

Article

Bayesian Inference for Zero-Modified Power Series Regression Models

Katiane S. Conceição¹, Marinho G. Andrade¹ , Victor Hugo Lachos² and Nalini Ravishanker^{2,*} 

¹ Department of Applied Mathematics and Statistics, Institute of Mathematics and Computer Science, University of São Paulo, São Carlos 13566-590, SP, Brazil; katiane@icmc.usp.br (K.S.C.); marinho@icmc.usp.br (M.G.A.)

² Department of Statistics, University of Connecticut, Storrs, CT 06269, USA; hlachos@uconn.edu

* Correspondence: nalini.ravishanker@uconn.edu

Abstract: Count data often exhibit discrepancies in the frequencies of zeros, which commonly occur across various application domains. These data may include excess zeros (zero inflation) or, less frequently, a scarcity of zeros (zero deflation). In regression models, both situations can arise at different levels of covariates. The zero-modified power series regression model provides an effective framework for modeling such count data, as it does not require prior knowledge of the type of zero modification, whether zero inflation or zero deflation, and can accommodate overdispersion, equidispersion, or underdispersion present in the data. This paper proposes a Bayesian estimation procedure based on the stochastic gradient Hamiltonian Monte Carlo algorithm, effectively addressing many challenges associated with estimating the model parameters. Additionally, we introduce a measure of Bayesian efficiency to evaluate the impact of prior information on parameter estimation. The practical utility of the proposed method is demonstrated through both simulated and real data across different types of zero modification.

Keywords: zero-modified power series models; SGHMC algorithm; Bayesian efficiency; Brazilian femicide notification data

MSC: 63P25; 62C20; 62-08; 60-08



Academic Editor: Heng Lian

Received: 13 October 2024

Revised: 24 December 2024

Accepted: 24 December 2024

Published: 27 December 2024

Citation: Conceição, K.S.; Andrade, M.G.; Lachos, V.H.; Ravishanker, N. Bayesian Inference for Zero-Modified Power Series Regression Models. *Mathematics* **2025**, *13*, 60. <https://doi.org/10.3390/math13010060>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Applied statistical data analysis of count (enumeration) responses generally requires assumptions about discrete probability distributions generating the response [1–4]. In many situations, count datasets are analyzed using the Poisson distribution (see [5–8]) when we can assume that the mean and variance of the counts are the same. When the variance is assumed to be greater than or smaller than the mean, several flexible probability distributions have been proposed as alternatives to the Poisson distribution for modeling overdispersion and underdispersion [9]. They include discrete probability distributions belonging to the class of power series (PS) distributions or their generalizations [1–4,9–12].

Another often encountered situation with count data involves discrepancies in the frequency of zeros from that corresponding to the Poisson distribution. While zero inflation is most commonly observed, zero deflation can also occur in some situations. Pioneering work on these topics includes M'Kendrick [13], who studied the number of cancer cases between 1875 and 1898 in two suburbs of Luckau in Hanover, the data corresponding to a cholera epidemic in a village in India. David and Johnson [14] analyzed the number of defective teeth in 11-year-old boys and the number of ticks found on sheep. Statistical

modeling of zero-inflated count data was proposed by [15,16] via the zero-inflated Poisson (ZIP) distribution, which is a mixture of a degenerate zero distribution and the Poisson distribution. For a review of ZIP models for count data, see [17]. In Dietz and Böhning [18], a zero-modified Poisson (ZMP) regression model was considered for analyzing zero-inflated or zero-deflated data. An application of the zero-inflated Poisson (ZIP) distribution for modeling data from natural calamities was presented by [19].

More recently, ref. [20] discussed zero-modified power series (ZMPS) distributions, which can adequately accommodate the discrepancy in the frequency of zeros without requiring any previous knowledge about the type of modification (zero-inflation or zero-deflation). As a special case, they illustrated the zero-modified negative binomial (ZMNB) distribution on simulated and real data, obtaining maximum likelihood estimates of model parameters. To our knowledge, the flexible and useful class of ZMPS distributions has not been employed in the literature for modeling dispersed count responses as a function of covariates, particularly in the Bayesian framework.

This paper presents a regression model for count data with the flexibility to handle different types of zero frequencies (inflation or deflation), as well as different dispersion characteristics (overdispersion or underdispersion) within a single dataset. Our framework includes, as special cases, standard regression models and zero-inflated models for count data. This paper addresses this issue in the Bayesian framework.

The sampling-based Bayesian framework using Markov Chain Monte Carlo (MCMC) methods [21] is attractive because the ZMPS likelihood function is complex, making exact Bayesian estimation and inference cumbersome. Most MCMC algorithms include an acceptance–rejection step of the generated candidate (the Metropolis–Hastings (MH) step) and may result in a high rejection rate in complex models, leading to slow convergence. To remedy this, several computationally efficient algorithms have been proposed in the literature, such as the Hamiltonian Monte Carlo (HMC) algorithm [22], the Metropolis Adjusted Langevin algorithm (MALA) [23–25], and an improvement of the HMC algorithm known as the No-U-Turn Sampler (NUTS) [26].

Further, the general class of algorithms using stochastic gradient MCMC (SGMCMC) [27–31] has gained recent popularity in the machine learning area but is still little used in the Bayesian context. An R package `sgmcmc` for stochastic gradient MCMC [30] calculates the gradients using a numerical computation library called `tensorflow` [32], which is a machine learning system with an R interface, operating on large, heterogeneous environments. In comparing algorithms in the SGMCMC class with the SGHMC algorithm, the latter is one of the most computationally efficient algorithms [31].

The objective of this paper is to develop a fully Bayesian approach for estimating the parameters of the ZMPS regression models via the computationally efficient stochastic gradient with friction HMC (SGHMC) algorithm proposed by [28]. A comprehensive study of various MCMC simulation methods is given by [33], along with a clear explanation of stochastic gradient methods. Since the required gradient equations can be obtained from the ZMPS models, we implement the SGHMC algorithm with the leapfrog integration method. We also derive and illustrate a relative Bayesian efficiency measure to compare the more general ZMPS distribution with the standard Power Series (PS) distribution.

The rest of the paper is organized as follows. Section 2.1 describes the ZMPS regression model framework for count responses. Section 2.3 presents the Bayesian approach for model parameter estimation, while Section 2.7 describes the SGHMC algorithm. Section 2.9 discusses the Bayesian information used to assess the efficiency loss incurred by fitting a ZMPS model when a PS distribution could have been applied. Section 3.1 presents simulation studies to evaluate the impact of the prior distribution on the quality of the Bayesian estimator. Section 3.2 reports the analysis of real datasets that consider different

characteristics of the ZMPS distribution to analyze the number of femicides in 642 São Paulo municipalities in 2019 and 2020; in this application, we want to assess whether there was an increase in the number of femicides during the COVID-19 pandemic period. Section 4 draws some final comments on the relevant results achieved.

2. Materials and Methods

A regression model for count data is developed here, incorporating the flexibility of the ZMPS distribution, along with a Bayesian framework for parameter estimation.

2.1. ZMPS Regression Model

In ZMPS regression, suppose that the observations to be analyzed are represented by the vector $\mathbf{y} = (y_1, \dots, y_n)'$. Let $Y_i, i = 1 \dots n$ be n independent random variables assuming values in the support set $\mathcal{A} = \{0, 1, \dots\}$ and having probability mass function (p.m.f.) defined by [20]

$$\pi_{ZMPS}(y_i; \mu_i, \phi, p_i) = (1 - p_i)I(y_i) + p_i\pi_{PS}(y_i; \mu_i, \phi), \tag{1}$$

where $\mu_i > 0, \phi > 0$, and the indicator function $I(y_i) = 1$ if $y_i = 0$ and $I(y_i) = 0$ if $y_i > 0$. The p.m.f. of the standard PS distribution is given by

$$\pi_{PS}(y_i; \mu_i, \phi) = \frac{a(y_i, \phi)(g(\mu_i, \phi))^{y_i}}{f(\mu_i, \phi)},$$

where $a(y_i, \phi), f(\mu_i, \phi)$ and $g(\mu_i, \phi)$ are positive, finite, and twice-differentiable functions, with $f(\mu_i, \phi) = \sum_{y_i \in \mathcal{A}_0} a(y_i, \phi)g(\mu_i, \phi)^{y_i}$, while the parameter p_i is responsible for modifying the probability of zero of the standard PS distribution, such that

$$0 \leq p_i \leq \frac{1}{1 - \pi_{PS}(0; \mu_i, \phi)},$$

where $\pi_{PS}(0; \mu_i, \phi)$ is the probability of zero in the PS distribution.

The Poisson, generalized Poisson, COM-Poisson, and negative binomial distributions are well-known members of the PS family. Table 1 presents the functions $f(\mu_i, \phi), g(\mu_i, \phi)$, and $a(y_i, \phi)$ for several distributions within the PS family.

Table 1. Distributions within the PS family.

PS	Distribution	$f(\mu_i, \phi)$	$g(\mu_i, \phi)$	$a(y_i, \phi)$
P	Poisson	e^{μ_i}	μ_i	$\frac{1}{y_i!}$
GP	Generalized Poisson	$e^{\mu_i(1+\mu_i\phi)^{-1}}$	$\frac{\mu_i e^{-\mu_i\phi(1+\mu_i\phi)^{-1}}}{1 + \mu_i\phi}$	$\frac{(1 + \phi y_i)^{y_i-1}}{y_i!}$
COMP	COM-Poisson	$\sum_{s=0}^{\infty} \left(\frac{\mu_i^s}{s!}\right)^\phi$	μ_i^ϕ	$\left(\frac{1}{y_i!}\right)^\phi$
NB	Negative Binomial	$\left(\frac{\phi}{\mu_i + \phi}\right)^{-\phi}$	$\frac{\mu_i}{\mu_i + \phi}$	$\frac{\Gamma(\phi + y_i)}{y_i! \Gamma(\phi)}$

For the COM-Poisson distribution, we consider the parameterization proposed in [34] to provide a clearly centering parameter. The conditional mean and variance of Y_i , $E(Y_i) = \mu_i^{PS}$ and $V(Y_i) = v_i^{PS}$, are given by

$$\mu_i^{PS} = \frac{f_\mu(\mu_i, \phi)g(\mu_i, \phi)}{f(\mu_i, \phi)g_\mu(\mu_i, \phi)} \quad \text{and} \quad v_i^{PS} = \frac{g(\mu_i, \phi)}{g_\mu(\mu_i, \phi)},$$

where f_μ and g_μ denote derivatives with respect to μ . For the distributions shown in Table 1, the mean $\mu_i^{PS} = \mu_i$, except for the COM-Poisson distribution, for which we can write the asymptotic approximations [34]:

$$\mu_i^{PS} \approx \mu_i + \frac{1 - \phi}{2\phi} \quad \text{and} \quad v_i^{PS} \approx \frac{\mu_i}{\phi}.$$

This parametrization also allows ϕ to keep its role as a shape parameter. That is, if $\phi < 1$, the variance is greater than the mean (overdispersion), while $\phi > 1$ leads to underdispersion. The mean and variance of random variable Y_i are, respectively,

$$\mu_i^{(ZMPS)} = p_i \mu_i^{PS} \quad \text{and} \quad \sigma_i^{2(ZMPS)} = p_i \{v_i^{PS} + (1 - p_i)(\mu_i^{PS})^2\},$$

Different values of p_i lead to different ZMPS distributions, as can be seen in the evaluation of the proportion of additional or missing zeros, given by

$$\begin{aligned} \pi_{ZMPS}(0; \mu_i, \phi, p_i) - \pi_{PS}(0; \mu_i, \phi) &= 1 - p_i + p_i \pi_{PS}(0; \mu_i, \phi) - \pi_{PS}(0; \mu_i, \phi) \\ &= (1 - p_i)(1 - \pi_{PS}(0; \mu_i, \phi)). \end{aligned} \tag{2}$$

According to (2), parameter p_i also controls the frequency of zeros, such as the following:

- (i) When $p_i = 0$ in (2), $\pi_{ZMPS}(0; \mu_i, \phi, p_i) = 1$, and (1) is the degenerate distribution with the entire mass at zero.
- (ii) For all $0 < p_i < 1$ in (2), $(1 - p_i)(1 - \pi_{PS}(0; \mu_i, \phi)) > 0$. Then, $\pi_{ZMPS}(0; \mu_i, \phi, p_i) > \pi_{PS}(0; \mu_i, \phi)$, and (1) is the zero-inflated PS (ZIPS) distribution, which has a proportion of additional zeros.
- (iii) When $p_i = 1$ in (2), $\pi_{ZMPS}(0; \mu_i, \phi, p_i) - \pi_{PS}(0; \mu_i, \phi) = 0$. Therefore, $\pi_{ZMPS}(0; \mu_i, \phi, p_i) = \pi_{PS}(0; \mu_i, \phi)$, and (1) is the standard PS distribution.
- (iv) For all $1 < p_i < \frac{1}{1 - \pi_{PS}(0; \mu_i, \phi)}$ in (2), $(1 - p_i)(1 - \pi_{PS}(0; \mu_i, \phi)) < 0$. Then, $\pi_{ZMPS}(0; \mu_i, \phi, p_i) < \pi_{PS}(0; \mu_i, \phi)$, and (1) is the zero-deflated negative binomial (ZDPS) distribution.
- (v) When $p_i = \frac{1}{1 - \pi_{PS}(0; \mu_i, \phi)}$ in (2), $\pi_{ZMPS}(0; \mu_i, \phi, p_i) = 0$. Then, (1) is the zero-truncated PS (ZTPS) distribution, whose p.m.f. is given by

$$\pi_{ZTPS}(y_i; \mu_i, \phi) = \frac{\pi_{PS}(y_i; \mu_i, \phi)}{1 - \pi_{PS}(0; \mu_i, \phi)} (1 - I(y_i)), \quad y_i \in \mathcal{A}. \tag{3}$$

Note that the parameter p_i controls the frequency of zeros and can assume different values at specific levels of covariates. Conditions (i)–(v) indicate that in zero inflation count data, deflation or standard zero frequency may occur in the same dataset at specific levels of covariates. Therefore, it is plausible that the ZMPS model can model these situations by considering different values of p_i . More details about the ZMPS distributions are provided in Conceição et al. [20].

2.2. Hurdle Model Versus ZMPS Model

Note that (1) can be rewritten as

$$\begin{aligned}
 \pi_{ZMPS}(y_i; \mu_i, \phi, p_i) &= \pi_{ZMPS}(0; \mu_i, \phi, p_i)I(y_i) + \pi_{ZMPS}(y_i; \mu_i, \phi, p_i)(1 - I(y_i)) \\
 &= \{1 - p_i + p_i\pi_{PS}(0; \mu_i, \phi)\}I(y_i) + \{p_i\pi_{PS}(y_i; \mu_i, \phi)\}(1 - I(y_i)) \\
 &= \{1 - p_i(1 - \pi_{PS}(0; \mu_i, \phi))\}I(y_i) + \\
 &\quad p_i(1 - \pi_{PS}(0; \mu_i, \phi)) \left\{ \frac{\pi_{PS}(y_i; \mu_i, \phi)}{1 - \pi_{PS}(0; \mu_i, \phi)} (1 - I(y_i)) \right\}. \tag{4}
 \end{aligned}$$

Since $0 \leq p_i \leq \frac{1}{1 - \pi_{PS}(0; \mu_i, \phi)}$, then $0 \leq p_i(1 - \pi_{PS}(0; \mu_i, \phi)) \leq 1$. Thus, setting $\omega_i = p_i(1 - \pi_{PS}(0; \mu_i, \phi))$, the p.m.f. given in (4) can be written as a mixture distribution, i.e.,

$$\pi_{ZMPS}(y_i; \mu_i, \phi, \omega_i) = (1 - \omega_i)I(y_i) + \omega_i\pi_{ZTPS}(y_i; \mu_i, \phi), \quad y_i \in \mathcal{A}, \tag{5}$$

where $0 \leq \omega_i \leq 1$

The ZMPS distribution in (5) is parameterized by ω_i and can be interpreted as a hurdle distribution [35], where the probability of the event $Y_i = 0$ is $1 - \omega_i$, while the probability of an observation $Y_i = y_i, y_i > 0$ is $\omega_i\pi_{ZTPS}(y_i; \mu_i, \phi)$.

Properties (i)–(v) can also be written in terms of ω_i . Therefore, the ZMPS model (1) and the hurdle model (5) are equivalent, and either can be used to model data with zero-frequency modifications (inflation or deflation). Ref. [36] compared the hurdle model to zero-inflated models, highlighting the advantages of hurdle models (or ZMPS) due to the viability of these models to consider zero-inflation ($0 < p_i < 1$) and zero-deflation ($1 < p_i \leq 1/(1 - \pi_{PS}(0; \mu_i, \phi))$) with different levels of covariates. Henceforth, we will consider the hurdle version of ZMPS given by (5).

However, it is essential to note that ZMPS models may have a drawback when $p_i = 1$ for all $i = 1, \dots, n$. A standard PS model could have been used without fitting the parameters p_i . We can only say that fitting a ZMPS model instead of a simple PS model in this scenario may result in efficiency losses for the parameter estimators. This paper offers a method to assess this efficiency loss from a Bayesian perspective (see Section 2.9).

Let $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,r})'$ with $x_{i,1} = 1$, and $\mathbf{v}_i = (v_{i,1}, \dots, v_{i,s})'$, with $v_{i,1} = 1$, be fixed predictors. Let $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_n)'$ and $\mathbf{V} = (\mathbf{v}'_1, \dots, \mathbf{v}'_n)'$, respectively, denote the $n \times r$ and $n \times s$ predictor matrices employed for modeling the μ_i and ω_i , respectively, via the dual-link function given by logarithmic and logit link functions, i.e.,

$$\begin{aligned}
 \log(\mu_i) &= \eta_1(\boldsymbol{\beta}_\mu, \mathbf{x}_i) = \beta_{\mu,1} + \sum_{j=2}^r \beta_{\mu,j}x_{i,j}, \\
 \text{logit}(\omega_i) &= \eta_2(\boldsymbol{\beta}_\omega, \mathbf{v}_i) = \beta_{\omega,1} + \sum_{j=2}^s \beta_{\omega,j}v_{i,j}.
 \end{aligned}$$

where $\boldsymbol{\beta}_\mu = (\beta_{\mu,1} \dots \beta_{\mu,r})'$, $\boldsymbol{\beta}_\omega = (\beta_{\omega,1} \dots \beta_{\omega,s})'$. It is also possible to employ alternate link functions for ω_i ; see [37].

The parameter ϕ accounts for under- or overdispersion in the data, in addition to any dispersion caused by zero inflation or zero deflation. In the case of the zero-modified Poisson distribution, $\phi = 1$, and all the overdispersion is attributed to zero inflation. In other models, some of the under- or overdispersion is explained by the parameters ω_i and μ_i . Therefore, we interpret ϕ as capturing any under- or overdispersion beyond that induced by zero-modification, without the need for additional covariates.

2.3. Bayesian Framework for the ZMPS Regression Model

Let $\mathcal{D} = \{\mathbf{y}, \mathbf{X}, \mathbf{V}\}$ denote the observed data, based on which we estimate the parameters $\beta'_\mu, \beta'_\omega, \phi$ under the model in (5) using the fully Bayesian framework.

2.4. Likelihood Function

The likelihood function is

$$\mathcal{L}(\beta_\mu, \beta_\omega, \phi | \mathcal{D}) = \prod_{i=1}^n (1 - \omega_i)^{I(y_i)} (\omega_i \pi_{ZTPS}(y_i; \mu_i, \phi))^{(1-I(y_i))}, \tag{6}$$

where $\mu_i > 0$ and $0 < \omega_i < 1$. It follows that the log-likelihood function is

$$\begin{aligned} \ell(\beta_\mu, \beta_\omega, \phi | \mathcal{D}) &= \sum_{i=1}^n I(y_i) \log(1 - \omega_i) + (1 - I(y_i)) \log(\omega_i \pi_{ZTPS}(y_i | \mu_i, \phi)), \\ &= \sum_{i=1}^n \left\{ I(y_i) \log\left(\frac{1 - \omega_i}{\omega_i}\right) + \log(\omega_i) \right\} + \\ &\quad \sum_{i=1}^n (1 - I(y_i)) \log(\pi_{ZTPS}(y_i | \mu_i, \phi)) \\ &= \ell_1(\beta_\mu, \phi | \mathcal{D}) + \ell_2(\beta_\omega | \mathcal{D}), \end{aligned}$$

where

$$\ell_1(\beta_\mu, \phi | \mathcal{D}) = \sum_{i=1}^n (1 - I(y_i)) \log(\pi_{ZTPS}(y_i | \mu_i, \phi, \mathcal{D})) \tag{7}$$

and

$$\ell_2(\beta_\omega | \mathcal{D}) = \sum_{i=1}^n I(y_i) \log\left(\frac{1 - \omega_i}{\omega_i}\right) + \log(\omega_i). \tag{8}$$

Here, $\ell_1(\beta_\mu, \phi | \mathcal{D})$ is the portion of the log-likelihood function from the associated ZTPS distribution, considering only the positive observations of \mathbf{y} , and $\ell_2(\beta_\omega | \mathcal{D})$ is the part associated with the hurdle parameter β_ω . Therefore, we can estimate the parameters β_μ and ϕ of the ZMPS distribution from its associated ZTPS distribution using only the positive observations of the sample \mathbf{y} . The following result states a relevant link between the estimation of the parameters of ZMPS and ZTPS distributions.

Result 1. *If Y_i follows the ZMPS distribution, we can estimate the β_μ and ϕ parameters by only considering the positive observations of \mathbf{y} and the associated ZTPS distribution.*

To verify this result, let $\mathbf{y}^+ = (y_1^+, \dots, y_m^+)'$ denote the positive observations in \mathbf{y} , where $m = n - n_0$, and n_0 is the number of zero observations in the vector \mathbf{y} . Let \mathbf{X}^+ denote the $m \times r$ predictor matrix associated with \mathbf{y}^+ . Assuming that \mathbf{y}^+ follows the associated ZTPS distribution and $\mathcal{D}^+ = (\mathbf{y}^+, \mathbf{X}^+)$, the log-likelihood function associated with \mathbf{y}^+ is given by

$$\ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+) = \sum_{i=1}^m \log(\pi_{ZTPS}(y_i^+ | \mu_i, \phi, \mathcal{D}^+)). \tag{9}$$

We can easily show that (9) is obtained directly from (7). We see from (7) and (8) that estimation of (β_μ, ϕ) independently of the parameter is independent of estimating β_ω . Substituting

$$\pi_{ZTPS}(y_i^+ | \mu_i, \phi) = \frac{a(y_i^+, \phi)g(\mu_i, \phi)^{y_i^+}}{f(\mu_i, \phi) - 1}, y_i^+ \in \{1, 2, \dots\}$$

in (9) and differentiating $\ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+)$ with respect to each parameter, $\beta_{\mu,j}, j = 1, \dots, r$, and ϕ , we obtain the score vectors \mathcal{U}_{β_μ} and \mathcal{U}_ϕ with elements:

$$\mathcal{U}_{\beta_{\mu,j}} = \frac{\partial \ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+)}{\partial \beta_{\mu,j}} = \sum_{i=1}^m \left[y_i^+ \frac{g_\mu(\mu_i, \phi)}{g(\mu_i, \phi)} - \frac{f_\mu(\mu_i, \phi)}{f(\mu_i, \phi) - 1} \right] \frac{\partial \mu_i}{\partial \beta_{\mu,j}}$$

and

$$\mathcal{U}_\phi = \frac{\partial \ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+)}{\partial \phi} = \sum_{i=1}^m \left[\frac{a_\phi(y_i^+, \phi)}{a(y_i, \phi)} + y_i^+ \frac{g_\phi(\mu_i, \phi)}{g(\mu_i, \phi)} - \frac{f_\phi(\mu_i, \phi)}{f(\mu_i, \phi) - 1} \right],$$

where

$g_\mu(\mu_i, \phi) = \partial g(\mu_i, \phi) / \partial \mu_i$, $f_\mu = \partial f(\mu_i, \phi) / \partial \mu_i$, $a_\phi = \partial a(y_i, \phi) / \partial \phi$, $g_\phi = \partial g(\mu_i, \phi) / \partial \phi$, and $f_\phi = \partial f(\mu_i, \phi) / \partial \phi$. Considering the logit link function $\text{logit}(\omega_i) = \log(\omega_i / (1 - \omega_i)) = v_i' \beta_\omega$ and differentiating $\ell_2(\beta_\omega | \mathcal{D})$ with respect to each parameter, $\beta_{\omega,j}, j = 1, \dots, s$, we obtain the score vector $\mathcal{U}_{\beta_\omega}$, whose elements are

$$\mathcal{U}_{\beta_{\omega,j}} = \frac{\partial \ell_2(\beta_\omega | \mathcal{D})}{\partial \beta_{\omega,j}} = \sum_{i=1}^n (1 - I(y_i)) v_{i,j} - \left(\frac{e^{v_i' \beta_\omega}}{1 + e^{v_i' \beta_\omega}} \right) v_{i,j}.$$

The score vector, denoted by $\mathcal{U}(\beta_\mu, \phi, \beta_\omega) = (\mathcal{U}_{\beta_\mu}(\beta_\mu, \phi), \mathcal{U}_\phi(\beta_\mu, \phi), \mathcal{U}_{\beta_\omega}(\beta_\omega))'$, will play a role in the MCMC simulation algorithm in the Bayesian approach.

Let the parameter vector be denoted by $\theta = (\beta_\mu', \beta_\omega', \phi)'$, where $\theta = (\theta_1, \dots, \theta_m)'$ and $m = r + s + 1$. The observed information matrix $\mathcal{J}(\theta)$ is an $(r + s + 1) \times (r + s + 1)$ matrix, with entries given by

$$\mathcal{J}_{j,k} = - \frac{\partial^2 \log \mathcal{L}(\theta | \mathcal{D})}{\partial \theta_j \partial \theta_k}, \quad j, k = 1, \dots, (r + s + 1).$$

Due to the orthogonality between the parameters (β_μ, ϕ) and β_ω , the information matrix $\mathcal{J}(\theta)$ is a block diagonal matrix given by

$$\mathcal{J}(\theta) = \begin{bmatrix} \mathcal{J}_{\beta_\mu, \beta_\mu} & \mathcal{J}_{\beta_\mu, \phi} & \mathbf{0} \\ \mathcal{J}'_{\beta_\mu, \phi} & \mathcal{J}_{\phi, \phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{J}_{\beta_\omega, \beta_\omega} \end{bmatrix}. \tag{10}$$

The elements of each block are defined by

$$\begin{aligned} \mathcal{J}_{\beta_{\mu,j}, \beta_{\mu,k}} &= - \frac{\partial^2 \ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+)}{\partial \beta_{\mu,j} \partial \beta_{\mu,k}} \\ &= \sum_{i=1}^m \left[-y_i^+ h_g(\beta_\mu, \phi) + h_f(\beta_\mu, \phi) \right] \left(\frac{\partial \mu_i}{\partial \beta_{\mu,j}} \right) \left(\frac{\partial \mu_i}{\partial \beta_{\mu,k}} \right) - \\ &\quad \sum_{i=1}^m \left[y_i^+ \frac{g_\mu(\mu_i, \phi)}{g(\mu_i, \phi)} - \frac{f_\mu(\mu_i, \phi)}{f(\mu_i, \phi) - 1} \right] \left(\frac{\partial^2 \mu_i}{\partial \beta_{\mu,j} \partial \beta_{\mu,k}} \right), \end{aligned}$$

$$\begin{aligned} \mathcal{J}_{\beta_\mu, \phi} &= \mathcal{J}_{\phi, \beta_\mu} = - \frac{\partial^2 \ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+)}{\partial \phi \partial \beta_\mu} \\ &= \sum_{i=1}^m \left[-y_i^+ t_g(\beta_\mu, \phi) + t_f(\beta_\mu, \phi) \right] \left(\frac{\partial \mu_i}{\partial \beta_{\mu,j}} \right), \end{aligned}$$

$$\begin{aligned} \mathcal{J}_{\phi, \phi} &= -\frac{\partial^2 \ell_{ZTPS}(\boldsymbol{\beta}_\mu, \phi | \mathcal{D}^+)}{\partial \phi^2} \\ &= \sum_{i=1}^m [-k_a(y_i^+, \phi) - y_i^+ k_g(\boldsymbol{\beta}_\mu, \phi) + k_f(\boldsymbol{\beta}_\mu, \phi)], \\ \mathcal{J}_{\boldsymbol{\beta}_{\omega_j}, \boldsymbol{\beta}_{\omega_k}} &= -\frac{\partial^2 \ell_2(\boldsymbol{\beta}_\omega | \mathcal{D})}{\partial \boldsymbol{\beta}_\omega \partial \boldsymbol{\beta}'_\omega} = \sum_{i=1}^n \frac{\mathbf{x}'_i \mathbf{x}_i \omega_i}{1 - e^{\mathbf{x}'_i \boldsymbol{\beta}_\omega}}, \end{aligned}$$

where

$$\begin{aligned} h_g(\boldsymbol{\beta}_\mu, \phi) &= \frac{g_{\mu\mu}(\mu_i, \phi)g(\mu_i, \phi) - g_\mu^2(\mu_i, \phi)}{g(\mu_i, \phi)^2}, \\ h_f(\boldsymbol{\beta}_\mu, \phi) &= \frac{f_{\mu\mu}(\mu_i, \phi)(f(\mu_i, \phi) - 1) - f_\mu^2(\mu_i, \phi)}{(f(\mu_i, \phi) - 1)^2}, \\ k_a(y_i^+, \phi) &= \frac{a_{\phi\phi}(y_i^+, \phi)a(y_i^+, \phi) - a_\phi^2(y_i^+, \phi)}{a(y_i^+, \phi)^2}, \\ k_g(\boldsymbol{\beta}_\mu, \phi) &= \frac{g_{\phi\phi}(\mu_i, \phi)g(\mu_i, \phi) - g_\phi^2(\mu_i, \phi)}{g(\mu_i, \phi)^2}, \\ k_f(\boldsymbol{\beta}_\mu, \phi) &= \frac{f_{\phi\phi}(\mu_i, \phi)(f(\mu_i, \phi) - 1) - f_\phi^2(\mu_i, \phi)}{(f(\mu_i, \phi) - 1)^2}, \\ t_g(\boldsymbol{\beta}_\mu, \phi) &= \frac{g_{\mu\phi}(\mu_i, \phi)g(\mu_i, \phi) - g_\mu(\mu_i, \phi)g_\phi(\mu_i, \phi)}{g(\mu_i, \phi)^2}, \\ t_f(\boldsymbol{\beta}_\mu, \phi) &= \frac{f_{\mu\phi}(\mu_i, \phi)(f(\mu_i, \phi) - 1) - f_\mu(\mu_i, \phi)f_\phi(\mu_i, \phi)}{(f(\mu_i, \phi) - 1)^2}, \end{aligned}$$

and, $g_{\mu\mu}(\mu_i, \phi) = \partial^2 g(\mu_i, \phi) / \partial \mu_i^2$, $f_{\mu\mu} = \partial^2 f(\mu_i, \phi) / \partial \mu_i^2$, $a_{\phi\phi} = \partial^2 a(y_i, \phi) / \partial \phi^2$, $g_{\phi\phi} = \partial^2 g(\mu_i, \phi) / \partial \phi^2$, and $f_{\phi\phi} = \partial^2 f(\mu_i, \phi) / \partial \phi^2$, $g_{\mu\phi}(\mu_i, \phi) = \partial^2 g(\mu_i, \phi) / \partial \mu_i \partial \phi$, $f_{\mu\phi} = \partial^2 f(\mu_i, \phi) / \partial \mu_i \partial \phi$.

2.5. Prior Distributions

Assuming prior independence, we have $\pi_0(\boldsymbol{\beta}_\mu, \phi, \boldsymbol{\beta}_\omega) = \pi_0(\boldsymbol{\beta}_\mu)\pi_0(\phi)\pi_0(\boldsymbol{\beta}_\omega)$. We propose a multivariate normal prior distribution for $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_\omega$, i.e., $\boldsymbol{\beta}_\mu \sim N(\boldsymbol{\beta}_0, \sigma_0^2 \mathbf{I}_r)$ and $\boldsymbol{\beta}_\omega \sim N(\boldsymbol{\beta}_1, \sigma_1^2 \mathbf{I}_s)$, where $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are hyperparameter vectors of suitable dimensions, and σ_0^2 and σ_1^2 are prior variances. We assume that $\phi \sim LN(\phi_0, \sigma_2^2)$, where LN denotes the lognormal prior distribution.

We specify the prior distribution $\pi_0(\boldsymbol{\beta}_\mu, \phi, \boldsymbol{\beta}_\omega)$ so that, even with moderate sample sizes, the information from the data outweighs the non-informative prior due to the “vague” nature of the prior knowledge.

2.6. Posterior Distribution

The likelihood function associated with the observations \mathcal{D} can be written in terms of $\ell_{ZTPS}(\boldsymbol{\beta}_\mu, \phi | \mathcal{D}^+)$ and $\ell_2(\boldsymbol{\beta}_\omega | \mathcal{D})$ as

$$\mathcal{L}(\boldsymbol{\beta}_\mu, \phi, \boldsymbol{\beta}_\omega | \mathcal{D}) = \exp\{\ell_{ZTPS}(\boldsymbol{\beta}_\mu, \phi | \mathcal{D}^+) + \ell_2(\boldsymbol{\beta}_\omega | \mathcal{D})\}.$$

The joint posterior distribution is

$$\pi(\beta_\mu, \phi, \beta_\omega | \mathcal{D}) \propto \exp\{\ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+) + \ell_2(\beta_\omega | \mathcal{D})\} \pi_0(\beta_\mu) \pi_0(\phi) \pi_0(\beta_\omega). \quad (11)$$

The Inference for each parameter is derived from its marginal posterior distribution, which can be obtained by integrating $\pi(\beta_\mu, \phi, \beta_\omega | \mathcal{D})$ for each parameter. For the parameters (β_μ, ϕ) , the joint posterior distribution is given by

$$\pi_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+) \propto \exp\{\ell_{ZTPS}(\beta_\mu, \phi | \mathcal{D}^+)\} \pi_0(\beta_\mu) \pi_0(\phi). \quad (12)$$

For the parameter β_ω , the posterior distribution is

$$\pi_\omega(\beta_\omega | \mathcal{D}) \propto \exp\{\ell_2(\beta_\omega | \mathcal{D})\} \pi_0(\beta_\omega).$$

Under quadratic loss, we calculate the expected value and variance of the posterior densities of these parameters. Since the posterior densities $\pi(\beta_\mu, \phi | \mathcal{D})$ and $\pi(\beta_\omega | \mathcal{D})$ do not have a standard analytic form, MCMC methods are useful. We employ the highly efficient stochastic gradient HMC (SGHMC) algorithm [31]. While Bayesian estimation based on SGHMC is similar in many aspects to a traditional MCMC approach, the SGHMC algorithm enhances sampling efficiency by eliminating the acceptance–rejection step typically found in other algorithms, such as MH and HMC. This streamlining allows for faster convergence and more effective sampling.

2.7. Stochastic Gradient HMC with Friction (SGHMC) Algorithm

This section reviews the SGHMC algorithm. Let $\mathcal{L}(\theta | \mathcal{D})$ and $\pi_0(\theta)$ respectively denote the likelihood function and prior distribution. To develop the HMC algorithm, the posterior distribution of $\theta = (\theta_1, \dots, \theta_m)'$ given a set of independent observations \mathcal{D} can be written as

$$\pi(\theta | \mathcal{D}) = \exp\{-E(\theta)\},$$

where $E(\theta)$ is a potential energy function given by

$$E(\theta) = -\log \mathcal{L}(\theta | \mathcal{D}) - \log \pi_0(\theta).$$

The HMC method generates samples from the joint distribution of (θ, z) defined by

$$\pi(\theta, z | \mathcal{D}) \propto \exp\{-E(\theta) + K(z)\}.$$

where z is an auxiliary latent random variable of the same dimension as θ , with a Gaussian distribution whose kernel $K(z)$ represents a kinetic energy function given by

$$K(z) = \frac{z' M^{-1} z}{2},$$

and the mass matrix M is often set to the identity matrix I .

Making an analogy with a physical system, the parameter vector θ denotes the generalized coordinate, and z is the generalized moment. The Hamiltonian function $H(\theta, z) = E(\theta) + K(z)$ represents the energy of the system, which is the sum of the potential energy $E(\theta)$ and kinetic energy $K(z)$.

Suppose τ refers to the iteration of the algorithm. The partial derivatives of the Hamiltonian determine how θ and z change over τ , according to Hamilton’s equations, i.e.,

$$\frac{d\theta}{d\tau} = \frac{\partial H(\theta, z)}{\partial z}, \tag{13}$$

$$\frac{dz}{d\tau} = -\frac{\partial H(\theta, z)}{\partial \theta}. \tag{14}$$

For any time interval of duration $\Delta\tau$, these equations define a transition from the state at time τ to the state at time $\tau + \Delta\tau$. The HMC algorithm simulates the Hamiltonian dynamics according to (13) and (14) for obtaining a sequence of random samples.

Ref. [28] modified Hamilton’s equations by adding a *friction* term to the momentum update, leading to the equations

$$\begin{aligned} d\theta &= M^{-1}z d\tau, \\ dz &= -\nabla E(\theta) d\tau - \mathbf{B}M^{-1}z d\tau + \mathcal{N}(\mathbf{0}, 2\mathbf{B}d\tau), \end{aligned}$$

where \mathbf{B} is a diffusion matrix, and with a slight abuse of notation, the last term in the second equation denotes the introduction of a multivariate Gaussian random variable $\mathcal{N}(\mathbf{0}, 2\mathbf{B}d\tau)$. In practice, we need an estimate $\hat{\mathbf{B}}$. Ref. [28] introduced a specified friction term $\mathbf{C} \succ \mathbf{B}$ and considered the following dynamics:

$$d\theta = M^{-1}z d\tau \tag{15}$$

$$dz = -\nabla E(\theta) d\tau - \mathbf{C}M^{-1}z d\tau + \mathcal{N}(\mathbf{0}, 2(\mathbf{C} - \hat{\mathbf{B}})d\tau) + \mathcal{N}(\mathbf{0}, 2\hat{\mathbf{B}}d\tau). \tag{16}$$

Under a realistic, simple choice $\hat{\mathbf{B}} = \mathbf{0}$, the momentum update simplifies to

$$dz = -\nabla E(\theta) d\tau - \mathbf{C}M^{-1}z d\tau + \mathcal{N}(\mathbf{0}, 2\mathbf{C}d\tau).$$

Therefore, the dynamics are governed by the controllable injected noise $\mathcal{N}(\mathbf{0}, 2\mathbf{C}d\tau)$ and the friction term $\mathbf{C}M^{-1}$. In practical applications, the differential Equations (15) and (16) cannot be solved analytically, and numerical methods are required. Considering the integration step size ϵ , we rewrite (15) and (16) in discretized form as

$$\Delta\theta = \epsilon M^{-1}z \tag{17}$$

$$\Delta z = -\epsilon \nabla E(\theta) - \alpha z + \mathcal{N}(\mathbf{0}, 2\alpha \mathbf{M}), \tag{18}$$

where $\alpha = \epsilon \mathbf{C}M^{-1}$. The constant α and the discretization step ϵ are tuning constants; α tends to be fixed at a small value in practice. In our implementation, we fixed $\alpha = 0.10$ and $\epsilon = 0.01$. To solve (17) and (18), we consider the *leapfrog* integrator method summarized in the following steps (Algorithm 1):

It should be noted that the SGHMC does not need a Metropolis–Hastings (MH) step to reach the target distribution, which avoids wasting computational processing [28]. However, SGHMC’s performance is highly sensitive to two user-specified parameters: a step size ϵ and a desired number of steps L . In particular, if L is too small, the algorithm exhibits undesirable random walk behavior, while if L is too large, the algorithm wastes computation time. We chose $\epsilon = 0.01$ and $L = 10$ to guarantee the good performance of the algorithm for the problems addressed in this paper.

Algorithm 1 Simulation of the discretized SGHMC dynamics with friction.

Starting position: $\theta^{(t)}$, **step size** ϵ , **number of steps** L , **constant** α , and **sample momentum** $z^{(t)} \sim N(\mathbf{0}, \mathbf{M})$.

Initialize:

$$\begin{aligned} \theta^{(0)} &= \theta^{(t)}, \\ z^{(0)} &= (1 - \alpha)z^{(t)} - \frac{\epsilon}{2}\nabla E(\theta^{(t)}) + \mathcal{N}(\mathbf{0}, 2\alpha\mathbf{M}). \end{aligned}$$

for $\tau \leftarrow 1$ **to** L **do**

$$\begin{aligned} \theta^{(\tau)} &= \theta^{(\tau-1)} + \epsilon\mathbf{M}^{-1}z^{(\tau-1)} \\ \text{if } \tau \neq L \quad z^{(\tau)} &= (1 - \alpha)z^{(\tau-1)} - \epsilon\nabla E(\theta^{(\tau)}) + \mathcal{N}(\mathbf{0}, 2\alpha\mathbf{M}) \end{aligned}$$

end for

$$z^{(L)} = (1 - \alpha)z^{(L-1)} - \frac{\epsilon}{2}\nabla E(\theta^{(L)}) + \mathcal{N}(\mathbf{0}, 2\alpha\mathbf{M})$$

return $(\theta^{(t+1)}, z^{(t+1)}) \leftarrow (\theta^{(L)}, z^{(L)})$.

2.8. Monte Carlo Posterior Estimates

We employ the SGHMC algorithm for the posterior sampling of the parameters $\theta = (\beta'_\mu, \beta'_\omega, \phi)'$ from the ZMPS regression. Suppose we denote these by $\theta = (\theta_1, \dots, \theta_m)$. Let $\{\theta^{(j)}, j = 1, \dots, G\}$, denote the θ samples generated from the posterior distribution $\pi(\theta|\mathcal{D})$ via the SGHMC procedure. Since we can approximate a function $E(\psi(\theta)|\mathcal{D})$ by

$$E(\psi(\theta)|\mathcal{D}) \approx \frac{1}{G} \sum_{g=1}^G \psi(\theta^{(g)}),$$

in particular, the Monte Carlo estimates for the posterior means and variances of θ_j , $j = 1, \dots, m$ are computed as

$$E(\theta_j|\mathcal{D}) \approx \hat{\theta}_j = \frac{1}{G} \sum_{g=1}^G \theta_j^{(g)}, \tag{19}$$

$$V(\theta_j|\mathcal{D}) \approx \hat{\sigma}_{\theta_j}^2 = \frac{1}{G} \sum_{g=1}^G (\theta_j^{(g)} - \hat{\theta}_j)^2. \tag{20}$$

Bayesian models can be evaluated and compared using various criteria; see [38] for a comprehensive overview. In this study, we employ three information criteria to select the best-fitting ZMPS regression model: the Conditional Predictive Ordinate (CPO) [39], the Expected Bayesian Information Criterion (EBIC) [40], and the Watanabe–Akaike Information Criterion (WAIC) [41]. All of these criteria can be estimated using Monte Carlo output; details are provided in Appendix C.

2.9. Posterior Information and Relative Efficiency

We use Bayesian information to quantify the efficiency loss from fitting a ZMPS model when a PS distribution could have been used. Ref. [42] originally introduced the concept of posterior information based on Fisher’s information matrix. In this work, we adapt this approach by using the observed information matrix, as the expected information matrix is intractable due to the model’s complexity.

2.10. Posterior Information

The information about the parameters $(\beta_\mu, \beta_\omega, \phi)$ contained in the joint distribution of \mathbf{y} and $(\beta_\mu, \beta_\omega, \phi)$ is written as

$$\mathcal{I} = E_{\pi_0(\beta_\mu, \beta_\omega, \phi)} \left\{ E_{\pi_{ZMPS}(\mathbf{y}|\beta_\mu, \beta_\omega, \phi, \mathbf{x})} \left[\mathcal{H}(\beta_\mu, \beta_\omega, \phi, \mathcal{D}) \right] \right\}, \tag{21}$$

where $\mathcal{H}(\beta_\mu, \beta_\omega, \phi, \mathcal{D})$ is the observed Bayesian information matrix, given by

$$\mathcal{H}(\beta_\mu, \beta_\omega, \phi, \mathcal{D}) = - \begin{bmatrix} \frac{\partial^2 \log \pi(\beta_\mu, \beta_\omega, \phi|\mathcal{D})}{\partial \beta_\mu \partial \beta_\mu'} & \frac{\partial^2 \log \pi(\beta_\mu, \beta_\omega, \phi|\mathcal{D})}{\partial \beta_\mu \partial \phi} & 0 \\ \frac{\partial^2 \log \pi(\beta_\mu, \beta_\omega, \phi|\mathcal{D})}{\partial \phi \partial \beta_\mu} & \frac{\partial^2 \log \pi(\beta_\mu, \beta_\omega, \phi|\mathcal{D})}{\partial \phi^2} & 0 \\ 0 & 0 & \frac{\partial^2 \log \pi(\beta_\mu, \beta_\omega, \phi|\mathcal{D})}{\partial \beta_\omega \partial \beta_\omega'} \end{bmatrix}. \tag{22}$$

Calculating the expected value in (21) is not always possible, so we consider the observed information matrix $\mathcal{H}(\beta_\mu, \beta_\omega, \phi, \mathcal{D})$ in this paper. From (11), we can write

$$\log \pi(\beta_\mu, \beta_\omega, \phi|\mathcal{D}) = \ell_{ZTPS}(\beta_\mu, \phi|\mathcal{D}^+) + \ell_2(\beta_\omega|\mathcal{D}) + \log \pi_0(\beta_\mu, \beta_\omega, \phi) + \log(C), \tag{23}$$

where $C > 0$ is a normalizing constant. Replacing (23) in (22), we can write the block of the observed Bayesian information matrix for β_μ and ϕ as

$$\mathcal{I}(\beta_\mu, \phi) = \mathcal{J}(\beta_\mu, \phi) + \mathbf{\Pi}(\beta_\mu, \phi),$$

where $\mathcal{J}(\beta_\mu, \phi)$ is the block of the observed information matrix (10), and $\mathbf{\Pi}(\beta_\mu, \phi)$ is the negative Hessian of the log-prior denoted by

$$\mathbf{\Pi}(\beta_\mu, \phi) = \begin{bmatrix} \mathbf{\Pi}_{\beta_\mu \beta_\mu} & \mathbf{\Pi}_{\beta_\mu \phi} \\ \mathbf{\Pi}_{\phi \beta_\mu} & \mathbf{\Pi}_{\phi \phi} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 \log \pi_0(\beta_\mu, \phi)}{\partial \beta_\mu \partial \beta_\mu'} & \frac{\partial^2 \log \pi_0(\beta_\mu, \phi)}{\partial \beta_\mu \partial \phi} \\ \frac{\partial^2 \log \pi_0(\phi, \beta_\mu)}{\partial \phi \partial \beta_\mu'} & \frac{\partial^2 \log \pi_0(\beta_\mu, \phi)}{\partial \phi^2} \end{bmatrix}.$$

The Hessian matrix $\mathbf{\Pi}(\beta_\mu, \phi)$ represents the amount of information about (β_μ, ϕ) contained in the proper prior distribution. Obviously $\mathbf{\Pi}(\beta_\mu, \phi) = \mathbf{0}$ only when the prior is uniform [42]. For simplicity, let us consider the notation.

$$\mathcal{I}(\beta_\mu, \phi) = \begin{bmatrix} \mathcal{I}_{\beta_\mu \beta_\mu} & \mathbf{b}_{\beta_\mu \phi} \\ \mathbf{b}'_{\phi \beta_\mu} & c_{\phi \phi} \end{bmatrix} = \begin{bmatrix} \mathcal{J}_{\beta_\mu \beta_\mu} + \mathbf{\Pi}_{\beta_\mu \beta_\mu} & \mathcal{J}_{\beta_\mu \phi} + \mathbf{\Pi}_{\beta_\mu \phi} \\ \left(\mathcal{J}_{\phi \beta_\mu} + \mathbf{\Pi}_{\phi \beta_\mu} \right)' & \mathcal{J}_{\phi \phi} + \mathbf{\Pi}_{\phi \phi} \end{bmatrix}, \tag{24}$$

where $\mathcal{I}_{\beta_\mu \beta_\mu}$ is a matrix $p \times p$, $\mathbf{b}_{\beta_\mu \phi}$ is a vector $p \times 1$, and $c_{\phi \phi}$ is a scalar.

Equation (24) defines the amount of observed information in the posterior distribution (21). We denote the inverse of the posterior observed information matrix $\mathcal{V} = \mathcal{I}^{-1}$ given by

$$\mathcal{V} = \begin{bmatrix} \mathcal{V}_{\beta_\mu \beta_\mu} & \mathcal{V}_{\beta_\mu \phi} \\ \mathcal{V}_{\phi \beta_\mu} & \mathcal{V}_{\phi \phi} \end{bmatrix} = \begin{bmatrix} \left(\mathcal{I}_{\beta_\mu \beta_\mu} - \frac{1}{c_{\phi \phi}} \mathbf{b}_{\beta_\mu \phi} \mathbf{b}'_{\beta_\mu \phi} \right)^{-1} & -\frac{1}{k_{\phi \phi}} \mathcal{I}_{\beta_\mu \beta_\mu}^{-1} \mathbf{b}_{\beta_\mu \phi} \\ -\frac{1}{k_{\phi \phi}} \mathbf{b}'_{\beta_\mu \phi} \mathcal{I}_{\beta_\mu \beta_\mu}^{-1} & \frac{1}{k_{\phi \phi}} \end{bmatrix}, \tag{25}$$

where $k_{\phi \phi} = c_{\phi \phi} - \mathbf{b}'_{\beta_\mu \phi} \mathcal{I}_{\beta_\mu \beta_\mu}^{-1} \mathbf{b}_{\beta_\mu \phi}$. The matrix \mathcal{V} in (25) allows us to compute the relative efficiency between the parameter estimators of the ZMPS regression model and those obtained from the PS regression model when $p = 1$.

Additionally, ref. [43] introduced a related inequality, demonstrating that the variance of the posterior distribution is bounded. Specifically, $E(V(\theta_j | \mathcal{D})) \geq E(V_{j,j} | \mathcal{D})$. This inequality will be employed in a simulation study to assess the efficiency of Bayesian estimators derived using the SGHMC algorithm.

2.11. Bayesian Relative Efficiency Between the ZMPS and PS Models

When fitting a ZMPS model to a dataset, we can assess whether a standard PS model associated with the ZMPS model is more appropriate by testing the hypothesis $H_0 : p = 1$. Tests such as FBST [44] can be used to assess the evidence in favor of H_0 . However, Result 1 ensures that the parameters $\theta = (\mu, \phi)$ can be estimated considering only the positive observations \mathbf{y}^+ present in the dataset and by fitting the ZTPS model associated with the PS model. Using only \mathbf{y}^+ with the ZTPS distribution can result in a loss of efficiency in the estimates $\hat{\theta} = (\hat{\mu}, \hat{\phi})$ when the true distribution of the data is a standard PS distribution. To evaluate this loss of efficiency, we consider the Laplace approximation for the posterior density $\pi_k(\theta_k | \mathcal{D}^{(k)})$, where $k = 0$ refers to the posterior density $\pi_{PS}(\theta_0 | \mathcal{D})$ and $k = 1$ is $\pi_{ZTPS}(\theta_1 | \mathcal{D}^+)$. Now, suppose we can apply the Laplace method; ref. [45] proves that under some conditions on $\pi_k(\theta_k | \mathcal{D}^{(k)})$ and for sufficiently large n , the posterior distribution can be approximated by a multivariate normal distribution $N(\theta_k^*, \mathcal{V}^{(k)}(\theta_k^*))$, where θ_k^* denotes the mode of the a posteriori distribution, i.e., the value of θ_k at which $\log \pi_k(\theta_k | \mathcal{D}^{(k)})$ achieves its maximum, such that

$$\left. \frac{\partial \log \pi_k(\theta_k | \mathcal{D}^{(k)})}{\partial \theta_k} \right|_{\theta_k = \theta_k^*} = 0$$

and $\mathcal{V}^{(k)}(\theta_k^*)$ is the inverse of the posterior observed information matrix given by (25), evaluated at the mode θ_k^* .

For ZTPS and most PS models, estimates of θ_k^* can only be obtained numerically. To calculate θ_k^* , we consider the Newton–Raphson iterative method, such that

$$\theta_k^{*(j+1)} = \theta_k^{*(j)} + \mathcal{V}(\theta_k^{*(j)}) \nabla \log \pi_k(\theta_k^{*(j)} | \mathcal{D}^{(k)}), \tag{26}$$

where $\nabla \log \pi_k(\theta_k | \mathcal{D}^{(k)}) = \partial \log \pi_k(\theta_k | \mathcal{D}^{(k)}) / \partial \theta_k$. As an initial condition for (26), we consider the Monte Carlo estimate of the $\hat{\theta}_k$'s components by finding the componentwise sample mean generated from the posterior distribution via MCMC. This initial condition is reasonable since θ_k^* and $\hat{\theta}_k$ should be close.

Let the parameter vector $\theta = (\beta'_\mu, \phi)'$, and we denote these by $\theta = (\theta_1, \dots, \theta_d)'$, which is the parameter vector in the ZTPS model, and two estimates of this parameter vector denoted by $\theta_0 = (\theta_{0,1}, \dots, \theta_{0,d})$ and $\theta_1 = (\theta_{1,1}, \dots, \theta_{1,d})$ are obtained, respectively, with the PS and ZTPS models. To evaluate the loss of efficiency caused by using Result 1, we propose to use a Bayesian relative efficiency (BRE) measure, which is defined by

$$\text{BRE}(\hat{\theta}_{0,i}, \hat{\theta}_{1,i}) = \frac{\mathcal{V}_{i,i}^{(0)}}{\mathcal{V}_{i,i}^{(1)}}. \tag{27}$$

Appendix A presents the computation of the loss of efficiency in estimating μ and ϕ assessed by (27) under two scenarios in which we fit the proposed ZMPS (or ZTPS) models when the observations come from a PS distribution.

3. Results

3.1. Simulation Study

We present a simulation study to evaluate Bayesian estimation of the ZMPS class of models using the SGHMC algorithm. In this study, we considered samples $\mathbf{y} = (y_1, \dots, y_n)$ from the distributions ZMP (Poisson), ZMNB (negative binomial), and ZMGP (generalized Poisson) models, with sample size $n = 100$, under the link functions

$$\begin{aligned} \log(\mu_i) &= \beta_{\mu,1} + \beta_{\mu,2}x_i, \\ \text{logit}(\omega_i) &= \beta_{\omega,1} + \beta_{\omega,2}x_i, \end{aligned}$$

where $x_i \sim U(0, 1)$, $i = 1, \dots, n$. We consider independent prior distributions, normal $N(\beta_{\mu 0}, \sigma_0^2)$ with $\beta_{\mu 0} = 0$ and $\sigma_0^2 = 1$ for informative prior and $\sigma_0^2 = 100$ for vague prior, and a gamma distribution $G(a_\phi, b_\phi)$ for ϕ . Let $\theta = (\beta_{\mu,1}, \beta_{\mu,2}, \beta_{\omega,1}, \beta_{\omega,2}, \phi)$.

In this simulation study, we conducted a sensitivity analysis to evaluate the effect of initial conditions on the convergence of the generated chains. We tested three different chains, each initialized with conditions based on the mean of the prior, (i) 1.4 times the mean (i.e., plus 40%), and (ii) 0.4 times the mean (i.e., minus 60%). After verifying the burn-in period and ensuring the method’s convergence, we selected the mean of the prior distribution as the reference point for subsequent analysis.

We employed the SGHMC algorithm to generate 50,000 samples of θ from its full posterior distribution, discarding the first 50% as burn-in. This resulted in a chain consisting of 25,000 samples of θ . The final sample was used to calculate the Monte Carlo estimates of each model parameter’s posterior mean and variance, as given by (19) and (20). We repeated this procedure $M = 100$ times. With M estimates of the model parameters, we evaluated their performance using metrics such as mean bias (B), the ratio of mean squared error (MSE) to variance (Var), and the average Bayesian efficiency (BE), calculated for each model parameter θ_j :

$$B(\hat{\theta}_j) = \frac{1}{M} \sum_{k=1}^M (\hat{\theta}_j^{(k)} - \theta_j), \tag{28}$$

$$MSE(\hat{\theta}_j) = \frac{1}{M} \sum_{k=1}^M (\hat{\theta}_j^{(k)} - \theta_j)^2, \tag{29}$$

$$BE(\hat{\theta}_j) = \frac{1}{M} \sum_{k=1}^M \frac{\mathcal{V}_{jj}^{(k)}}{\hat{\sigma}_{\theta_j}^{2(k)}}, \tag{30}$$

where $\mathcal{V}_{jj}^{(k)}$ is given by (25) and $\hat{\sigma}_{\theta_j}^{2(k)}$ is given by (20) for each $\hat{\theta}^{(k)}$, $k = 1, \dots, L$. Additionally, we have also computed the coverage probability (CP) of the Bayesian credible intervals expressed as percentages as

$$CP(\theta_j) = \frac{100}{M} \sum_{k=1}^M \delta_\theta^{(k)}, \tag{31}$$

where $\delta_\theta^{(k)}$ assumes 1 if the k th Bayesian credible interval (BCI) contains the true value θ_j and 0 otherwise.

A summary of the $M = 100$ simulations is shown in Table 2. In this summary, we can see the effect of the prior information on the calculation of the standard deviation (s.d.) and the amplitude of the highest posterior density (HPD) credibility interval for each parameter.

Table 2. Simulation study. Estimation for the ZMPS(μ, ϕ, p) distribution using sample size $N = 100$.

ZMPS	Parameters	N(0,1)			N(0,100)		
		θ	$\hat{\theta}$	s.d.	HPD (95%)	$\hat{\theta}$	s.d.
ZMP	$\beta_{\mu,1} = -1.0$	-1.061	0.306	(-1.573, -0.427)	-1.111	0.426	(-1.863, -0.299)
	$\beta_{\mu,2} = 3.0$	3.078	0.371	(2.341, 3.678)	3.135	0.511	(2.185, 4.025)
	$\beta_{\omega,1} = -1.0$	-1.025	0.312	(-1.604, -0.424)	-1.076	0.465	(-1.883, -0.212)
	$\beta_{\omega,2} = 2.5$	2.582	0.446	(1.915, 3.393)	2.686	0.768	(1.654, 4.262)
ZMNB	$\beta_{\mu,1} = -1.0$	-1.024	0.298	(-1.560, -0.380)	-1.102	0.559	(-2.261, -0.129)
	$\beta_{\mu,2} = 3.0$	3.035	0.403	(2.236, 3.811)	3.127	0.729	(1.668, 4.549)
	$\beta_{\omega,1} = -1.5$	-1.485	0.296	(-2.095, -0.883)	-1.507	0.476	(-2.443, -0.552)
	$\beta_{\omega,2} = 2.0$	1.967	0.459	(1.146, 2.889)	1.993	0.788	(0.661, 3.666)
	$\phi = 3.0$	2.954	0.266	(2.423, 3.388)	2.939	0.265	(2.423, 3.398)
ZMGP	$\beta_{\mu,1} = -1.0$	-1.022	0.300	(-1.571, -0.412)	-1.124	0.605	(-2.377, -0.057)
	$\beta_{\mu,2} = 3.0$	3.053	0.419	(2.212, 3.829)	3.186	0.816	(1.519, 4.757)
	$\beta_{\omega,1} = -1.5$	-1.485	0.296	(-2.095, -0.883)	-1.507	0.476	(-2.443, -0.551)
	$\beta_{\omega,2} = 2.0$	1.967	0.459	(1.146, 2.889)	1.993	0.788	(0.661, 3.666)
	$\phi = 0.2$	0.206	0.027	(0.153, 0.256)	0.208	0.027	(0.156, 0.256)
ZMCOMP	$\beta_{\mu,1} = -3.5$	-3.500	0.011	(-3.523, -3.478)	-3.523	0.098	(-3.716, -3.334)
	$\beta_{\mu,2} = 2.0$	1.999	0.012	(1.976, 2.022)	1.986	0.099	(1.785, 2.173)
	$\beta_{\omega,1} = -1.0$	-0.999	0.011	(-1.022, -0.976)	-0.998	0.090	(-1.103, -0.747)
	$\beta_{\omega,2} = 0.5$	0.500	0.011	(0.478, 0.523)	0.499	0.096	(0.345, 0.726)
	$\phi = 0.2$	0.198	0.027	(0.125, 0.230)	0.196	0.027	(0.124, 0.231)

As previously discussed, the performance of the estimators using the metrics in (28)–(31) is presented in Table 3. The results show improved efficiency of the estimators when the prior distribution is more informative.

Table 3. Simulation study. Results of the Bayesian efficiency for the ZMPS(μ, ϕ, p) distribution.

ZMPS	θ	N(0,1)					N(0,100)				
		B	MSE	PC (%)	$\frac{MSE}{Var}$	BE	B	MSE	CP (%)	$\frac{MSE}{Var}$	BE
ZMP	$\beta_{\mu,1}$	-0.061	0.097	97	1.039	1.00	-0.111	0.193	91	1.068	1.00
	$\beta_{\mu,2}$	0.078	0.144	99	1.044	1.00	0.135	0.279	89	1.069	1.00
	$\beta_{\omega,1}$	-0.025	0.098	98	1.006	1.00	-0.076	0.221	94	1.027	1.00
	$\beta_{\omega,2}$	0.082	0.205	99	1.033	1.00	0.186	0.625	94	1.059	1.00
ZMNB	$\beta_{\mu,1}$	-0.024	0.089	99	1.006	1.00	-0.102	0.323	94	1.033	1.00
	$\beta_{\mu,2}$	0.035	0.164	98	1.008	1.00	0.127	0.548	94	1.030	1.00
	$\beta_{\omega,1}$	0.014	0.088	97	1.002	1.00	-0.007	0.226	93	1.005	1.00
	$\beta_{\omega,2}$	-0.033	0.212	98	1.005	1.00	-0.007	0.621	94	1.007	1.00
	ϕ	-0.046	0.073	99	1.029	1.00	-0.061	0.074	99	1.052	1.00
ZMGP	$\beta_{\mu,1}$	-0.022	0.091	99	1.005	1.00	-0.124	0.382	93	1.042	0.97
	$\beta_{\mu,2}$	0.053	0.178	98	1.016	1.00	0.186	0.700	93	1.052	0.94
	$\beta_{\omega,1}$	0.014	0.088	97	1.002	1.00	-0.007	0.226	93	1.000	1.00
	$\beta_{\omega,2}$	-0.033	0.212	98	1.005	1.00	-0.007	0.621	94	1.000	1.00
	ϕ	0.006	0.001	99	1.052	0.83	0.008	0.001	99	1.083	0.82
ZMCP	$\beta_{\mu,1}$	0.002	0.001	99	1.004	1.00	0.023	0.010	99	1.056	0.76
	$\beta_{\mu,2}$	0.001	0.001	99	1.001	1.00	0.014	0.010	99	1.019	0.74
	$\beta_{\omega,1}$	0.003	0.001	99	1.000	0.93	0.002	0.008	99	1.000	0.74
	$\beta_{\omega,2}$	0.002	0.001	99	1.000	0.96	0.002	0.009	99	1.000	0.75
	ϕ	0.021	0.011	98	1.039	0.41	0.003	0.015	98	1.094	0.42

3.2. Application: Number of Femicide Cases

Violence against women encompasses a wide range of physical, psychological, sexual, and property-related attacks, often occurring on a continuum that can tragically culminate in murder, the most extreme manifestation of violence inflicted upon women.

Femicide refers to the murder of a woman specifically because she is a woman, typically driven by hatred, contempt, or a sense of lost control or ownership over women. Brazil’s Femicide Law (Law 13.104 of 9 March 2015) classifies homicide as femicide when committed against a woman due to her gender. The law defines such crimes as those

involving domestic or family violence, contempt, or discrimination against women and also includes femicide in the category of heinous crimes.

Of the five measures to combat gender-based violence recommended by the United Nations, Brazil has implemented only one: online services. This includes the “Red Light” campaign launched by the National Council of Justice and the Brazilian Association of Magistrates, which functions similarly to emergency alert systems. However, there has yet to be an assessment of the campaign’s impact on protecting women in violent situations. Factors contributing to the increase in gender-based violence include victims’ difficulty in reporting, a reduction in crime reporting at police stations, and a decline in the number of emergency protective measures granted to women.

This study aims to determine whether femicides increased during the COVID-19 pandemic. We analyze data from 642 municipalities in São Paulo for the years 2019 and 2020 [46], fitting ZMPS regression models with the Municipal Human Development Index (MHDI) as an explanatory variable. The MHDI, a key measure of municipal development in Brazil, evaluates three dimensions of human development: longevity, education, and income. It ranges from 0 to 1, with higher values indicating greater development. By interpreting the mean μ_i^{ZMPS} of the model and the parameter p_i (which accounts for zero inflation or deflation) as functions of the MHDI, we gain insights into femicide patterns. A summary of the sample is presented in Table 4.

Table 4. Frequency distribution and descriptive statistics of each sample referring to the number of femicides in 2019 and 2020 in 642 São Paulo municipalities.

Year		Frequency															
2019	y_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	f_i	254	132	56	45	27	23	16	13	7	7	3	1	5	2	2	4
2020	y_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	f_i	262	121	67	40	36	17	11	10	8	6	4	3	2	3	5	-
2019	y_i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	>30
	f_i	2	3	2	3	3	2	3	2	-	2	3	1	2	1	1	15
2020	y_i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	>30
	f_i	1	2	3	4	2	1	2	3	2	1	1	1	3	1	3	17

We apply the proposed Bayesian approach to fit ZMPS models to the femicide data, using the MHDI as a regressor. Specifically, we consider the zero-modified Poisson (ZMP), zero-modified negative binomial (ZMNB), zero-modified generalized Poisson (ZMGP), and zero-modified COM-Poisson (ZMCOMP) distributions. For the parameters $\beta\mu$ and βw , we use normal priors $N(0, 10^2)$, while for ϕ , we apply a gamma prior $G(a_\phi, b_\phi)$ with $a_\phi/b_\phi = 5$ and $a_\phi/b_\phi^2 = 10^2$. We employ three information criteria, LogCPO (A8), EBIC (A9), and WAIC (A10), to identify the distribution that best fits the data, with higher LogCPO values and lower WAIC and EBIC values indicating better model performance. The values of these criteria for each fitted model are presented in Table 5, with the selected model for each sample highlighted in bold.

Table 5. Criteria for selecting the models that were fitted to the data on the number of femicide cases in 2019 and 2020.

Year	2019				2020			
	Criteria	ZMP	ZMNB	ZMGP	ZMCOMP	ZMP	ZMNB	ZMGP
LogCPO	2296	1376	1372	1481	2392	1372	1368	1492
WAIC	4658	2758	2767	2972	4866	2750	2750	3032
EBIC	4648	2783	2791	3000	4847	2775	2783	3044

The three criteria identify the ZMNB model as the best fit for both 2019 and 2020. Posterior summaries and credible intervals for the parameters μ_i , ϕ , and p_i are presented in Table 6. The results are derived using three Markov chains, each containing 100,000 samples,

initialized with distinct starting conditions. To minimize the effect of the burn-in period, we discard 25% of each chain and select every 75th value from the remaining 75% (thinning), which helps reduce autocorrelation within the chains. We use the resulting final sample of 3000 observations to compute the Bayesian Monte Carlo estimates. The convergence details of the SGHMC algorithm are provided in Appendix B. Figure 1 displays the estimated mean $\mu_i^{(ZMPS)}$ and the parameter p_i for the ZMNB models based on data from 2019 and 2020.

Table 6. Posterior summaries and credible intervals of the parameters of ZMNB fitted to the number of femicide cases in 2019 and 2020.

Year	2019				2020			
	Parameter	Mean	Median	s.d.	CI (95%)	Mean	Median	s.d.
$\beta_{\mu,0}$	-15.665	-15.594	1.579	(-18.883, -12.757)	-17.597	-17.528	1.592	(-20.839, -14.647)
$\beta_{\mu,1}$	22.629	22.558	2.098	(18.497, 26.587)	25.293	25.226	2.117	(21.200, 29.377)
$\beta_{\omega,0}$	-10.441	-10.443	1.808	(-14.178, -7.086)	-7.602	-7.613	1.751	(-11.150, -4.259)
$\beta_{\omega,1}$	14.724	14.692	2.454	(10.032, 19.653)	10.798	10.798	2.375	(6.339, 15.658)
ϕ	0.556	0.554	0.085	(0.396, 0.729)	0.574	0.572	0.086	(0.403, 0.741)

In Figure 1a, we examine the mean number of reported femicides in relation to the MHDI. The graphs demonstrate that the mean number of notifications increases with the MHDI and is slightly higher during the COVID-19 pandemic. This trend indicates that municipalities with higher levels of development tend to report a greater mean number of femicides. This pattern is anticipated as more developed municipalities are likely to offer better protective measures for women and more effective crime reporting systems. The increase in reported cases in 2020, compared with the averages in 2019, can be attributed to the impacts of the pandemic.

The p_i parameters are depicted as a function of the MHDI in Figure 1b. These parameters help interpret the frequency of zero observations within the sample. Typically, municipalities with lower mean notification numbers, μ_i , are expected to exhibit zero inflation, meaning $p_i \leq 1$. However, Figure 1b shows that for municipalities with an MHDI below 0.7 (i.e., those with low average μ_i), the estimates of p_i exceed 1. This indicates a deflation of zero observations at this level of the covariate, suggesting that municipalities with a low MHDI experience less zero inflation than expected. In fact, they report more cases than the standard model would predict ($p_i = 1$). Furthermore, the increase in registrations during 2020, marked by a rise in deflation from zero, can be attributed to the impact of the COVID-19 pandemic, as seen in the comparison of p_i estimates between 2019 and 2020.

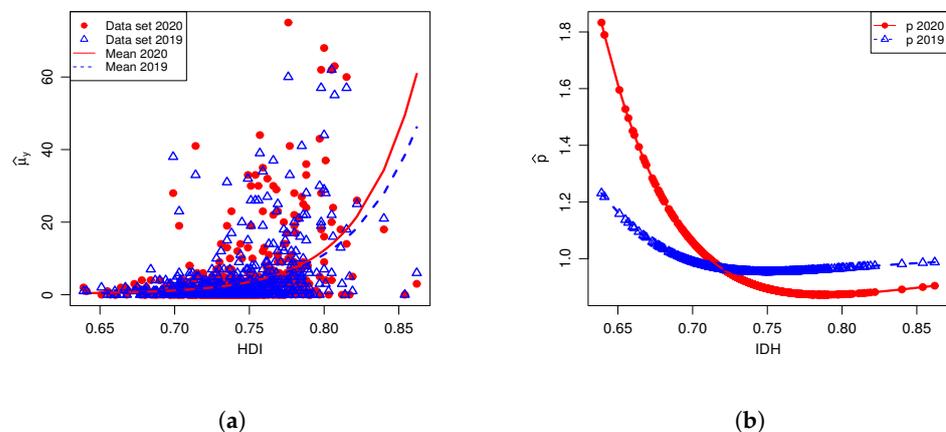


Figure 1. The mean μ_{ZMNB} (a) and the parameters p (b) fitted according to the MHDI for the 642 municipalities of São Paulo in 2019 and 2020.

4. Conclusions

The ZMPS distribution is constructed by modifying the zero probability of standard PS distributions. This flexible model accommodates count data with varying characteristics related to the frequency of zeros, including zero inflation and deflation, as well as different types of dispersion, such as overdispersion, equidispersion, and underdispersion. This property is particularly significant for ZMPS regression models, where zero inflation or deflation can manifest at different levels of covariates. As a result, the ZMPS class can be applied to model count datasets without requiring prior knowledge of zero inflation or deflation. Importantly, the ZMPS distribution includes PS distributions as a specific case.

This paper presents a Bayesian approach using the SGHMC algorithm to fit a ZMPS regression model. We demonstrate that estimates of μ_i and ϕ can be obtained by focusing exclusively on the positive observations in a dataset and assuming a ZTPS distribution for these data. Based on this approach, we can evaluate relative efficiency losses by considering a measure of Bayesian information across different levels of prior information.

The estimation of the parameter p_i enables the characterization of zero inflation ($0 < p_i < 1$) or zero deflation ($1 < p_i < \frac{1}{1 - \pi_{PS}(0; \mu_i, \phi)}$). This parameter plays a key role in analyzing and interpreting the data, as discussed in Section 3.2. The ability to estimate p_i is a significant strength of the ZMPS regression models.

Finally, as a practical application of the proposed ZMPS model, we analyzed the number of femicides in 642 municipalities in São Paulo for the years 2019 and 2020. In this dataset, zero observations are particularly significant, as the counts represent the number of women murdered. We fitted ZMPS regression models to these counts, utilizing the Municipal Human Development Index (MHDI) as an explanatory variable and estimating the p_i parameter. This allowed for a more comprehensive analysis of the zero observations.

A limitation of the proposed model is the significant computational time and effort required to assess the impact of different initial conditions for a given dataset and to determine the appropriate chain length to ensure convergence and efficiency of the Monte Carlo estimators for the model parameters. Furthermore, the computational cost of fitting multiple models and applying a model selection criterion to identify the best PS distribution must be considered. In this study, we have used three criteria, LogCPO, EBIC, and WAIC, for Bayesian model selection.

A formulation of ZMPS regression models for longitudinal datasets that consider spatial correlations represents a natural extension for future research on the models presented in this paper.

Author Contributions: Conceptualization, K.S.C., M.G.A., V.H.L. and N.R.; Methodology, K.S.C., M.G.A., V.H.L. and N.R.; Formal analysis, K.S.C., M.G.A., V.H.L. and N.R.; Writing—review & editing, K.S.C., M.G.A., V.H.L. and N.R. All authors have read and agreed to the published version of the manuscript.

Funding: Marinho G. Andrade is supported by the Brazilian organization FAPESP (2019/21766-8); Katiane S. Conceição is supported by the Brazilian organization FAPESP (2019/22412-5).

Data Availability Statement: The data presented in this study are openly available at [statista] <https://www.statista.com/statistics/1102041/number-femicides-brazil-state/> (accessed on 4 January 2020).

Acknowledgments: The authors are grateful to the referees for their useful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Bayesian Relative Efficiency: Two Examples

Scenario 1: Large sample size: In the case of a large sample size, where the data information outweighs that of the prior densities, this scenario is referred to as the asymptotic case ($n \rightarrow \infty$). To assess the loss of efficiency when fitting the proposed ZMPS (or ZTPS) models to observations from a PS distribution, we consider Equation (27) under this assumption.

Scenario 2: Small sample size: In this scenario, the goal is to evaluate the impact of prior information about the model parameters on the loss of efficiency. To achieve this, we consider a small sample size (sufficient for fitting the ZMPS models) and two prior densities that differ only in their level of information (prior variance).

Appendix A.1. Poisson Distribution (P) Versus ZMP (or ZTP) Distribution

In this example, we consider the Poisson distribution $P(y | \mu)$ without the regression model for simplicity ($\theta = \mu$). Let us consider (27) to assess the effect of the information level of the prior distribution and the effect of the sample size on the loss of efficiency when we consider fitting a more general ZMP (or ZTP) for data coming from a Poisson distribution (P).

Let $\mathcal{D} = \mathbf{y} = (y_1, \dots, y_n)$ be the observed dataset coming from a Poisson distribution $P(\mathbf{y} | \mu) = e^{-\mu} \mu^{\mathbf{y}} / \mathbf{y}!$. Let $\mathcal{D}^+ = \mathbf{y}^+ = (y_1^+, \dots, y_m^+)$ be the positive observations in \mathcal{D} ($\mathcal{D}^+ \subset \mathcal{D}$). Considering the Poisson distribution with parameters μ and a prior density gamma $G(\alpha, \beta)$ for μ , we can write the posterior density as

$$\pi_0(\mu | \mathcal{D}) \propto \mu^{n\bar{y} + \alpha - 1} e^{-(n + \beta)\mu},$$

where \bar{y} is the sample mean, and the mode of the posterior distribution is given by

$$\mu_0^* = \frac{n\bar{y} + \alpha - 1}{n + \beta}.$$

The observed posterior information is given by

$$\mathcal{I}^{(0)}(\mu) = -\frac{\partial^2 \log \pi_0(\mu | \mathcal{D})}{\partial \mu^2} = \frac{n\bar{y} + \alpha - 1}{\mu^2},$$

and the inverse of the posterior observed information matrix $\mathcal{V}^{(0)}(\mu) = \mathcal{I}^{(0)}{}^{-1}(\mu)$ evaluated at μ_0^* is given by

$$\mathcal{V}^{(0)}(\mu_0^*) = \frac{n\bar{y} + \alpha - 1}{(n + \beta)^2}.$$

Remark A1. Note that the posterior variance is $\sigma_0^2 = n\bar{y} + \alpha / (n + \beta)^2$ and $\mathcal{V}^{(0)}(\mu^*) = \sigma_0^2 - 1 / (n + \beta)^2$. So $\mathcal{V}^{(0)}(\mu^*) \leq \sigma_0^2$. Therefore, paralleling the classic Cramér–Rao inequality, we can see that $\mathcal{V}^{(0)}(\mu^*)$ provides the posterior Cramér–Rao lower bound type for the posterior variance.

Let \mathcal{D}^+ be the positive observations in \mathcal{D} , and the ZTP model be

$$P(\mathbf{y}^+ | \mu) = \frac{e^{-\mu} \mu^{\mathbf{y}^+}}{\mathbf{y}^+! (1 - e^{-\mu})}.$$

Considering a gamma prior density $G(\alpha, \beta)$ for μ , the posterior density is given by

$$\pi_1(\mu | \mathcal{D}^+) \propto \frac{\mu^{m\bar{y}^+ + \alpha - 1} e^{-(m + \beta)\mu}}{(1 - e^{-\mu})^m},$$

where \bar{y}^+ is the sample mean of the positive observations, and the mode of the posterior distribution can be found by solving numerically the following equation:

$$\frac{\partial \log \pi_1(\mu \mid \mathcal{D}^+)}{\partial \mu} = \frac{(m\bar{y}^+ + \alpha - 1)}{\mu} - (m + \beta) - \frac{m}{e^\mu - 1} = 0.$$

The observed posterior information for the ZTP distribution is given by

$$\mathcal{I}^{(1)}(\mu) = -\frac{\partial^2 \log \pi_1(\mu \mid \mathcal{D}^+)}{\partial \mu^2} = \frac{m\bar{y}^+ + \alpha - 1}{\mu^2} - \frac{me^\mu}{(e^\mu - 1)^2},$$

while the inverse of the posterior observed information matrix $\mathcal{V}^{(1)}(\mu) = \mathcal{I}^{(1)-1}(\mu)$ is given by

$$\mathcal{V}^{(1)}(\mu) = \frac{\mu^2(e^\mu - 1)^2}{(m\bar{y}^+ + \alpha - 1)(e^\mu - 1)^2 - m\mu^2e^\mu}.$$

Considering the mode of the posterior distribution μ_1^* calculated via the Newton–Raphson procedure (26), we can evaluate the loss of relative efficiency $BRE(\widehat{\mu}_0, \widehat{\mu}_1)$ given in (27) when only the positive dataset \mathcal{D}^+ and the zero-truncated Poisson distribution are considered:

$$BRE(\widehat{\mu}_0, \widehat{\mu}_1) = \left(\frac{\mu_0^*}{\mu_1^*} \right)^2 \left(\frac{m\bar{y}^+ + \alpha - 1}{n\bar{y} + \alpha - 1} - \frac{\mu_1^{*2}e^{-\mu_1^*}}{(n\bar{y} + \alpha - 1)(1 - e^{-\mu_1^*})^2} \right). \tag{A1}$$

Since $n\bar{y} = m\bar{y}^+$, we can write (A1) as

$$BRE(\widehat{\mu}_0, \widehat{\mu}_1) = \left(\frac{\mu_0^*}{\mu_1^*} \right)^2 \left(1 - \frac{\mu_1^{*2}e^{-\mu_1^*}}{(m\bar{y}^+ + \alpha - 1)(1 - e^{-\mu_1^*})^2} \right). \tag{A2}$$

Prior Information

In this example, let us consider a gamma prior distribution $G(\alpha, \beta)$ for μ , i.e.,

$$\pi_0(\mu) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\beta\mu}, \quad \alpha > 1, \beta > 0.$$

Let $\mu^* = (\alpha - 1)/\beta^2$ represent the mode of the prior density that gives the Bayesian information conveyed by the prior density.

$$\mathcal{I}_{\mu\mu}^{(\pi_0)}(\mu^*) = -\frac{\partial^2 \log \pi_0(\mu)}{\partial \mu^2} \Big|_{\mu=\mu^*} = \frac{\beta^2}{\alpha - 1}.$$

The prior Bayesian information for μ is proportional to β^2 , meaning that the shape parameter β of the prior density is linked to the accuracy of the prior information about μ .

A simulation study was conducted to assess the effect of prior information and sample size (n) on the loss of relative efficiency (BRE), as calculated by (A2).

Figure A1a illustrates the asymptotic Bayesian relative efficiency (BRE) as a function of the parameters of the gamma prior distribution $G(\alpha, \beta)$ for $\beta = 5$ and $\beta = 50$ with $\alpha = \beta\mu$ for $\mu = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.

In Figure A1a, we consider $n = 300$. The figure illustrates that the information provided by the data outweighs that from the prior densities, resulting in BRE curves that are nearly identical for both prior densities. This figure demonstrates that the ZTP estimator is nearly fully efficient for $\mu \geq 6$. Figure A1b,c illustrate the sample Bayesian relative efficiency (BRE).

The individual points were derived from simulated Poisson samples of size $n = 30$ for each true parameter μ . The BRE was calculated using (A2). For $\mu \geq 5$, the sample BRE (with $n = 30$) and the asymptotic BRE (with $n = 300$) are nearly identical when $\beta = 5$ and $\beta = 50$.

However, as shown in Figure A1b,c, prior information enhances Bayesian efficiency for $\mu \leq 5$.

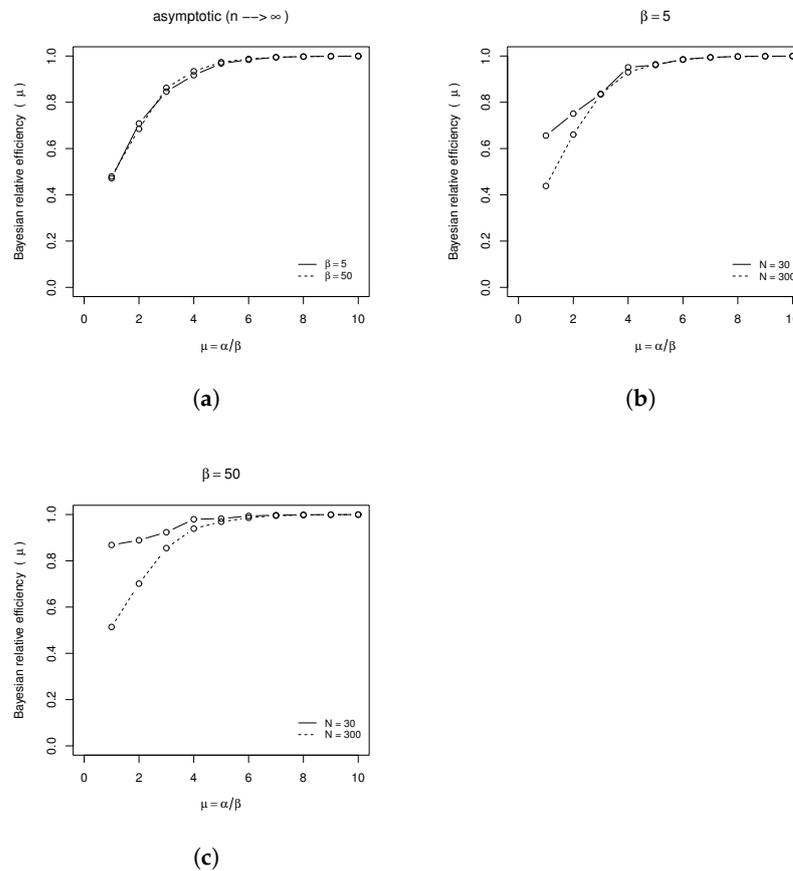


Figure A1. Bayesian relative efficiency of the ZTP estimation if a Poisson distribution is true: (a) asymptotic BRE ($n = 300$); (b) Sample BRE for $\beta = 5$ ($n = 30$); (c) Sample BRE for $\beta = 50$ ($n = 30$).

Appendix A.2. Negative Binomial Distribution (NB) Versus ZMNB (or ZTNB) Distribution

Using (24) and (25), we can assess the loss of efficiency in estimating μ and ϕ when observations are drawn from a traditional negative binomial distribution, but parameter estimates are obtained from the ZTNB distribution. From (25), we can express the lower bounds for the variance of the Bayesian estimates of μ and ϕ , considering only the positive samples and the ZTNB distribution, as follows:

$$V_{\mu\mu}^{(ZTNB)} = \frac{\mathcal{J}_{\phi\phi} + \Pi_{\phi\phi}}{(\mathcal{J}_{\mu\mu} + \Pi_{\mu\mu})(\mathcal{J}_{\phi\phi} + \Pi_{\phi\phi}) - (\mathcal{J}_{\mu\phi} + \Pi_{\mu\phi})^2}, \tag{A3}$$

$$V_{\phi\phi}^{(ZTNB)} = \frac{\mathcal{J}_{\mu\mu} + \Pi_{\mu\mu}}{(\mathcal{J}_{\mu\mu} + \Pi_{\mu\mu})(\mathcal{J}_{\phi\phi} + \Pi_{\phi\phi}) - (\mathcal{J}_{\mu\phi} + \Pi_{\mu\phi})^2}. \tag{A4}$$

The elements of the matrix $\mathcal{J}(\mu, \phi)$ are given by

$$\begin{aligned} \mathcal{J}_{\mu\mu} &= \sum_{i=1}^m \left[-y_i^+ h_g(\mu, \phi) + h_f(\mu, \phi) \right], \\ \mathcal{J}_{\mu\phi} &= \mathcal{J}_{\phi\mu} = \sum_{i=1}^m \left[-y_i^+ t_g(\mu, \phi) + t_f(\mu, \phi) \right], \\ \mathcal{J}_{\phi\phi} &= \sum_{i=1}^m \left[-k_a(y_i^+, \phi) - y_i^+ k_g(\mu, \phi) + k_f(\mu, \phi) \right], \end{aligned}$$

The evaluation of (A3) and (A4) depends on the choice of the prior density $\pi_0(\mu, \phi)$. Here, we illustrate this assessment by considering the loss in efficiency through the BRE, using independent prior gamma distributions $G(\alpha, \beta)$ for both μ and ϕ , respectively.

Using a similar procedure, we derived the observed information matrix for the parameters μ and ϕ of the traditional NB distribution. We denote this matrix by

$$\mathcal{J}^{(NB)}(\mu, \phi) = \begin{bmatrix} \mathcal{J}_{\mu\mu}^{(NB)} & \mathcal{J}_{\mu\phi}^{(NB)} \\ \mathcal{J}_{\phi\mu}^{(NB)} & \mathcal{J}_{\phi\phi}^{(NB)} \end{bmatrix},$$

and its elements are given by

$$\mathcal{J}_{\mu\mu}^{(NB)} = \sum_{i=1}^n \left[-y_i h_g(\mu, \phi) + h_f^{(NB)}(\mu, \phi) \right],$$

where

$$h_f^{(NB)}(\mu, \phi) = \frac{f_{\mu\mu}(\mu, \phi)f(\mu, \phi) - f_{\mu}^2(\mu, \phi)}{f^2(\mu, \phi)}.$$

$$\mathcal{J}_{\mu\phi}^{(NB)} = \mathcal{J}_{\phi\mu}^{(NB)} = \sum_{i=1}^n \left[y_i t_g(\mu, \phi) + t_f^{(NB)}(\mu, \phi) \right],$$

where

$$t_f^{(NB)}(\mu, \phi) = \frac{f_{\mu\phi}(\mu, \phi)f(\mu, \phi) - f_{\mu}(\mu, \phi)f_{\phi}(\mu, \phi)}{f^2(\mu, \phi)}.$$

$$\mathcal{J}_{\phi\phi}^{(NB)} = \sum_{i=1}^n \left[-k_a(y_i, \phi) - \mu k_g(\mu, \phi) + k_f^{(NB)}(\mu, \phi) \right],$$

where

$$k_f^{(NB)}(\mu, \phi) = \frac{f_{\phi\phi}(\mu, \phi)f(\mu, \phi) - f_{\phi}^2(\mu, \phi)}{f^2(\mu, \phi)}.$$

From the inverse Bayesian information matrix \mathcal{V} for the NB distribution, we can derive the variability of the posterior distribution for μ and ϕ as follows:

$$\mathcal{V}_{\mu\mu}^{(NB)} = \frac{\mathcal{J}_{\phi\phi}^{(NB)} + \Pi_{\phi\phi}}{(\mathcal{J}_{\mu\mu}^{(NB)} + \Pi_{\mu\mu})(\mathcal{J}_{\phi\phi}^{(NB)} + \Pi_{\phi\phi}) - (\mathcal{J}_{\mu\phi}^{(NB)} + \Pi_{\mu\phi})^2}, \tag{A5}$$

$$\mathcal{V}_{\phi\phi}^{(NB)} = \frac{\mathcal{J}_{\mu\mu}^{(NB)} + \Pi_{\mu\mu}}{(\mathcal{J}_{\mu\mu}^{(NB)} + \Pi_{\mu\mu})(\mathcal{J}_{\phi\phi}^{(NB)} + \Pi_{\phi\phi}) - (\mathcal{J}_{\mu\phi}^{(NB)} + \Pi_{\mu\phi})^2}. \tag{A6}$$

By considering the Bayesian relative efficiency (BRE), we can assess the loss of efficiency in estimating μ and ϕ when observations are drawn from a standard NB distribution, but parameter estimates are obtained from the ZTNB distribution. Using (A3)–(A6), the BRE is given by:

$$\text{BRE}(\mu) = \frac{\mathcal{V}_{\mu\mu}^{(NB)}}{\mathcal{V}_{\mu\mu}^{(ZTNB)}} \quad \text{and} \quad \text{BRE}(\phi) = \frac{\mathcal{V}_{\phi\phi}^{(NB)}}{\mathcal{V}_{\phi\phi}^{(ZTNB)}}.$$

Figure A2a displays the BRE for μ considering an infinite sample size (here, $n = 1000$). Figure A2b,c compare the asymptotic BRE ($n \rightarrow \infty$) with empirical BRE values calculated for $n = 50$. Figure A2b shows the BRE with a less informative prior density ($\beta = 5$), where the curve closely matches that obtained for a larger sample size ($n = 1000$). In Figure A2c, a more informative prior density ($\beta = 50$) is used, showing that prior information improves the BRE compared with that provided by the large sample size ($n = 1000$).

The plots in Figure A3 demonstrate that prior information enhances the BRE compared with that provided by a large sample size. Figure A3a displays the BRE for ϕ when considering an infinite sample size ($n = 1000$). Figure A3b,c compare the asymptotic BRE ($n = 1000$) with empirical BRE values calculated for $n = 50$.

As shown in Figures A2 and A3, for both parameters μ and ϕ of the negative binomial distribution, there is virtually no loss of efficiency from using the zero-truncated distribution when $\mu > 6$ and $\phi > 4$. However, the prior densities can still enhance the efficiency of estimating these parameters.

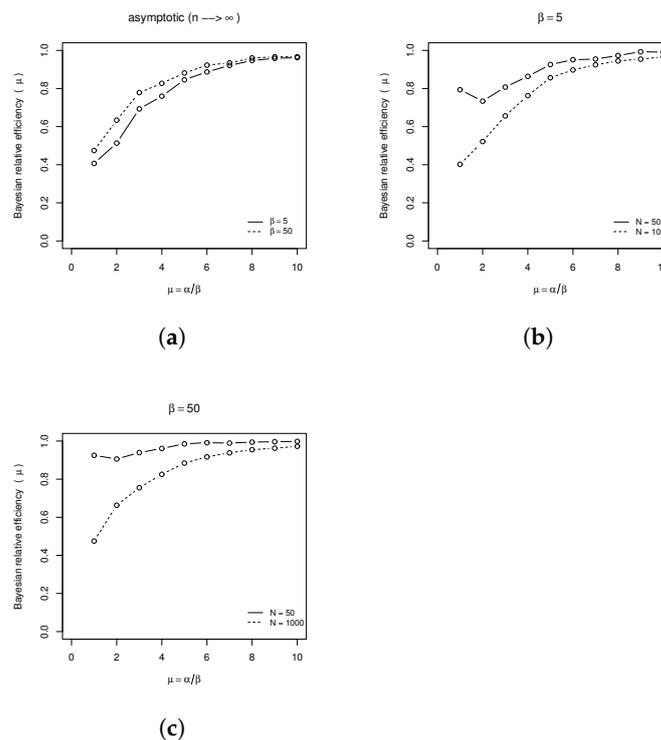


Figure A2. Bayesian relative efficiency of the ZTNB estimation for μ if a negative binomial distribution is true: (a) asymptotic BRE (μ); (b) empirical BRE (μ) for $\beta = 5$; (c) empirical BRE (μ) for $\beta = 50$.

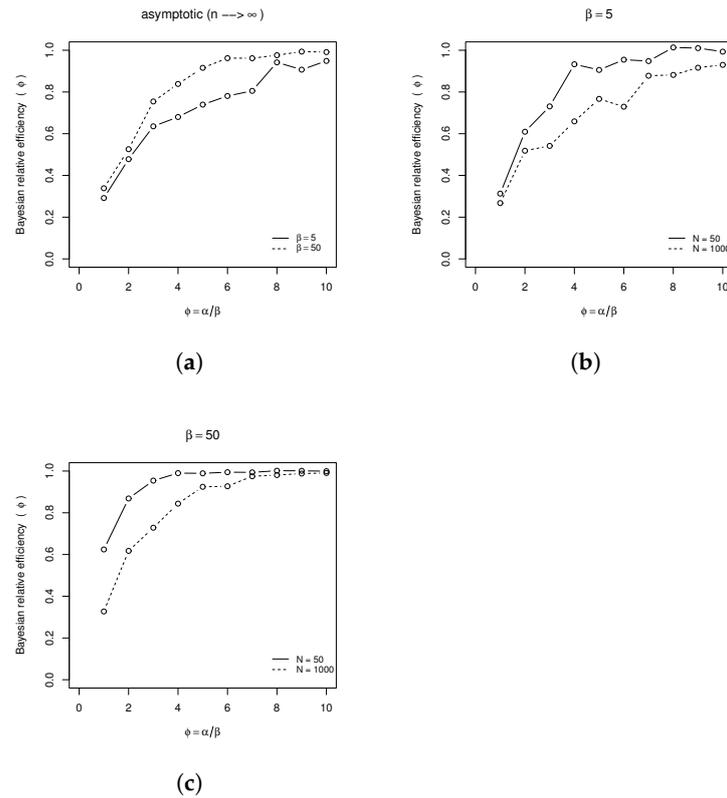


Figure A3. Bayesian relative efficiency of the ZTNB estimation for ϕ in a negative binomial distribution is true: (a) asymptotic BRE(ϕ); (b) empirical BRE(ϕ) for $\beta = 5$; (c) empirical BRE(ϕ) for $\beta = 50$.

Appendix B. Results of the SGHMC Algorithm Applied to the Data in Section 3.2

The results were obtained using the SGHMC algorithm outlined in Section 2.7. Three Markov chains, each consisting of 100,000 samples, were generated with distinct initial conditions. To minimize the effect of the burn-in period, 25% of each chain was discarded. From the remaining 75%, a subsample was taken for every 75th generated value (a process known as thinning) to reduce autocorrelation within the chains. This process yielded a final sample of 3000 observations, which was used to compute Bayesian Monte Carlo estimates. In addition, the Geweke [47] and Gelman–Rubin [48] convergence diagnostics were calculated using the R CODA package [49] (see Tables A1 and A2). The graphs in Figures A4 and A5 show the generated chains, the final combined chain, the histogram, and the autocorrelation function for the samples from the final chain.

Table A1. The Gelman–Rubin and Geweke convergence diagnostics to the data of 2019.

Parameters	Gelman & Rubin		Geweke
	Point Est.	Upper C.I.	$ Z < 1.96$
$\beta_{\mu,1}$	1.03	1.08	1.3895
$\beta_{\mu,2}$	1.03	1.08	1.3664
$\beta_{\omega,1}$	1.01	1.04	0.4394
$\beta_{\omega,2}$	1.01	1.04	0.4320
ϕ	1.00	1.00	1.1438

Multivariate potential scale reduction factor: 1.02.

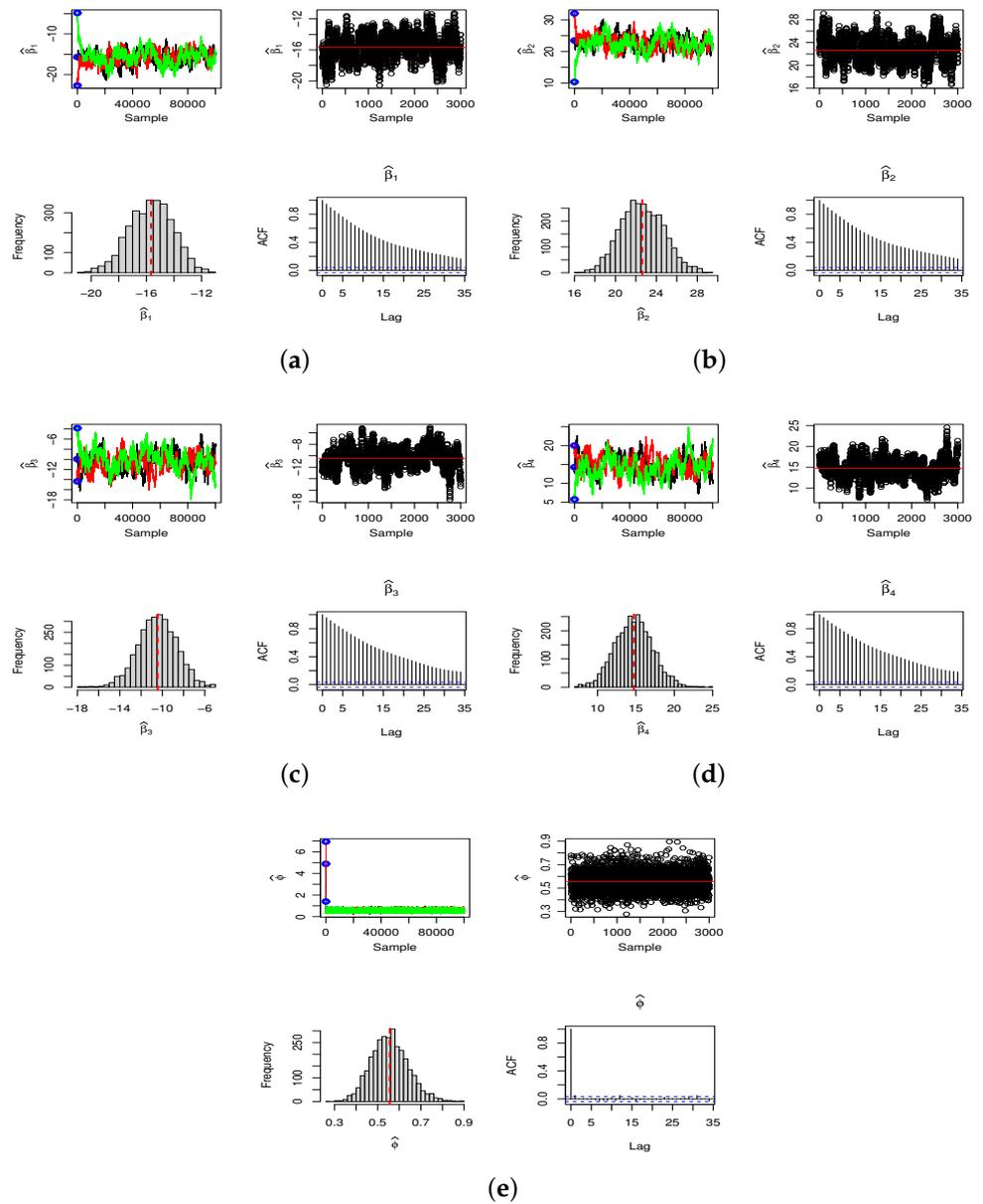


Figure A4. Results of the SGHMC algorithm for estimating the parameters of the ZMNB regression model applied to the data of 2019: (a) $\beta_1 = \beta_{\mu,1}$; (b) $\beta_2 = \beta_{\mu,2}$; (c) $\beta_3 = \beta_{\omega,1}$ (d) $\beta_4 = \beta_{\omega,2}$; (e) ϕ .

Table A2. the Gelman–Rubin and Geweke convergence diagnostics to the data of 2020.

Parameters	Gelman & Rubin		Geweke
	Point Est.	Upper C.I.	$ Z < 1.96$
$\beta_{\mu,1}$	1.03	1.08	1.3863
$\beta_{\mu,2}$	1.03	1.08	1.3648
$\beta_{\omega,1}$	1.01	1.03	0.4278
$\beta_{\omega,2}$	1.01	1.03	0.4206
ϕ	1.00	1.00	1.2430

Multivariate potential scale reduction factor: 1.02.

A sensitivity analysis was performed to assess the impact of the prior density parameters on the estimation of the model parameters. This analysis varied the prior mean parameters within the interval $[-50, 50]$ and considered prior variances of $\sigma^2 = 1, 10, 100,$ and $10,000$. The results showed that, under the same generation procedure used in the SGHMC algorithm as detailed in this paper, the parameter estimates did not exhibit sig-

nificant changes. However, the variance of the prior densities had a minimal effect on the parameter variances and, consequently, on the credibility intervals. Given the large dataset ($N = 642$), the prior parameters had little impact on the estimates. However, for smaller sample sizes, more informative prior densities could have a greater influence on the estimates. This is a fundamental aspect of Bayesian inference and highlights one of the key advantages of this approach.

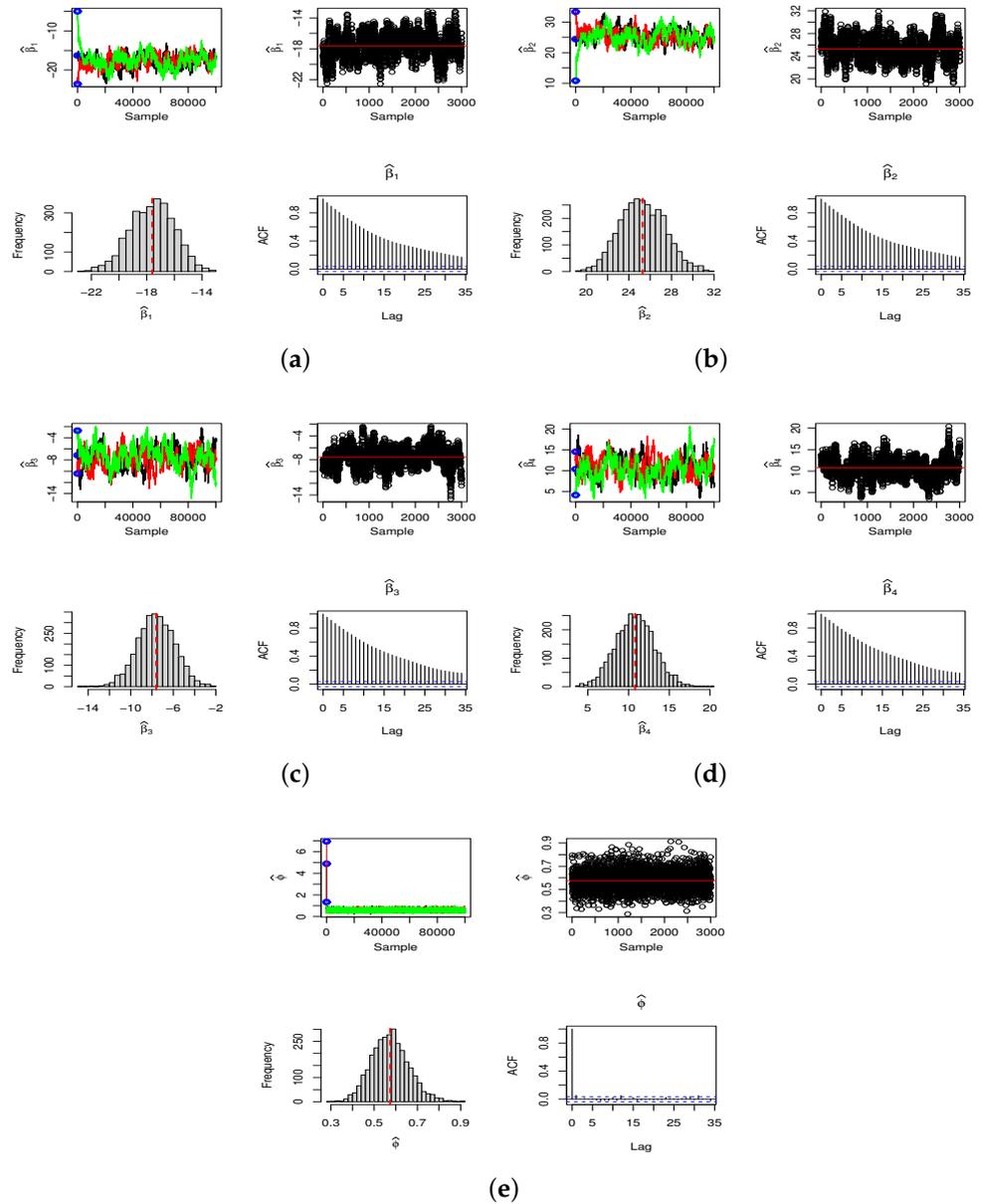


Figure A5. Results of the SGHMC algorithm for estimating the parameters of the ZMNB regression model applied to the data of 2020: (a) $\beta_1 = \beta_{\mu,1}$; (b) $\beta_2 = \beta_{\mu,2}$; (c) $\beta_3 = \beta_{\omega,1}$ (d) $\beta_4 = \beta_{\omega,2}$; (e) ϕ .

Appendix C. Model Selection

Bayesian models can be evaluated and compared in several ways. Details on various criteria for comparing models in the Bayesian context are presented by [38]. We use three criteria to select the best-fitting ZMPS regression model: the conditional predictive ordinate (CPO) [39], expected Bayesian information criterion (EBIC) [40], and Watanabe–Akaike information criterion (WAIC) [41].

The Conditional Predictive Ordinate (CPO) criterion [39] offers a functional cross-validation approach and serves as a computationally efficient measure of model fit. It is

calculated using the predictive probability density function of a future observation y_i , given the data excluding the i -th observation, $\mathcal{D}_{(-i)}$, as follows:

$$\pi_{ZMPS}(y_i | \mathcal{D}_{-i}) = \int_{\Omega_{\Theta}} \pi_{ZMPS}(y_i | \Theta, \mathcal{D}_{-i}) \pi(\Theta | \mathcal{D}_{-i}) d\Theta, \tag{A7}$$

where Ω_{Θ} is the parameter space. From (A7) and the generated chain from the SGHMC procedure $\{\Theta^{(g)}, g = 1, \dots, G\}$, we can estimate CPO_i by

$$\widehat{CPO}_i = \left[\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi_{ZMPS}(y_i | \Theta^{(g)}, \mathcal{D}_{-i})} \right]^{-1}. \tag{A8}$$

In many cases, it is computationally more convenient to calculate $\log(CPO_i)$ rather than CPO_i . In such situations, $\log(\widehat{CPO})$ is obtained by summing the estimates of $\log(\widehat{CPO}_i)$ see [38]. The model with the largest $\log(\widehat{CPO})$ is then selected as the best-fitting model for the data.

The expected Bayesian information criterion (EBIC) [40] is estimated from the SGHMC output as

$$\widehat{EBIC} = \frac{1}{G} \sum_{g=1}^G -2 \log(\mathcal{L}(\Theta^{(g)} | \mathcal{D}_n)) + p \log(n), \tag{A9}$$

where $\mathcal{L}(\Theta^{(g)} | \mathcal{D}_n)$ is the likelihood function defined in (6) evaluated using the generated chain from the SGHMC procedure $\{\Theta^{(g)}, g = 1, \dots, G\}$, p is the number of parameters in the model, and n is the total number of observations.

The Watanabe-Akaike information criterion (WAIC) was introduced by [41]. The WAIC criterion can be estimated using SGHMC output as

$$\widehat{WAIC} = -2 \left\{ \sum_{i=1}^n \log \left(\frac{1}{G} \sum_{g=1}^G \pi_{ZMPS}(y_i | \theta^{(g)}, \mathcal{D}) \right) - p_{WAIC} \right\}, \tag{A10}$$

where $p_{WAIC} = \sum_{i=1}^n V_{g=1}^G \log(\pi_{ZMPS}(y_i | \theta^{(g)}, \mathcal{D}))$, with $V_{g=1}^G$ representing the sample variance, i.e., $V_{g=1}^G a_g = \frac{1}{G-1} \sum_{g=1}^G (a_g - \bar{a})^2$.

To assess the performance of these model selection criteria for ZMPS regression models, we conducted a simulation study. In this study, we generated samples of size $N = 100$ from each of the ZMP, ZMNB, ZMGP, and ZM-COMP distributions and fitted the four models to each generated dataset. The SGHMC algorithm was applied with three different initial conditions. For each chain, 100,000 samples were generated, with the first 25% discarded as the burn-in period. From the remaining 75,000 samples, one was selected every 75 iterations (thinned sample), resulting in a joint sample of 3000 values. We then used this sample to calculate the Bayesian estimates for the logCPO, EBIC, and WAIC criteria. The results are presented in Tables A3–A6.

As shown in Table A3, for the data generated with the ZMP distribution, only the LogCPO criterion identified ZM-COMP as the best model, while the other criteria correctly identified the ZMP distribution as the best model.

As shown in Table A4, only the LogCPO criterion incorrectly identifies the best model. In contrast, both the WAIC and EBIC criteria correctly select the ZMNB model as the best fit.

Table A3. Data generated from the ZMP distribution.

Model	Criteria		
	LogCPO	WAIC	EBIC
ZMP	162	333	347
ZMNB	164	336	355
ZMGP	163	334	353
ZM-COMP	178	361	382

The selected model is highlighted in bold.

Table A4. Data generated from the ZMNB distribution.

Model	Criteria		
	LogCPO	WAIC	EBIC
ZMP	156	323	335
ZMNB	150	308	327
ZMGP	150	309	328
ZM-COMP	157	321	341

The selected model is highlighted in bold.

Table A5. Data generated from the ZMGP distribution.

Model	Criteria		
	LogCPO	WAIC	EBIC
ZMP	162	339	349
ZMNB	152	314	332
ZMGP	152	312	330
ZM-COMP	159	325	344

The selected model is highlighted in bold.

In Table A5, we observe that the LogCPO criterion incorrectly identifies the best model. However, both the WAIC and the EBIC criteria correctly select the ZMGP model as the best fit.

Table A6. Data generated from the ZM-COMP distribution.

Model	Criteria		
	LogCPO	WAIC	EBIC
ZMP	116	240	254
ZMNB	116	240	258
ZMGP	116	241	260
ZM-COMP	117	239	259

The selected model is highlighted in bold.

In Table A6, we observe that for data generated with the ZM-COMP distribution, both the LogCPO and WAIC criteria correctly identify the best model. However, the EBIC criterion incorrectly selects the ZMP model as the best fit. This is likely due to the lower complexity of the ZMP model, as the EBIC criterion tends to favor models with greater parsimony.

As a final analysis of this simulation study with artificial data, we conclude that the WAIC criterion is the most robust, consistently identifying the correct model as the best in all cases. However, using all three criteria provides a valuable approach to resolving ties and selecting the best model.

References

- Jain, G.C.; Consul, P.C. A Generalized Negative Binomial Distribution. *SIAM J. Appl. Math.* **1971**, *21*, 501–513. [[CrossRef](#)]
- Ng, T. A new class of modified binomial distributions with applications to certain toxicological experiments. *Commun. Stat. Theory Methods* **1989**, *18*, 3477–3492. [[CrossRef](#)]

3. Johnson, N.L.; Kemp, A.W.; Kotz, S. *Univariate Discrete Distributions*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2005.
4. Consul, P.C.; Famoye, F. *Lagrangian Probability Distributions*; Birkhäuser: Boston, MA, USA, 2006.
5. Frome, E.L. The analysis of rates using Poisson regression model. *Biometrics* **1983**, *39*, 665–674. [[CrossRef](#)]
6. Frome, E.L.; Checkoway, H. Use of Poisson regression models in estimating incidence rates and ratios. *Am. J. Epidemiol.* **1985**, *121*, 309–323. [[CrossRef](#)] [[PubMed](#)]
7. Conigliani, C.; Castro, J.I.; O'Hagan, A. Bayesian Assessment of Goodness of Fit against Nonparametric Alternatives. *Can. J. Stat.* **2000**, *28*, 327–342. [[CrossRef](#)]
8. Bayarri, M.J.; Berger, J.O.; Datta, G.S. Objective Bayes Testing of Poisson Versus Inflated Poisson Models. *Inst. Math. Stat.* **2008**, *3*, 105–121.
9. Hinde, J.; Demetrio, C.G.B. Overdispersion: Models and estimation. *Comput. Stat. Data Anal.* **1998**, *27*, 151–170. [[CrossRef](#)]
10. Gupta, R.C. Modified Power Series Distribution and Some of its Applications. *Indian J. Stat.* **1974**, *36*, 288–298.
11. Consul, P.C. New Class of Location-Parameter Discrete Probability Distributions and Their Characterizations. *Commun. Stat. Theory Methods* **1990**, *19*, 4653–4666. [[CrossRef](#)]
12. Cordeiro, G.M.; Andrade, M.G.; de Castro, M. Power Series Generalized Nonlinear Models. *Comput. Stat. Data Anal.* **2009**, *53*, 1155–1166. [[CrossRef](#)]
13. M'Kendrick, A.G. Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* **1926**, *44*, 98–103.
14. David, F.N.; Johnson, N.I. The truncated Poisson. *Biometrics* **1952**, *8*, 275–285. [[CrossRef](#)]
15. Cohen, A.C. An extension of a truncated Poisson distribution. *Biometrics* **1960**, *16*, 447–450. [[CrossRef](#)]
16. Umbach, D. On inference for a mixture of a Poisson and a degenerate distribution. *Commun. Stat. Theory Methods* **1981**, *10*, 299–306. [[CrossRef](#)]
17. Ridout, M.; Demétrio, C.G.B.; Hinde, J. Models for count data with many zeros. In Proceedings of the XIXth International Biometrics Conference, Cape Town, South Africa, 13–16 December 1998; pp. 179–192.
18. Dietz, E.; Böhning, D. On Estimation of the Poisson Parameter in Zero-Modified Poisson Models. *Comput. Stat. Data Anal.* **2000**, *34*, 441–459. [[CrossRef](#)]
19. Beckett, S.; Jee, J.; Ncube, T.; Pompilus, S.; Washington, Q.; Singh, A.; Pal, N. Zero-inflated Poisson (ZIP) distribution: Parameter estimation and applications to model data from natural calamities. *Involv. A J. Math.* **2014**, *7*, 751–767. [[CrossRef](#)]
20. Conceição, K.S.; Louzada, F.; Andrade, M.G.; Helou, E. Zero-modified power series distribution and its hurdle distribution version. *J. Stat. Comput. Simul.* **2017**, *87*, 1842–1862. [[CrossRef](#)]
21. Gelfand, A.; Smith, A. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409. [[CrossRef](#)]
22. Neal, R.M. Chapter 5: MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*; CRC Press: New York, NY, USA, 2011; pp. 113–162.
23. Roberts, G.O.; Rosenthal, J.S. Optimal Scaling of Discrete Approximations to Langevin Diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1998**, *60*, 255–268. [[CrossRef](#)]
24. Roberts, G.; Stramer, O. Langevin diffusions and Metropolis-Hastings algorithms. *Methodol. Comput. Appl. Probab.* **2003**, *4*, 337–358. [[CrossRef](#)]
25. Girolami, M.; Calderhead, B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 123–214. [[CrossRef](#)]
26. Hoffman, M.D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623.
27. Welling, M.; Teh, Y.W. Bayesian learning via stochastic gradient Langevin dynamics. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 681–688.
28. Chen, T.; Fox, E.; Guestrin, C. Stochastic Gradient Hamiltonian Monte Carlo. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 22–24 June 2014; Xing, E.P., Jebara, T., Eds.; ACM Digital Library: New York, NY, USA, 2014; pp. 1683–1691.
29. Ding, N.; Fang, Y.; Babbush, R.; Chen, C.; Skeel, R.D.; Neven, H. Bayesian sampling using stochastic gradient thermostats. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
30. Baker, J.; Fearnhead, P.; Fox, E.B.; Nemeth, C. sgmcmc: An R package for stochastic gradient Markov chain Monte Carlo. *J. Stat. Softw.* **2019**, *91*, 1–27. [[CrossRef](#)]
31. Nemeth, C.; Fearnhead, P. Stochastic Gradient Markov Chain Monte Carlo. *J. Am. Stat. Assoc.* **2021**, *116*, 433–450. [[CrossRef](#)]
32. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *arXiv* **2016**, arXiv:1603.04467.
33. Luo, R. Approximate Inference on Structured Distributions Using Stochastic Dynamics. Doctoral Thesis, UCL (University College London), London, UK, 2024.

34. Guikema, S.D.; Coffelt, J.P. A flexible count data regression model for risk analysis. *Risk Anal.* **2008**, *28*, 213–223. [[CrossRef](#)] [[PubMed](#)]
35. Mullahy, J. Specification and testing of some modified count data models. *J. Econom.* **1986**, *33*, 341–365. [[CrossRef](#)]
36. Feng, C.X. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *J. Stat. Distrib. Appl.* **2021**, *8*, 1–19. [[CrossRef](#)] [[PubMed](#)]
37. Huayanay, A.; Bazán, J.; Cancho, V.; Dey, D. Performance of asymmetric links and correction methods for imbalanced data in binary regression. *J. Stat. Comput. Simul.* **2019**, *89*, 1694–1714. [[CrossRef](#)]
38. Gelman, A.; Hwang, J.; Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **2014**, *24*, 997–1016. [[CrossRef](#)]
39. Gelfand, A.; Dey, K. Bayesian model choice: Asymptotics and exact calculations. *J. R. Stat. Soc.* **1994**, *56*, 501–514. [[CrossRef](#)]
40. Carlin, B.; Louis, T. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed.; Chapman & Hall/CRC Texts in Statistical Science; Taylor & Francis: London, UK, 2010.
41. Watanabe, S. A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **2013**, *14*, 867–897.
42. Ferreira, P. Extending Fisher’s measure of information. *Biometrika* **1981**, *68*, 695–698. [[CrossRef](#)]
43. Schützenberger, M.P. A generalization of the Frechet-Cramer inequality to the case of Bayes estimation. *Bull. Amer. Math. Soc.* **1957**, *63*, 142.
44. Pereira, C.A.B.; Stern, J.M. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy* **1999**, *1*, 99–110. [[CrossRef](#)]
45. Bernardo, J.; Smith, A. *Bayesian Theory*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2009.
46. Datasus. Ministério da Saúde. DATASUS. Tabnet. 2022. Available online: <https://datasus.saude.gov.br/mortalidade-desde-1996-pela-cid-10> (accessed on 4 January 2020).
47. Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *J. R. Stat. Soc.* **1994**, *56*, 501–514.
48. Gelman, A.; Rubin, D. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, *7*, 457–511. [[CrossRef](#)]
49. Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **2006**, *6*, 7–11.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.