}_{\varphi}^{L}$	$d_{\varphi}(\zeta)$	$\hat{\zeta}_{\varphi}^{L}$
KL_{ij}	$\frac{1}{M}\sum_{t=1}^{M}\log\left(f_{ij}^{t}\right)-\log\left(\widehat{\mathrm{CPO}}_{ij}^{L}\right)$	$-\frac{1}{2}\log\left[4\zeta_{ij}\left(1-\zeta_{ij}\right)\right]$	$\frac{1}{2}\left[1+\sqrt{1-e^{-2\hat{d}_{ij}^{L}}}\right]$
J_{ij}	$\frac{1}{M}\sum_{t=1}^{M}\left(\frac{\widehat{\mathrm{CPO}}_{ij}^{L}}{f_{ij}^{t}}-1\right)\log\left(\frac{\widehat{\mathrm{CPO}}_{ij}^{L}}{f_{ij}^{t}}\right)$	$-\frac{\left(1-2\zeta_{ij}\right)}{2}\log\left(\frac{\zeta_{ij}}{1-\zeta_{ij}}\right)$	no closed-form
L_{ij}^{1}	$\frac{1}{2M}\sum_{t=1}^{M}\frac{\left\lvert\widehat{\mathrm{CPO}}_{ij}^{L}-f_{ij}^{t}\right\rvert}{f_{ij}^{t}}$	$\frac{1}{2}\left\lvert 1-2\zeta_{ij}\right\rvert$	$\frac{1}{2}+\hat{d}_{ij}^{L}$
CS_{ij}	$\frac{1}{M}\sum_{t=1}^{M}\frac{\left(\widehat{\mathrm{CPO}}_{ij}^{L}-f_{ij}^{t}\right)^{2}}{\left(f_{ij}^{t}\right)^{2}}$	$\left(1-2\zeta_{ij}\right)^{2}$	$\frac{1}{2}\left[1+\sqrt{\hat{d}_{ij}^{L}}\right]$
H_{ij}	$\frac{1}{2M}\sum_{t=1}^{M}\left(\sqrt{\frac{\widehat{\mathrm{CPO}}_{ij}^{L}}{f_{ij}^{t}}}-1\right)^{2}$	$1-\frac{1}{\sqrt{2}}\left(\sqrt{\zeta_{ij}}+\sqrt{1-\_ij}\right)$	$\frac{1}{2}+\left[\sqrt{\hat{d}_{ij}^{L}}-\sqrt{\hat{d}_{ij}^{3L}}\right]\sqrt{2-\hat{d}_{ij}^{L}}$

Source: Elaborated by the author.

D.2 Algorithms

D.2.1 Adaptive Metropolis

Algorithm 9 – Adaptive Metropolis

procedure AM($N, \boldsymbol{\theta}_0, \psi$)
  Set $t \leftarrow 1$
  while $t \leqslant N$ **do**
    if $t \leqslant 2d$ **then**
      Generate $\boldsymbol{\theta}' \sim \mathcal{N}_d(\boldsymbol{\theta}_{t-1}, 0.10^2 d^{-1} \mathcal{I}_d)$
    else
      Set $\overline{\boldsymbol{\theta}}_t \leftarrow t^{-1} \sum_{i=0}^{t-1} \boldsymbol{\theta}_i$ and $\mathcal{H}_{t-1} \leftarrow t^{-1} \sum_{j=0}^{t-1} (\boldsymbol{\theta}_j - \overline{\boldsymbol{\theta}}_t)(\boldsymbol{\theta}_j - \overline{\boldsymbol{\theta}}_t)^\intercal$
      Generate $\boldsymbol{\theta}' \sim (1-\psi)\mathcal{N}_d(\boldsymbol{\theta}_{t-1}, 0.10^2 d^{-1} \mathcal{I}_d) + \psi \mathcal{N}_d(\boldsymbol{\theta}_{t-1}, 2.38^2 d^{-1} \mathcal{H}_{t-1})$
    end if
    Set $\alpha' \leftarrow \exp\{\log[\pi(\boldsymbol{\theta}'; \boldsymbol{y})] - \log[\pi(\boldsymbol{\theta}_{t-1}; \boldsymbol{y})]\}$ and $\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1}$
    Generate $u \sim \mathcal{U}(0,1)$
    if $u \leqslant \min\{\alpha', 1\}$ **then**
      Set $\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}'$
    end if
    Set $t \leftarrow t+1$
  end while
  return $\{\boldsymbol{\theta}_t\}_{t=1}^N$
end procedure

D.2.2 Sequential-Search

Algorithm 10 – Sequential-Search

procedure SEQSEA($\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$)
  Set $y \leftarrow 0$
  for $t \leftarrow 1$ to M **do**
    Set $\tilde{\boldsymbol{u}} \leftarrow \arg\min_{\boldsymbol{u} \in \mathbb{R}^2} r(\boldsymbol{u}, y, \boldsymbol{\theta}_t)$ and $\tilde{\mathcal{V}} \leftarrow \frac{\partial^2 r}{\partial \boldsymbol{u} \partial \boldsymbol{u}^\intercal}(\tilde{\boldsymbol{u}}; y, \boldsymbol{\theta}_t)$
    Set $s_t \leftarrow (\det \tilde{\mathcal{V}})^{-1/2} \exp\{-r(\tilde{\boldsymbol{u}}, y, \boldsymbol{\theta}_t)\}$
  end for
  Set $k \leftarrow M^{-1} \sum_{t=1}^M s_t$
  Generate $u \sim \mathcal{U}(0,1)$
  while $u > k$ **do**
    Set $y \leftarrow y+1$
    Repeat lines 3-6
    Set $k \leftarrow k + M^{-1} \sum_{t=1}^M s_t$
  end while
  return y
end procedure

D.2.3 Inverse transform sampling

Algorithm 11 – Inverse transform sampling

procedure ITS($\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_1, \sigma_2, \rho$)
  Generate $u_{1i}, u_{2i} \sim \mathcal{N}(0,1)$
  Set $b_{1i} \leftarrow u_{1i}\sigma_1$ and $b_{2i} \leftarrow \sigma_2(u_{1i}\rho + u_{2i}\sqrt{1-\rho^2})$
  for $j \leftarrow 1$ to m_i **do**
    Generate $u \sim \mathcal{U}(0,1)$
    Set $\mu_{ij} \leftarrow \exp\{\beta_{10} + \beta_{11}x_{1ij} + b_{1i}\}$ and $\omega_{ij} \leftarrow g_2^{-1}(\beta_{20} + \beta_{21}x_{2ij} + b_{2i})$
    Set $\theta_{ij} \leftarrow -(2\mu_{ij})^{-1}[(\mu_{ij}-1) - \sqrt{(\mu_{ij}-1)^2 + 8\mu_{ij}}]$ and $y_{ij} \leftarrow Q(u, \theta_{ij})$
    Generate $z \sim \mathcal{B}(1, 1-\omega_{ij})$ (Binomial with size 1 and prob. of success $1-\omega_{ij}$)
    if $z = 1$ **then**
      Set $y_{ij} \leftarrow 0$
    end if
  end for
  return $(y_{i1}, \ldots, y_{im_i})$
end procedure

D.3 Results from the simulation study

D.3.1 Zero-inflated artificial data

Table 128 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 1 and 2).

n (m)	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 1		
25 (100)	β_{10}	0.038	0.119	1.006	18.134
	β_{11}	−0.294	0.471	1.107	21.966
	β_{20}	−0.506	0.797	1.214	71.284
	β_{21}	0.638	2.044	1.118	74.180
	σ_1	−0.003	0.037	1.000	31.694
	σ_2	0.054	0.426	1.003	33.356
	ρ	0.414	0.265	1.685	86.390
50 (200)	β_{10}	0.092	0.067	1.069	13.825
	β_{11}	−0.263	0.238	1.187	15.680
	β_{20}	−0.506	0.508	1.420	59.901
	β_{21}	0.637	1.211	1.226	60.016
	σ_1	−0.031	0.021	1.024	23.006
	σ_2	−0.130	0.164	1.056	21.782
	ρ	0.277	0.175	1.332	65.269
100 (400)	β_{10}	0.144	0.047	1.339	11.874
	β_{11}	−0.261	0.146	1.370	12.680
	β_{20}	−0.503	0.360	1.831	52.528
	β_{21}	0.638	0.782	1.444	48.921
	σ_1	−0.051	0.013	1.119	17.590
	σ_2	−0.156	0.090	1.172	16.302
	ρ	0.173	0.103	1.188	50.726
			Scenario 2		
25 (100)	β_{10}	0.007	0.133	1.000	18.926
	β_{11}	−0.224	0.433	1.063	20.608
	β_{20}	−0.462	0.683	1.206	68.074
	β_{21}	0.553	1.768	1.099	71.485
	σ_1	0.055	0.049	1.033	35.431
	σ_2	0.086	0.386	1.010	32.294
	ρ	0.131	0.112	1.087	-
50 (200)	β_{10}	0.048	0.064	1.018	13.579
	β_{11}	−0.176	0.203	1.087	14.467
	β_{20}	−0.419	0.417	1.315	52.498
	β_{21}	0.526	1.083	1.158	55.379
	σ_1	0.013	0.022	1.004	23.798
	σ_2	−0.055	0.157	1.010	20.980
	ρ	0.116	0.133	1.055	-
100 (400)	β_{10}	0.064	0.033	1.070	9.473
	β_{11}	−0.156	0.105	1.140	10.491
	β_{20}	−0.425	0.310	1.550	46.167
	β_{21}	0.545	0.714	1.308	45.466
	σ_1	0.001	0.008	1.000	14.707
	σ_2	−0.123	0.075	1.120	14.731
	ρ	0.106	0.093	1.067	-

Source: Elaborated by the author.

Table 129 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 1 and 2).

n (m)	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
25 (100)	β_{10}	1.538	1.545	0.343	0.817	2.249
	β_{11}	2.206	2.207	0.620	0.982	3.429
	β_{20}	0.494	0.487	0.735	−1.044	2.047
	β_{21}	−0.862	−0.849	1.279	−3.564	1.814
	σ_1	0.497	0.476	0.192	0.111	0.912
	σ_2	1.554	1.489	0.651	0.492	2.715
	ρ	−0.086	−0.100	0.306	−0.857	0.724
50 (200)	β_{10}	1.592	1.594	0.243	1.110	2.068
	β_{11}	2.237	2.236	0.411	1.421	3.061
	β_{20}	0.494	0.488	0.502	−0.520	1.523
	β_{21}	−0.863	−0.853	0.897	−2.663	0.920
	σ_1	0.469	0.464	0.142	0.195	0.744
	σ_2	1.370	1.347	0.383	0.690	2.088
	ρ	−0.223	−0.249	0.314	−0.859	0.472
100 (400)	β_{10}	1.644	1.645	0.162	1.319	1.970
	β_{11}	2.239	2.238	0.279	1.689	2.789
	β_{20}	0.497	0.494	0.328	−0.208	1.208
	β_{21}	−0.862	−0.856	0.612	−2.098	0.366
	σ_1	0.449	0.448	0.101	0.264	0.635
	σ_2	1.344	1.332	0.256	0.889	1.820
	ρ	−0.327	−0.345	0.270	−0.835	0.214
Scenario 2						
25 (100)	β_{10}	1.507	1.518	0.364	0.742	2.254
	β_{11}	2.276	2.279	0.618	0.984	3.568
	β_{20}	0.538	0.531	0.686	−1.014	2.115
	β_{21}	−0.947	−0.933	1.209	−3.689	1.773
	σ_1	0.555	0.532	0.214	0.147	0.998
	σ_2	1.586	1.520	0.616	0.518	2.767
	ρ	0.131	0.161	0.308	−0.679	0.860
50 (200)	β_{10}	1.548	1.552	0.248	1.046	2.042
	β_{11}	2.324	2.324	0.415	1.480	3.170
	β_{20}	0.581	0.574	0.491	−0.474	1.644
	β_{21}	−0.974	−0.962	0.898	−2.831	0.856
	σ_1	0.513	0.506	0.148	0.238	0.795
	σ_2	1.445	1.419	0.392	0.746	2.189
	ρ	0.116	0.130	0.346	−0.549	0.747
100 (400)	β_{10}	1.564	1.566	0.169	1.221	1.905
	β_{11}	2.344	2.343	0.285	1.773	2.917
	β_{20}	0.575	0.571	0.359	−0.140	1.299
	β_{21}	−0.955	−0.949	0.646	−2.207	0.290
	σ_1	0.501	0.498	0.092	0.320	0.687
	σ_2	1.377	1.365	0.244	0.917	1.859
	ρ	0.106	0.113	0.285	−0.424	0.622

Source: Elaborated by the author.

Table 130 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 1 and 2).

n (m)	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
25 (100)	β_{10}	3.80	95.00	1.20	2.60	96.00	1.40
	β_{11}	0.40	91.20	8.40	1.40	94.60	4.00
	β_{20}	0.40	88.00	11.60	0.00	92.20	7.80
	β_{21}	9.00	90.60	0.40	5.60	94.00	0.40
	σ_1	0.40	95.40	4.20	1.80	96.20	2.00
	σ_2	1.20	89.80	9.00	0.80	93.40	5.80
	ρ	8.00	91.80	0.20	4.80	94.60	0.60
50 (200)	β_{10}	7.00	92.00	1.00	5.20	94.00	0.80
	β_{11}	0.80	88.80	10.40	0.60	93.80	5.60
	β_{20}	0.00	81.80	18.20	0.40	88.60	11.00
	β_{21}	10.80	88.80	0.40	8.00	91.60	0.40
	σ_1	1.20	92.00	6.80	2.80	92.40	4.80
	σ_2	0.00	90.80	9.20	1.00	93.00	6.00
	ρ	8.20	91.80	0.00	9.60	88.00	2.40
100 (400)	β_{10}	14.60	85.40	0.00	7.00	91.80	1.20
	β_{11}	0.20	85.20	14.60	0.80	91.20	8.00
	β_{20}	0.00	70.00	30.00	0.00	76.80	23.20
	β_{21}	18.80	81.20	0.00	16.20	83.60	0.20
	σ_1	0.80	91.80	7.40	1.40	95.80	2.80
	σ_2	0.20	87.20	12.60	0.20	90.20	9.60
	ρ	6.60	92.40	1.00	9.40	88.80	1.80

Source: Elaborated by the author.

Table 131 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 3 and 4).

$n\ (m)$	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 3		
25 (100)	β_{10}	−0.062	0.151	1.013	20.274
	β_{11}	−0.079	0.407	1.008	20.320
	β_{20}	−0.272	0.540	1.076	57.662
	β_{21}	0.298	1.571	1.030	64.685
	σ_1	0.056	0.047	1.035	34.731
	σ_2	0.103	0.384	1.014	31.854
	ρ	−0.157	0.098	1.155	49.002
50 (200)	β_{10}	−0.017	0.061	1.002	12.960
	β_{11}	−0.067	0.188	1.012	13.707
	β_{20}	−0.258	0.343	1.113	47.729
	β_{21}	0.289	0.907	1.049	50.749
	σ_1	0.020	0.022	1.009	23.857
	σ_2	−0.006	0.143	1.000	20.179
	ρ	−0.069	0.077	1.033	43.322
100 (400)	β_{10}	−0.005	0.028	1.000	8.900
	β_{11}	−0.046	0.077	1.014	8.749
	β_{20}	−0.250	0.182	1.234	34.261
	β_{21}	0.302	0.454	1.119	36.014
	σ_1	0.007	0.009	1.003	15.238
	σ_2	−0.073	0.076	1.037	14.617
	ρ	0.014	0.050	1.002	36.408
			Scenario 4		
25 (100)	β_{10}	0.494	0.493	1.409	38.835
	β_{11}	−1.245	2.345	1.718	52.835
	β_{20}	−0.142	0.231	1.047	38.791
	β_{21}	0.191	0.717	1.026	45.051
	σ_1	−0.237	0.106	1.461	17.933
	σ_2	0.114	0.088	1.083	41.333
	ρ	0.372	0.218	1.655	78.357
50 (200)	β_{10}	0.541	0.432	1.756	37.851
	β_{11}	−1.021	1.550	1.748	43.070
	β_{20}	−0.118	0.145	1.052	30.322
	β_{21}	0.155	0.400	1.032	34.083
	σ_1	−0.212	0.073	1.609	15.012
	σ_2	−0.013	0.040	1.002	31.920
	ρ	0.284	0.157	1.433	64.260
100 (400)	β_{10}	0.610	0.455	2.341	40.968
	β_{11}	−0.959	1.221	2.015	39.224
	β_{20}	−0.158	0.080	1.207	23.217
	β_{21}	0.194	0.209	1.104	24.330
	σ_1	−0.218	0.063	2.028	14.847
	σ_2	−0.089	0.031	1.161	30.329
	ρ	0.225	0.112	1.352	54.081

Source: Elaborated by the author.

Table 132 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 3 and 4).

n (m)	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
25 (100)	β_{10}	1.438	1.452	0.384	0.689	2.161
	β_{11}	2.421	2.424	0.633	1.150	3.689
	β_{20}	0.728	0.719	0.683	−0.817	2.306
	β_{21}	−1.202	−1.184	1.218	−3.980	1.521
	σ_1	0.556	0.533	0.208	0.150	0.996
	σ_2	1.603	1.536	0.611	0.536	2.781
	ρ	0.343	0.408	0.271	−0.455	0.956
50 (200)	β_{10}	1.483	1.489	0.246	0.979	1.975
	β_{11}	2.433	2.433	0.429	1.586	3.276
	β_{20}	0.742	0.734	0.526	−0.342	1.845
	β_{21}	−1.211	−1.195	0.908	−3.136	0.673
	σ_1	0.520	0.512	0.147	0.242	0.805
	σ_2	1.494	1.466	0.378	0.796	2.243
	ρ	0.431	0.474	0.268	−0.187	0.939
100 (400)	β_{10}	1.495	1.498	0.167	1.155	1.832
	β_{11}	2.454	2.454	0.273	1.882	3.025
	β_{20}	0.750	0.745	0.346	0.009	1.501
	β_{21}	−1.198	−1.189	0.602	−2.495	0.089
	σ_1	0.507	0.504	0.096	0.324	0.696
	σ_2	1.427	1.414	0.265	0.961	1.917
	ρ	0.514	0.540	0.224	0.064	0.909
Scenario 4						
25 (100)	β_{10}	1.994	2.006	0.498	0.864	3.108
	β_{11}	1.255	1.261	0.891	−0.775	3.266
	β_{20}	0.858	0.846	0.459	−0.155	1.899
	β_{21}	−1.309	−1.294	0.825	−3.139	0.484
	σ_1	1.263	1.229	0.222	0.765	1.828
	σ_2	0.614	0.570	0.274	0.027	1.283
	ρ	−0.128	−0.153	0.282	−0.888	0.698
50 (200)	β_{10}	2.041	2.045	0.374	1.233	2.844
	β_{11}	1.479	1.481	0.712	0.035	2.918
	β_{20}	0.882	0.874	0.362	0.204	1.571
	β_{21}	−1.345	−1.335	0.613	−2.538	−0.164
	σ_1	1.288	1.273	0.168	0.951	1.652
	σ_2	0.487	0.467	0.200	0.038	0.959
	ρ	−0.216	−0.249	0.276	−0.905	0.549
100 (400)	β_{10}	2.110	2.111	0.288	1.542	2.678
	β_{11}	1.541	1.540	0.548	0.526	2.557
	β_{20}	0.842	0.838	0.234	0.384	1.307
	β_{21}	−1.306	−1.301	0.414	−2.106	−0.513
	σ_1	1.282	1.275	0.124	1.049	1.526
	σ_2	0.411	0.403	0.151	0.049	0.771
	ρ	−0.275	−0.302	0.247	−0.916	0.409

Source: Elaborated by the author.

Table 133 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 3 and 4).

$n\ (m)$	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
25 (100)	β_{10}	2.00	94.80	3.20	11.00	89.00	0.00
	β_{11}	2.00	94.20	3.80	0.20	78.20	21.60
	β_{20}	0.20	94.60	5.20	0.00	97.40	2.60
	β_{21}	4.40	95.00	0.60	2.00	98.00	0.00
	σ_1	2.20	95.20	2.60	0.00	87.40	12.60
	σ_2	1.00	91.40	7.60	0.40	99.60	0.00
	ρ	0.40	97.80	1.80	5.00	95.00	0.00
50 (200)	β_{10}	2.20	95.20	2.60	25.80	74.20	0.00
	β_{11}	2.00	95.20	2.80	0.00	72.00	28.00
	β_{20}	0.40	92.80	6.80	1.00	92.20	6.80
	β_{21}	4.80	94.40	0.80	5.60	93.60	0.80
	σ_1	2.00	93.60	4.40	0.00	75.00	25.00
	σ_2	0.60	94.80	4.60	1.00	98.80	0.20
	ρ	2.60	96.60	0.80	5.00	94.80	0.20
100 (400)	β_{10}	2.00	95.40	2.60	56.40	43.60	0.00
	β_{11}	1.40	96.00	2.60	0.00	54.00	46.00
	β_{20}	0.20	91.20	8.60	0.60	89.20	10.20
	β_{21}	6.00	93.80	0.20	8.60	90.40	1.00
	σ_1	2.20	95.20	2.60	0.00	56.60	43.40
	σ_2	1.00	91.60	7.40	0.00	95.00	5.00
	ρ	8.40	89.60	2.00	3.40	96.00	0.60

Source: Elaborated by the author.

Table 134 – Empirical properties of the Bayesian estimators using zero-inflated samples (Scenarios 5 and 6).

n (m)	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 5		
25 (100)	β_{10}	0.386	0.422	1.242	34.955
	β_{11}	−1.111	2.159	1.528	48.528
	β_{20}	−0.007	0.233	1.000	36.819
	β_{21}	−0.054	0.743	1.002	44.128
	σ_1	−0.221	0.099	1.406	17.143
	σ_2	0.122	0.095	1.088	42.353
	ρ	0.043	0.081	1.012	-
50 (200)	β_{10}	0.462	0.360	1.569	33.993
	β_{11}	−0.938	1.462	1.585	41.095
	β_{20}	−0.040	0.121	1.007	27.343
	β_{21}	0.003	0.366	1.000	31.195
	σ_1	−0.177	0.066	1.383	13.952
	σ_2	−0.001	0.049	1.000	35.196
	ρ	0.050	0.084	1.015	-
100 (400)	β_{10}	0.501	0.334	2.008	34.106
	β_{11}	−0.815	0.942	1.839	33.776
	β_{20}	−0.065	0.062	1.036	19.671
	β_{21}	0.080	0.174	1.019	21.928
	σ_1	−0.170	0.045	1.653	11.993
	σ_2	−0.096	0.032	1.187	30.588
	ρ	0.063	0.082	1.025	-
			Scenario 6		
25 (100)	β_{10}	0.284	0.376	1.129	32.825
	β_{11}	−0.973	2.016	1.374	46.015
	β_{20}	0.052	0.215	1.006	36.386
	β_{21}	−0.168	0.698	1.021	44.176
	σ_1	−0.214	0.095	1.386	16.550
	σ_2	0.106	0.078	1.080	39.536
	ρ	−0.321	0.182	1.518	69.562
50 (200)	β_{10}	0.380	0.305	1.380	30.991
	β_{11}	−0.822	1.247	1.478	37.626
	β_{20}	0.027	0.128	1.003	28.176
	β_{21}	−0.094	0.383	1.012	32.757
	σ_1	−0.159	0.059	1.319	13.304
	σ_2	0.008	0.045	1.001	33.477
	ρ	−0.191	0.114	1.212	53.417
100 (400)	β_{10}	0.406	0.245	1.748	28.089
	β_{11}	−0.705	0.784	1.651	29.971
	β_{20}	0.020	0.053	1.004	17.974
	β_{21}	−0.043	0.160	1.006	21.425
	σ_1	−0.139	0.036	1.454	10.629
	σ_2	−0.047	0.029	1.040	28.299
	ρ	−0.063	0.055	1.038	35.809

Source: Elaborated by the author.

Table 135 – *Posterior* estimates of model parameters using zero-inflated samples (Scenarios 5 and 6).

n (m)	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 5						
25 (100)	β_{10}	1.886	1.898	0.523	0.737	3.006
	β_{11}	1.389	1.397	0.961	−0.654	3.421
	β_{20}	0.993	0.981	0.482	−0.039	2.048
	β_{21}	−1.554	−1.536	0.860	−3.397	0.252
	σ_1	1.279	1.243	0.224	0.770	1.855
	σ_2	0.622	0.577	0.283	0.026	1.306
	ρ	0.043	0.053	0.282	−0.775	0.834
50 (200)	β_{10}	1.962	1.967	0.382	1.137	2.785
	β_{11}	1.562	1.563	0.763	0.083	3.042
	β_{20}	0.960	0.951	0.346	0.274	1.659
	β_{21}	−1.497	−1.486	0.605	−2.708	−0.300
	σ_1	1.323	1.307	0.186	0.974	1.699
	σ_2	0.499	0.478	0.221	0.040	0.984
	ρ	0.050	0.058	0.286	−0.710	0.796
100 (400)	β_{10}	2.001	2.002	0.288	1.414	2.582
	β_{11}	1.685	1.685	0.528	0.641	2.725
	β_{20}	0.935	0.932	0.240	0.472	1.402
	β_{21}	−1.420	−1.414	0.409	−2.219	−0.622
	σ_1	1.330	1.323	0.129	1.089	1.585
	σ_2	0.404	0.394	0.151	0.036	0.772
	ρ	0.063	0.070	0.280	−0.648	0.765
Scenario 6						
25 (100)	β_{10}	1.784	1.798	0.544	0.640	2.909
	β_{11}	1.527	1.536	1.034	−0.538	3.580
	β_{20}	1.052	1.040	0.460	0.032	2.095
	β_{21}	−1.668	−1.649	0.818	−3.503	0.135
	σ_1	1.286	1.251	0.223	0.775	1.874
	σ_2	0.606	0.562	0.259	0.022	1.274
	ρ	0.179	0.218	0.281	−0.658	0.908
50 (200)	β_{10}	1.880	1.885	0.400	1.047	2.704
	β_{11}	1.678	1.681	0.756	0.187	3.164
	β_{20}	1.027	1.020	0.356	0.338	1.727
	β_{21}	−1.594	−1.583	0.612	−2.812	−0.400
	σ_1	1.341	1.325	0.184	0.985	1.724
	σ_2	0.508	0.489	0.211	0.054	0.982
	ρ	0.309	0.351	0.278	−0.430	0.934
100 (400)	β_{10}	1.906	1.908	0.283	1.314	2.499
	β_{11}	1.795	1.797	0.536	0.734	2.853
	β_{20}	1.020	1.016	0.229	0.549	1.498
	β_{21}	−1.543	−1.536	0.397	−2.365	−0.731
	σ_1	1.361	1.354	0.131	1.113	1.623
	σ_2	0.453	0.447	0.165	0.091	0.809
	ρ	0.437	0.472	0.225	−0.159	0.961

Source: Elaborated by the author.

Table 136 – Coverage probabilities (%) of the HPDIs using zero-inflated samples (Scenarios 5 and 6).

n (m)	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 5			Scenario 6		
25 (100)	β_{10}	10.60	88.60	0.80	7.60	92.20	0.20
	β_{11}	0.00	82.00	18.00	0.20	84.60	15.20
	β_{20}	0.60	96.80	2.60	1.60	97.00	1.40
	β_{21}	2.00	96.40	1.60	0.60	98.00	1.40
	σ_1	0.00	88.00	12.00	0.00	89.60	10.40
	σ_2	1.00	99.00	0.00	0.60	99.40	0.00
	ρ	1.60	97.60	0.80	0.00	97.00	3.00
50 (200)	β_{10}	18.80	81.20	0.00	15.40	84.20	0.40
	β_{11}	0.00	75.60	24.40	0.60	80.00	19.40
	β_{20}	1.20	95.20	3.60	2.20	94.60	3.20
	β_{21}	3.00	95.00	2.00	2.40	95.80	1.80
	σ_1	0.00	81.80	18.20	0.00	84.00	16.00
	σ_2	0.80	98.40	0.80	1.20	97.80	1.00
	ρ	2.60	95.60	1.80	1.20	96.20	2.60
100 (400)	β_{10}	39.60	60.40	0.00	25.20	74.80	0.00
	β_{11}	0.00	65.20	34.80	0.00	71.80	28.20
	β_{20}	1.00	93.60	5.40	2.20	95.80	2.00
	β_{21}	4.40	94.40	1.20	1.80	96.60	1.60
	σ_1	0.00	70.80	29.20	0.00	79.60	20.40
	σ_2	0.40	94.80	4.80	1.20	95.20	3.60
	ρ	4.80	93.80	1.40	2.40	96.60	1.00

Source: Elaborated by the author.

D.3.2 Zero-deflated artificial data

Table 137 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 1 and 2).

n (m)	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 1		
25 (100)	β_{10}	−0.023	0.374	1.001	24.126
	β_{11}	0.567	1.457	1.133	95.998
	β_{20}	−0.067	0.493	1.005	37.208
	β_{21}	0.306	1.680	1.029	51.540
	σ_1	0.336	0.198	1.527	67.758
	σ_2	−0.070	0.327	1.008	30.726
	ρ	0.309	0.132	1.901	64.178
50 (200)	β_{10}	−0.141	0.291	1.036	20.653
	β_{11}	0.328	0.966	1.061	77.834
	β_{20}	−0.044	0.288	1.003	28.836
	β_{21}	0.293	0.895	1.052	38.060
	σ_1	0.155	0.081	1.192	36.760
	σ_2	−0.117	0.141	1.052	19.884
	ρ	0.306	0.138	1.759	64.877
100 (400)	β_{10}	−0.121	0.204	1.038	16.980
	β_{11}	0.164	0.689	1.020	64.828
	β_{20}	0.012	0.137	1.000	19.426
	β_{21}	0.168	0.446	1.034	26.568
	σ_1	0.084	0.059	1.066	33.881
	σ_2	−0.188	0.093	1.268	16.711
	ρ	0.254	0.126	1.429	57.939
			Scenario 2		
25 (100)	β_{10}	0.046	0.346	1.003	23.268
	β_{11}	0.583	1.508	1.136	95.943
	β_{20}	−0.129	0.557	1.015	38.141
	β_{21}	0.300	1.610	1.029	48.802
	σ_1	0.287	0.156	1.456	57.966
	σ_2	0.051	0.389	1.003	32.895
	ρ	−0.133	0.051	1.238	-
50 (200)	β_{10}	−0.070	0.255	1.010	20.229
	β_{11}	0.383	0.984	1.084	79.824
	β_{20}	−0.033	0.327	1.002	30.002
	β_{21}	0.162	0.952	1.014	38.670
	σ_1	0.117	0.059	1.141	31.768
	σ_2	−0.051	0.158	1.008	21.056
	ρ	−0.112	0.059	1.127	-
100 (400)	β_{10}	−0.032	0.164	1.003	16.284
	β_{11}	0.151	0.577	1.020	60.887
	β_{20}	−0.004	0.153	1.000	20.614
	β_{21}	0.139	0.454	1.022	26.508
	σ_1	0.041	0.042	1.021	28.506
	σ_2	−0.142	0.084	1.145	15.862
	ρ	−0.088	0.078	1.054	-

Source: Elaborated by the author.

Table 138 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 1 and 2).

n (m)	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 1						
25 (100)	β_{10}	−2.023	−1.985	0.611	−3.732	−0.404
	β_{11}	−0.433	−0.431	1.066	−3.523	2.659
	β_{20}	1.433	1.413	0.699	−0.077	2.994
	β_{21}	−1.694	−1.669	1.260	−4.348	0.895
	σ_1	0.836	0.677	0.292	0.002	2.113
	σ_2	1.430	1.369	0.567	0.395	2.555
	ρ	−0.191	−0.263	0.191	−0.977	0.787
50 (200)	β_{10}	−2.141	−2.108	0.521	−3.394	−0.946
	β_{11}	−0.672	−0.665	0.926	−3.001	1.632
	β_{20}	1.456	1.438	0.535	0.389	2.559
	β_{21}	−1.707	−1.685	0.899	−3.574	0.112
	σ_1	0.655	0.574	0.239	0.007	1.522
	σ_2	1.383	1.359	0.356	0.684	2.129
	ρ	−0.194	−0.256	0.212	−0.968	0.754
100 (400)	β_{10}	−2.121	−2.098	0.436	−3.031	−1.258
	β_{11}	−0.836	−0.828	0.813	−2.547	0.854
	β_{20}	1.512	1.500	0.370	0.773	2.269
	β_{21}	−1.831	−1.818	0.646	−3.104	−0.573
	σ_1	0.584	0.538	0.227	0.021	1.251
	σ_2	1.312	1.301	0.241	0.852	1.795
	ρ	−0.246	−0.306	0.248	−0.963	0.645
Scenario 2						
25 (100)	β_{10}	−1.954	−1.918	0.586	−3.618	−0.375
	β_{11}	−0.417	−0.412	1.081	−3.478	2.623
	β_{20}	1.371	1.351	0.735	−0.202	2.992
	β_{21}	−1.700	−1.676	1.233	−4.496	1.023
	σ_1	0.787	0.637	0.271	0.002	1.999
	σ_2	1.551	1.483	0.622	0.476	2.732
	ρ	−0.133	−0.184	0.182	−0.962	0.833
50 (200)	β_{10}	−2.070	−2.037	0.500	−3.300	−0.897
	β_{11}	−0.617	−0.612	0.915	−2.885	1.636
	β_{20}	1.467	1.448	0.571	0.355	2.613
	β_{21}	−1.838	−1.816	0.962	−3.778	0.054
	σ_1	0.617	0.539	0.213	0.005	1.444
	σ_2	1.449	1.422	0.395	0.735	2.209
	ρ	−0.112	−0.149	0.215	−0.939	0.821
100 (400)	β_{10}	−2.032	−2.009	0.404	−2.909	−1.192
	β_{11}	−0.849	−0.844	0.744	−2.481	0.782
	β_{20}	1.496	1.484	0.391	0.740	2.269
	β_{21}	−1.861	−1.847	0.660	−3.164	−0.578
	σ_1	0.541	0.495	0.201	0.013	1.179
	σ_2	1.358	1.346	0.254	0.888	1.847
	ρ	−0.088	−0.111	0.264	−0.899	0.790

Source: Elaborated by the author.

Table 139 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 1 and 2).

n (m)	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 1			Scenario 2		
25 (100)	β_{10}	1.00	99.00	0.00	1.00	99.00	0.00
	β_{11}	1.20	98.80	0.00	1.40	98.60	0.00
	β_{20}	0.60	95.80	3.60	0.60	95.20	4.20
	β_{21}	4.40	94.80	0.80	2.60	96.40	1.00
	σ_1	0.00	100.00	0.00	0.00	100.00	0.00
	σ_2	0.40	89.20	10.40	1.20	89.60	9.20
	ρ	0.00	100.00	0.00	0.00	99.80	0.20
50 (200)	β_{10}	1.20	97.80	1.00	1.20	98.00	0.80
	β_{11}	2.40	97.60	0.00	3.20	96.80	0.00
	β_{20}	1.00	95.60	3.40	3.60	93.60	2.80
	β_{21}	4.00	95.40	0.60	4.20	94.00	1.80
	σ_1	0.00	100.00	0.00	0.00	100.00	0.00
	σ_2	0.40	92.60	7.00	1.20	92.80	6.00
	ρ	0.00	100.00	0.00	0.00	99.60	0.40
100 (400)	β_{10}	0.80	96.80	2.40	1.80	97.20	1.00
	β_{11}	3.00	95.80	1.20	2.80	96.40	0.80
	β_{20}	1.40	97.00	1.60	1.40	95.60	3.00
	β_{21}	5.60	93.20	1.20	5.20	93.60	1.20
	σ_1	0.20	99.80	0.00	0.20	99.80	0.00
	σ_2	0.20	85.40	14.40	0.60	89.60	9.80
	ρ	0.40	99.20	0.40	0.00	98.00	2.00

Source: Elaborated by the author.

Table 140 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 3 and 4).

n (m)	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 3		
25 (100)	β_{10}	0.118	0.353	1.020	23.497
	β_{11}	0.551	1.451	1.125	96.551
	β_{20}	−0.130	0.520	1.017	36.849
	β_{21}	0.248	1.649	1.019	50.910
	σ_1	0.277	0.166	1.363	56.250
	σ_2	0.112	0.401	1.016	32.773
	ρ	−0.590	0.386	3.190	118.034
50 (200)	β_{10}	−0.045	0.262	1.004	20.370
	β_{11}	0.416	0.984	1.101	78.989
	β_{20}	−0.098	0.328	1.015	29.713
	β_{21}	0.180	0.910	1.018	37.190
	σ_1	0.106	0.049	1.140	29.100
	σ_2	−0.006	0.167	1.000	21.290
	ρ	−0.526	0.327	2.557	105.197
100 (400)	β_{10}	−0.011	0.151	1.000	15.512
	β_{11}	0.234	0.550	1.054	59.962
	β_{20}	−0.047	0.164	1.007	21.406
	β_{21}	0.128	0.478	1.018	27.661
	σ_1	0.018	0.032	1.005	26.643
	σ_2	−0.090	0.078	1.056	15.163
	ρ	−0.434	0.253	1.976	87.993
			Scenario 4		
25 (100)	β_{10}	0.200	0.548	1.038	29.315
	β_{11}	0.363	1.610	1.044	101.983
	β_{20}	0.007	0.276	1.000	27.377
	β_{21}	0.073	0.828	1.003	36.264
	σ_1	−0.420	0.400	1.336	37.421
	σ_2	0.159	0.113	1.135	46.947
	ρ	0.386	0.184	2.291	78.177
50 (200)	β_{10}	0.115	0.345	1.020	23.659
	β_{11}	0.112	0.942	1.007	77.399
	β_{20}	0.075	0.157	1.018	21.419
	β_{21}	−0.049	0.394	1.003	25.241
	σ_1	−0.454	0.366	1.514	34.741
	σ_2	0.013	0.045	1.002	32.666
	ρ	0.356	0.170	1.970	73.033
100 (400)	β_{10}	0.061	0.229	1.008	19.172
	β_{11}	−0.048	0.675	1.002	65.886
	β_{20}	0.019	0.065	1.003	13.685
	β_{21}	0.028	0.182	1.002	17.079
	σ_1	−0.367	0.256	1.451	27.746
	σ_2	−0.066	0.033	1.072	31.055
	ρ	0.288	0.142	1.559	62.178

Source: Elaborated by the author.

Table 141 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 3 and 4).

n (m)	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 3						
25 (100)	β_{10}	−1.883	−1.845	0.583	−3.523	−0.327
	β_{11}	−0.449	−0.445	1.071	−3.460	2.542
	β_{20}	1.370	1.348	0.709	−0.226	3.017
	β_{21}	−1.752	−1.724	1.260	−4.592	1.022
	σ_1	0.777	0.633	0.299	0.002	1.960
	σ_2	1.612	1.542	0.624	0.515	2.830
	ρ	−0.090	−0.123	0.195	−0.945	0.855
50 (200)	β_{10}	−2.045	−2.012	0.510	−3.263	−0.888
	β_{11}	−0.584	−0.577	0.900	−2.813	1.621
	β_{20}	1.402	1.385	0.564	0.298	2.550
	β_{21}	−1.820	−1.798	0.937	−3.767	0.076
	σ_1	0.606	0.531	0.194	0.003	1.414
	σ_2	1.494	1.466	0.408	0.775	2.268
	ρ	−0.026	−0.034	0.224	−0.896	0.876
100 (400)	β_{10}	−2.011	−1.988	0.388	−2.878	−1.183
	β_{11}	−0.766	−0.759	0.704	−2.343	0.804
	β_{20}	1.453	1.442	0.403	0.685	2.241
	β_{21}	−1.872	−1.858	0.679	−3.202	−0.566
	σ_1	0.518	0.475	0.178	0.011	1.129
	σ_2	1.410	1.397	0.265	0.933	1.907
	ρ	0.066	0.085	0.255	−0.816	0.889
Scenario 4						
25 (100)	β_{10}	−1.800	−1.749	0.713	−3.515	−0.194
	β_{11}	−0.637	−0.639	1.216	−3.747	2.501
	β_{20}	1.507	1.488	0.525	0.396	2.655
	β_{21}	−1.927	−1.902	0.907	−3.874	−0.034
	σ_1	1.080	0.948	0.474	0.070	2.388
	σ_2	0.659	0.614	0.296	0.027	1.370
	ρ	−0.114	−0.155	0.187	−0.954	0.832
50 (200)	β_{10}	−1.885	−1.841	0.576	−3.188	−0.669
	β_{11}	−0.888	−0.883	0.964	−3.228	1.441
	β_{20}	1.575	1.562	0.389	0.828	2.344
	β_{21}	−2.049	−2.033	0.626	−3.328	−0.798
	σ_1	1.046	0.995	0.400	0.196	1.973
	σ_2	0.513	0.494	0.213	0.035	1.008
	ρ	−0.144	−0.188	0.210	−0.950	0.781
100 (400)	β_{10}	−1.939	−1.906	0.475	−2.926	−1.010
	β_{11}	−1.048	−1.041	0.820	−2.770	0.663
	β_{20}	1.519	1.512	0.255	1.015	2.032
	β_{21}	−1.972	−1.963	0.426	−2.827	−1.132
	σ_1	1.133	1.111	0.348	0.430	1.856
	σ_2	0.434	0.428	0.169	0.052	0.807
	ρ	−0.212	−0.260	0.241	−0.948	0.657

Source: Elaborated by the author.

Table 142 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 3 and 4).

n (m)	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 3			Scenario 4		
25 (100)	β_{10}	2.00	97.80	0.20	4.20	95.80	0.00
	β_{11}	0.80	99.20	0.00	1.80	97.80	0.40
	β_{20}	1.00	95.60	3.40	0.60	97.40	2.00
	β_{21}	4.60	94.60	0.80	2.80	96.20	1.00
	σ_1	0.00	100.00	0.00	0.00	91.20	8.80
	σ_2	1.00	92.80	6.20	1.00	99.00	0.00
	ρ	0.00	96.80	3.20	0.00	100.00	0.00
50 (200)	β_{10}	1.00	98.60	0.40	3.60	95.80	0.60
	β_{11}	1.60	98.20	0.20	1.80	97.60	0.60
	β_{20}	0.80	94.40	4.80	2.60	94.80	2.60
	β_{21}	4.20	94.60	1.20	2.00	96.40	1.60
	σ_1	0.00	100.00	0.00	0.00	78.20	21.80
	σ_2	1.60	94.00	4.40	1.40	98.40	0.20
	ρ	0.00	97.20	2.80	0.00	100.00	0.00
100 (400)	β_{10}	2.20	97.40	0.40	5.40	94.20	0.40
	β_{11}	3.00	96.60	0.40	1.80	96.80	1.40
	β_{20}	1.40	93.20	5.40	2.00	95.40	2.60
	β_{21}	5.20	93.20	1.60	2.80	95.20	2.00
	σ_1	0.40	99.60	0.00	0.00	79.40	20.60
	σ_2	0.00	91.40	8.60	0.80	95.20	4.00
	ρ	0.00	97.00	3.00	2.00	98.00	0.00

Source: Elaborated by the author.

Table 143 – Empirical properties of the Bayesian estimators using zero-deflated samples (Scenarios 5 and 6).

n (m)	Parameter	Bias	MSE	$\sqrt{\frac{\text{MSE}}{\text{Var}}}$	MAPE (%)
			Scenario 5		
25 (100)	β_{10}	0.269	0.580	1.069	30.479
	β_{11}	0.404	1.791	1.049	106.520
	β_{20}	−0.009	0.306	1.000	29.361
	β_{21}	0.027	0.912	1.000	37.902
	σ_1	−0.392	0.368	1.311	34.688
	σ_2	0.164	0.130	1.123	48.473
	ρ	−0.055	0.039	1.042	-
50 (200)	β_{10}	0.218	0.428	1.060	26.059
	β_{11}	0.140	1.131	1.009	85.405
	β_{20}	0.019	0.145	1.001	19.917
	β_{21}	−0.009	0.400	1.000	24.991
	σ_1	−0.426	0.353	1.434	33.761
	σ_2	0.022	0.047	1.005	32.895
	ρ	−0.048	0.053	1.022	-
100 (400)	β_{10}	0.090	0.215	1.019	18.615
	β_{11}	0.035	0.583	1.001	60.230
	β_{20}	0.004	0.069	1.000	14.035
	β_{21}	0.013	0.188	1.000	17.237
	σ_1	−0.304	0.204	1.355	24.136
	σ_2	−0.065	0.033	1.071	30.677
	ρ	−0.016	0.066	1.002	-
			Scenario 6		
25 (100)	β_{10}	0.312	0.572	1.098	30.850
	β_{11}	0.484	1.673	1.078	104.066
	β_{20}	0.000	0.260	1.000	26.636
	β_{21}	0.015	0.810	1.000	35.589
	σ_1	−0.331	0.384	1.182	35.521
	σ_2	0.138	0.097	1.116	42.243
	ρ	−0.497	0.287	2.667	99.387
50 (200)	β_{10}	0.228	0.412	1.070	25.865
	β_{11}	0.260	1.175	1.030	85.575
	β_{20}	−0.007	0.129	1.000	18.664
	β_{21}	0.003	0.366	1.000	23.664
	σ_1	−0.329	0.264	1.302	28.259
	σ_2	0.015	0.048	1.002	33.470
	ρ	−0.397	0.212	1.974	81.440
100 (400)	β_{10}	0.124	0.229	1.036	19.004
	β_{11}	0.124	0.629	1.012	62.452
	β_{20}	0.008	0.070	1.000	14.003
	β_{21}	−0.010	0.185	1.000	17.136
	σ_1	−0.242	0.159	1.258	21.272
	σ_2	−0.046	0.035	1.032	31.055
	ρ	−0.303	0.163	1.515	66.123

Source: Elaborated by the author.

Table 144 – *Posterior* estimates of model parameters using zero-deflated samples (Scenarios 5 and 6).

n (m)	Parameter	Mean	Median	Std. Dev.	95% HPDI	
					Lower	Upper
Scenario 5						
25 (100)	β_{10}	−1.731	−1.679	0.713	−3.470	−0.112
	β_{11}	−0.596	−0.600	1.276	−3.704	2.521
	β_{20}	1.491	1.473	0.553	0.364	2.652
	β_{21}	−1.973	−1.949	0.954	−3.931	−0.060
	σ_1	1.108	0.984	0.463	0.079	2.401
	σ_2	0.664	0.618	0.321	0.034	1.381
	ρ	−0.055	−0.075	0.189	−0.928	0.868
50 (200)	β_{10}	−1.782	−1.740	0.617	−3.062	−0.594
	β_{11}	−0.860	−0.856	1.054	−3.123	1.401
	β_{20}	1.519	1.506	0.381	0.774	2.282
	β_{21}	−2.009	−1.993	0.632	−3.288	−0.770
	σ_1	1.074	1.027	0.414	0.248	1.968
	σ_2	0.522	0.504	0.216	0.037	1.021
	ρ	−0.048	−0.061	0.224	−0.897	0.843
100 (400)	β_{10}	−1.910	−1.880	0.454	−2.893	−0.985
	β_{11}	−0.965	−0.958	0.763	−2.672	0.727
	β_{20}	1.504	1.497	0.263	1.000	2.022
	β_{21}	−1.987	−1.979	0.434	−2.843	−1.144
	σ_1	1.196	1.172	0.333	0.538	1.888
	σ_2	0.435	0.428	0.170	0.047	0.811
	ρ	−0.016	−0.019	0.255	−0.838	0.812
Scenario 6						
25 (100)	β_{10}	−1.688	−1.640	0.689	−3.408	−0.070
	β_{11}	−0.516	−0.516	1.200	−3.596	2.571
	β_{20}	1.500	1.482	0.510	0.399	2.638
	β_{21}	−1.985	−1.961	0.900	−3.902	−0.112
	σ_1	1.169	1.050	0.524	0.155	2.436
	σ_2	0.638	0.593	0.280	0.024	1.342
	ρ	0.003	0.005	0.201	−0.892	0.897
50 (200)	β_{10}	−1.772	−1.731	0.600	−3.068	−0.556
	β_{11}	−0.740	−0.734	1.052	−3.022	1.523
	β_{20}	1.493	1.480	0.359	0.756	2.254
	β_{21}	−1.997	−1.982	0.605	−3.263	−0.755
	σ_1	1.171	1.127	0.394	0.353	2.050
	σ_2	0.515	0.496	0.218	0.038	1.011
	ρ	0.103	0.131	0.233	−0.786	0.920
100 (400)	β_{10}	−1.876	−1.847	0.462	−2.855	−0.955
	β_{11}	−0.876	−0.869	0.783	−2.557	0.789
	β_{20}	1.508	1.501	0.265	1.001	2.028
	β_{21}	−2.010	−2.001	0.430	−2.872	−1.163
	σ_1	1.258	1.235	0.316	0.636	1.918
	σ_2	0.454	0.449	0.181	0.064	0.832
	ρ	0.197	0.234	0.266	−0.630	0.923

Source: Elaborated by the author.

Table 145 – Coverage probabilities (%) of the HPDIs using zero-deflated samples (Scenarios 5 and 6).

$n\ (m)$	Parameter	BNCP	CP	ANCP	BNCP	CP	ANCP
		Scenario 5			Scenario 6		
25 (100)	β_{10}	5.40	94.60	0.00	3.80	96.20	0.00
	β_{11}	1.80	98.20	0.00	2.80	97.00	0.20
	β_{20}	1.60	95.60	2.80	0.60	97.20	2.20
	β_{21}	3.20	95.40	1.40	2.20	97.00	0.80
	σ_1	0.00	92.40	7.60	0.00	90.20	9.80
	σ_2	1.80	98.20	0.00	1.00	99.00	0.00
	ρ	0.00	99.60	0.40	0.00	97.40	2.60
50 (200)	β_{10}	5.80	93.00	1.20	5.20	94.20	0.60
	β_{11}	2.40	96.60	1.00	4.00	95.80	0.20
	β_{20}	2.60	95.20	2.20	1.40	96.00	2.60
	β_{21}	3.60	94.20	2.20	3.00	94.80	2.20
	σ_1	0.00	77.40	22.60	0.00	84.40	15.60
	σ_2	1.40	98.60	0.00	1.80	98.20	0.00
	ρ	0.00	100.00	0.00	0.00	98.00	2.00
100 (400)	β_{10}	6.20	93.40	0.40	6.00	93.40	0.60
	β_{11}	1.60	97.60	0.80	2.80	95.60	1.60
	β_{20}	1.60	95.40	3.00	2.00	95.20	2.80
	β_{21}	3.00	94.60	2.40	2.60	95.20	2.20
	σ_1	0.00	84.80	15.20	0.00	87.00	13.00
	σ_2	1.00	96.20	2.80	1.60	95.20	3.20
	ρ	0.80	98.40	0.80	0.00	97.40	2.60

Source: Elaborated by the author.