





Article

Bayesian Reference Analysis for the Generalized Normal Linear Regression Model

Vera Lucia Damasceno Tomazella ^{1,*}, Sandra Rêgo Jesus ^{2,†}, Amanda Buosi Gazon ^{1,†} , Francisco Louzada ^{3,†}, Saralees Nadarajah ^{4,†} , Diego Carvalho Nascimento ^{5,†}  and Francisco Aparecido Rodrigues ^{3,†} and Pedro Luiz Ramos ^{3,*} 

- ¹ Department of Statistics, Universidade Federal de São Carlos, São Paulo 13565-905, Brazil; amandagazon@alumni.usp.br
- ² Multidisciplinary Health Institute, Federal University of Bahia, Vitória da Conquista, Bahia 45029-094, Brazil; sandrarj@ufba.br
- ³ Institute of Mathematical Science and Computing, University of São Paulo, São Carlos 13566-590, Brazil; louzada@icmc.usp.br (F.L.); francisco@icmc.usp.br (F.A.R.)
- ⁴ School of Mathematics, University of Manchester, Manchester M13 9PR, UK; saralees.nadarajah@manchester.ac.uk
- ⁵ Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; diego.nascimento@uda.cl
- * Correspondence: vera@ufscar.br (V.L.D.T.); pedrolramos@usp.br (P.L.R.)
- † These authors contributed equally to this work.



Citation: Tomazella, V.L.D.; Jesus, S.R.; Gazon, A.B.; Louzada, F.; Nadarajah, S.; Nascimento, D.C.; Rodrigues, F.A.; Ramos P.L. Bayesian Reference Analysis for the Generalized Normal Linear Regression Model. *Symmetry* **2021**, *13*, 856. <https://doi.org/10.3390/sym13050856>

Academic Editor: Christophe Chesneau

Received: 9 April 2021
Accepted: 29 April 2021
Published: 12 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Abstract: This article proposes the use of the Bayesian reference analysis to estimate the parameters of the generalized normal linear regression model. It is shown that the reference prior led to a proper posterior distribution, while the Jeffreys prior returned an improper one. The inferential purposes were obtained via Markov Chain Monte Carlo (MCMC). Furthermore, diagnostic techniques based on the Kullback–Leibler divergence were used. The proposed method was illustrated using artificial data and real data on the height and diameter of Eucalyptus clones from Brazil.

Keywords: Bayesian inference; generalized normal linear regression model; normal linear regression model; reference prior; Jeffreys prior; Kullback–Leibler divergence

1. Introduction

Although the normal distribution is used in many fields for symmetrical data modeling, when the data come from a distribution with lighter or heavier tails, the assumption of normality becomes inappropriate. Such circumstances show the need for more flexible models such as the Generalized Normal (GN) distribution [1], which encompasses various distributions such as the Laplace, normal, and uniform distributions.

In the presence of covariates, the normal linear regression model can be used to investigate and model the relationship between variables, assuming that the observations follow a normal distribution. However, it is well known that a normal linear regression model can be influenced by the presence of outliers [2,3]. In these circumstances, as discussed above, it is necessary to use models in which the error distribution presents heavier or lighter tails than normal such as the GN distribution.

Due to its flexibility, the GN distribution is considered a tool to reduce the impact of the outliers and obtain robust estimates [4–6]. This distribution has been used in different contexts, with different parameterizations, but the main difficulties to adopting GN distribution modeling have been computational problems since there are no explicit expressions for estimators of the shape parameter. The estimation of the shape parameter should be done through numerical methods. However, in the classical context, there are problems of convergence, as demonstrated in the studies [7,8].

In the Bayesian context, the methods presented in the literature for the estimation of the parameters of the GN distribution are restricted and applied to particular cases. West [9] proved that a scale mixture of the normal distribution could represent the GN distribution. Choy and Smith [10] used the prior GN distribution for the location parameter in the Gaussian model. They obtained the summaries of the posterior distribution, estimated through the Laplace method, and examined their robustness properties. Additionally, the authors used the representation of the GN distribution as a scale mixture of the normal distribution in random effect models and considered the Markov Chain Monte Carlo method (MCMC) to obtain the posterior summaries of the parameter of interest.

Bayesian procedures for regression models with GN errors have been discussed earlier. Box and Tiao [4] used the GN distribution from the Bayesian approach in which they proposed robust regression models as an alternative to the assumptions of normality of errors in regression models. Salazar et al. [11] considered an objective Bayesian analysis for exponential power regression models, i.e., a reparametrized version of the GDN. They derived the Jeffreys prior and showed that such a prior results in an improper posterior. To overcome this limitation, they considered a modified version of the Jeffreys prior under the assumption of independence of the parameters. This assumption does not hold since the scale and the shape parameters correlate with the proposed distribution. Additionally, the use of the Jeffreys prior is not appropriate in many cases and can cause strong inconsistencies and marginalization paradoxes (see Bernardo [12], p. 41).

Reference priors, also called objective priors, can be used to overcome this problem. This method was introduced by Bernardo [13] and enhanced by Berger and Bernardo [14]. For the proposed methodology, the prior information is dominated by the information provided by the data, resulting in a vague influence of the prior distribution. Estimations are made based on priors through the maximization of the expected Kullback–Leibler (KL) divergence between the posterior and prior distributions. The resulting reference prior affords a posterior distribution that has interesting features, such as consistent marginalization, one-to-one invariance, and consistent sampling properties [12]. Some applications of reference priors can be seen for other distributions in [15–20].

In this paper, we considered the reference approach for estimating the parameters of the GN linear regression model. We showed that the reference prior led to a proper posterior distribution, whereas the Jeffreys prior brought an improper one and should not be used. The posterior summaries were obtained via Markov Chain Monte Carlo (MCMC) methods. Furthermore, diagnostic techniques based on the Kullback–Leibler divergence were used.

To exemplify the proposed model, we considered the 1309 entries on the height and diameter of Eucalyptus clones (more details are given in Section 8). For these data, a linear regression model using the normal distribution for the residuals was not adequate, and so, we used the GN linear regression approach.

This paper is composed as follows. Section 2 presents the GN distribution with some special cases, and in the following Section 3, we introduce the GN linear regression model. Section 4 shows the reference and Jeffreys priors, respectively, for a GN linear regression model. Then, the model selection criteria, in Section 6, are discussed, as well as their applications. Section 7 shows the proposed method for analyzing compelling cases considering the Bayesian reference prior approach through the use of the Kullback–Leibler divergence. Section 8 presents studies with an artificial and a real application, respectively. Finally, we discuss the conclusions in Section 9.

2. Generalized Normal Distribution

The Generalized Normal distribution (GN distribution) has been referred to in the literature under different names and parametrizations such as the exponential power distribution or the generalized error distribution. The first formulation of this distribution [1] as a generalization of the normal distribution was characterized by the location, scale, and shape parameters.

Here, we considered the form presented by Nadarajah [21]. It is understood that the random variable Y is the GN distribution given its probability density function (pdf) as:

$$f(y|\mu, \tau, s) = \frac{s}{2\tau\Gamma\left(\frac{1}{s}\right)} \exp\left\{-\left(\frac{|y-\mu|}{\tau}\right)^s\right\}, \quad y, \mu \in \mathfrak{R}, \quad \tau, s > 0. \quad (1)$$

The parameter μ is the mean; τ is the scale factor; and s is the shape parameter. In particular, the mean, variance, and coefficient of the kurtosis of Y are given by:

$$E(Y) = \mu, \quad \text{Var}(Y) = \frac{\tau^2\Gamma\left(\frac{3}{s}\right)}{\Gamma\left(\frac{1}{s}\right)} \quad \text{and} \quad \gamma = \frac{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{5}{s}\right)}{\Gamma\left(\frac{3}{s}\right)^2},$$

respectively.

The GN distribution characterizes leptokurtic distributions if $s < 2$ and platykurtic distributions if $s > 2$. In particular, the GN distribution displays the Laplace distribution when $s = 1$ and the normal distribution when $s = 2$ and τ is equal to $\sqrt{2}\sigma$, where σ is the standard deviation, and when $s \rightarrow \infty$, the pdf converges to a uniform distribution in (1).

This distribution is flexible-symmetrical concerning the average and unimodality. Moreover, it allows a more flexible fit for the kurtosis than a normal distribution. Furthermore, the ability of the GN distribution to provide a precise fit for the data depends on its shape.

Zhu and Zinde-Walsh [22] proposed a reparameterization of the asymmetric exponential power distribution that allows us to observe the effect of the shape parameter on the distribution, which was adapted for the GN distribution. Using a similar reparameterization, $\sigma = \tau\Gamma\left(1 + \frac{1}{s}\right)$, the GN distribution in (1) is given by:

$$f(y|\mu, \sigma, s) = 2^{-1}\sigma^{-1} \exp\left\{-\left(\frac{\Gamma\left(1 + \frac{1}{s}\right)|y-\mu|}{\sigma}\right)^s\right\} \quad y, \mu \in \mathfrak{R}, \sigma > 0, s > 0. \quad (2)$$

The reparameterization above is used throughout the paper. Figure 1 shows the density functions of the GN distribution in (2) for various parameter values.

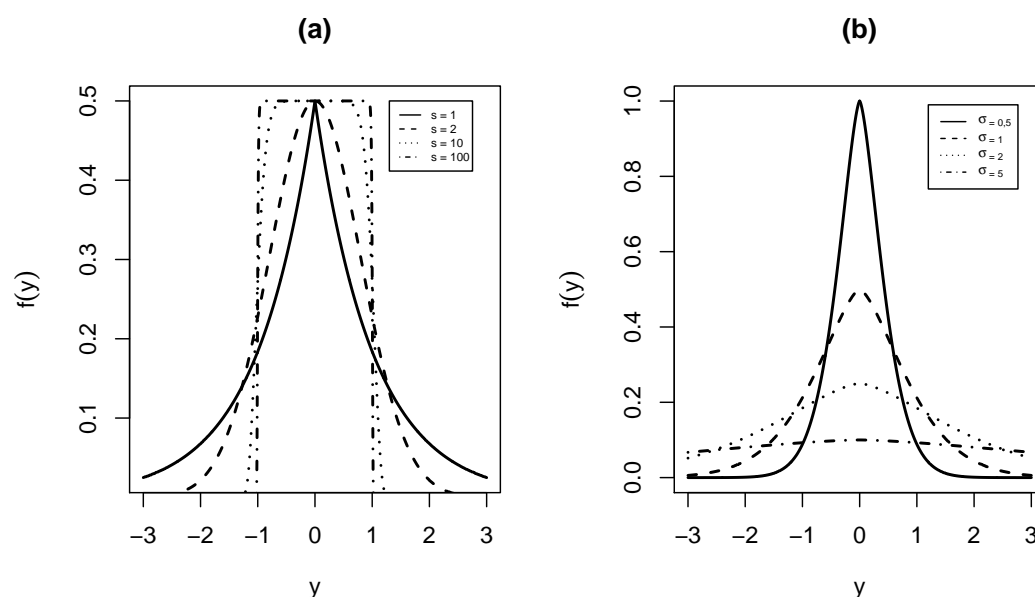


Figure 1. Density functions of the GN distribution with the parameters (a) $\mu = 0$ and $\sigma = 1$ fixed and varying s and (b) $\mu = 0$ and $s = 1.5$ fixed and varying σ .

3. Generalized Normal Linear Regression Model

The GN linear regression model is defined as:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n, \quad (3)$$

where Y_i is the vector of the response for the i th case, $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ contains the values of the explanatory variables, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the vector of regression coefficients, and ϵ_i is the vector of random errors that follows a GN distribution with mean zero, scale parameter σ , and shape parameter s .

Therefore, the likelihood function is:

$$L(\mathbf{y}|\boldsymbol{\beta}, \sigma, s) = 2^{-n} \sigma^{-n} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s \right\}. \quad (4)$$

The log-likelihood function (4) is given by:

$$\log L(\mathbf{y}|\boldsymbol{\beta}, \sigma, s) = -n \log 2 - n \log \sigma - \sum_{i=1}^n \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s. \quad (5)$$

The first-order derivatives of the log-likelihood function in (5) are given by:

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \frac{s \Gamma\left(1 + \frac{1}{s}\right)}{\sigma} \sum_{i=1}^n \mathbf{x}_i^\top \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^{s-1} \text{sign}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), \quad (6)$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{s}{\sigma} \sum_{i=1}^n \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s, \quad (7)$$

$$\frac{\partial \log L}{\partial s} = - \sum_{i=1}^n \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s \left[\log \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right) - \frac{\Psi\left(1 + \frac{1}{s}\right)}{s} \right], \quad (8)$$

where $\Psi(s) = \frac{\Gamma'(s)}{\Gamma(s)}$ is the digamma function.

The score function obtains the Fisher information matrix. This matrix is helpful to get the reference priors for the model parameters. The following proposition obtains the elements of the Fisher information matrix for the model in (3).

Proposition 1. Let $I(\boldsymbol{\theta})$ be the Fisher information matrix, with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, s)$. The elements of the Fisher information matrix,

$$I_{ij}(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right) = E \left(\frac{\partial \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_j} \right), \quad i, j = 1, 2, 3,$$

with $I_{ij}(\boldsymbol{\theta}) = I_{ji}(\boldsymbol{\theta})$ and θ_j the j th element of $\boldsymbol{\theta}$, are given by,

$$\begin{aligned}
I_{11}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \boldsymbol{\beta}^\top} \right) \right] = \frac{\Gamma\left(\frac{1}{s}\right)\Gamma\left(2 - \frac{1}{s}\right)}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top, \\
I_{12}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = \mathbf{0}, \\
I_{13}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = \mathbf{0}, \\
I_{22}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] = \frac{ns}{\sigma^2}, \\
I_{23}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = -\frac{n}{\sigma s}, \\
I_{33}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] = \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right\},
\end{aligned}$$

where $s > 1$ and $\Psi'(s) = \frac{\partial \Psi(s)}{\partial s}$ is the trigamma function. The restriction $s > 1$ ensures that the elements $I_{ij}(\boldsymbol{\theta})$, calculated for $i, j = 1, 2, 3$, are finite and the information matrix $I(\boldsymbol{\theta})$ is positive definite.

For further details of this proof, please see Proposition 5 in Zhu and Zinde-Walsh [22]. Then, Fisher's information matrix is given by:

$$I(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\Gamma\left(\frac{1}{s}\right)\Gamma\left(2 - \frac{1}{s}\right)}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{ns}{\sigma^2} & -\frac{n}{\sigma s} \\ \mathbf{0} & -\frac{n}{\sigma s} & \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right\} \end{bmatrix}. \quad (9)$$

The corresponding inverse Fisher's information matrix is given by:

$$B(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma^2}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(2 - \frac{1}{s}\right) \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2}{ns \left[1 - \frac{s}{(1+s)\Psi'\left(1 + \frac{1}{s}\right)} \right]} & \frac{\sigma s^2}{n \left[-s + (1+s)\Psi'\left(1 + \frac{1}{s}\right) \right]} \\ \mathbf{0} & \frac{\sigma s^2}{n \left[-s + (1+s)\Psi'\left(1 + \frac{1}{s}\right) \right]} & \frac{s^4}{n \left[-s + (1+s)\Psi'\left(1 + \frac{1}{s}\right) \right]} \end{bmatrix}. \quad (10)$$

The matrix in (9) coincides with the Fisher information matrix found by Salazar et al. [11] due to the one-to-one invariance property.

4. Objective Bayesian Analysis

An important class of objective priors was introduced by Bernardo [13] and later developed by Berger and Bernardo [14]. This class of prior is known as reference priors. A vital feature of the method developed by Berger–Bernardo is the specific treatment given to interest and nuisance parameters. The construction of the reference prior, in the case of nuisance parameters, must be made using an ordered parameterization. The parameter of interest is selected, and the procedure is followed below (see Bernardo [12], for a detailed discussion).

Proposition 2. (Reference priors under asymptotic normality) Let a probability model $f(y|\boldsymbol{\theta}, \boldsymbol{\omega}), y \in \mathbb{R}^n, \boldsymbol{\omega} = (\omega_1, \dots, \omega_m), \boldsymbol{\theta} \in \Phi, \boldsymbol{\omega} \in \Omega = \prod_{j=1}^m \Omega_j$, with $m + 1$ real-valued parameters, and let $\boldsymbol{\theta}$ be the quantity of interest. For instance, the posterior distribution of the parameters $(\boldsymbol{\theta}, \omega_1, \dots, \omega_m)$ is asymptotically normal with covariance matrix $B(\hat{\boldsymbol{\theta}}, \hat{\omega}_1, \dots, \hat{\omega}_m)$, where $H = B^{-1}$, B_j is the upper left $j \times j$ submatrix of B , $H_j = B_j^{-1}$, and $h_{jj}(\boldsymbol{\theta}, \omega_1, \dots, \omega_m)$ is the lower right element of H_j .

Then, it holds that the conditional reference priors can be represented as:

$$\pi^R(\omega_m|\theta, \omega_1, \dots, \omega_{m-1}) \propto h_{m+1, m+1}^{\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m),$$

and:

$$\pi^R(\omega_i|\theta, \omega_1, \dots, \omega_{i-1}) \propto \exp \left\{ \int_{\Omega_{i+1}} \dots \int_{\Omega_m} \log h_{i+1, i+1}^{\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m) \times \left\{ \prod_{j=i+1}^m \pi^R(\omega_j|\theta, \omega_1, \dots, \omega_{j-1}) \right\} d\omega_{i+1} \right\},$$

where $d\omega_j = d\omega_j \times \dots \times d\omega_m$ and $\pi^R(\omega_i|\theta, \omega_1, \dots, \omega_{i-1}), i = 1, \dots, m$ are all proper. A compact approximation will be only required, for the corresponding integrals, if any of those conditional reference priors is not proper.

Furthermore, the marginal reference prior of θ is:

$$\pi^R(\theta) \propto \exp \left\{ \int_{\Omega_1} \dots \int_{\Omega_m} \log B_{11}^{-\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m) \left\{ \prod_{j=1}^m \pi^R(\omega_j|\theta, \omega_1, \dots, \omega_{j-1}) \right\} d\omega_1 \right\},$$

where $B_{11}^{-\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m) = h_{11}^{\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m)$.

The reference posterior distribution associated with θ , after \mathbf{y} , is given by:

$$\pi^R(\theta|\mathbf{y}) \propto \pi^R(\theta) \int_{\Omega_1} \dots \int_{\Omega_m} f(\mathbf{y}|\theta, \omega_1, \dots, \omega_m) \left\{ \prod_{j=1}^m \pi^R(\omega_j|\omega_1, \dots, \omega_{j-1}) \right\} d\omega_1, \dots, d\omega_m.$$

The proposition is first presented for the presence of one nuisance parameter and further extended to a vector of nuisance parameters (see Bernardo [12] for a detailed discussion and proofs). Here, as our model has a special structure, we also considered an additional result presented in the following corollary that will be used to construct the reference prior to the GN distribution.

Corollary 1. *If the nuisance parameter spaces Ω_i do not depend on $\{\theta, \omega_1, \dots, \omega_{i-1}\}$ and the functions $h_{11}, h_{22}, \dots, h_{mm}, h_{m+1, m+1}$ factorize in the form:*

$$h_{11}^{\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m) = f_0(\theta)g_0(\omega_1, \dots, \omega_m) \text{ and}$$

$$h_{i+1, i+1}^{\frac{1}{2}}(\theta, \omega_1, \dots, \omega_m) = f_i(\omega_i)g_i(\theta, \omega_i, \dots, \omega_{i-1}, \omega_{i+1}, \dots, \omega_m), i = 1, \dots, m,$$

then $\pi^R(\theta) \propto f_0(\theta)$ and $\pi^R(\omega_i|\theta, \omega_1, \dots, \omega_{i-1}) \propto f_i(\omega_i), i = 1, \dots, m$, and there is no need for compact approximations, even if $\pi^R(\omega_i|\theta, \omega_1, \dots, \omega_{i-1})$ are not proper.

Under appropriate regularity conditions (see Bernardo [12]), the posterior distribution of $(\theta, \omega_1, \dots, \omega_m)$ is asymptotically normal with mean $(\hat{\theta}, \hat{\omega}_1, \dots, \hat{\omega}_m)$, the corresponding MLEs and covariance matrix $B(\hat{\theta}, \hat{\omega}_1, \dots, \hat{\omega}_m)$, where $B(\hat{\theta}, \hat{\omega}_1, \dots, \hat{\omega}_m) = I^{-1}(\hat{\theta}, \hat{\omega}_1, \dots, \hat{\omega}_m)$ and $I(\theta, \omega_1, \dots, \omega_m)$ is the corresponding $(m + 1) \times (m + 1)$ Fisher information matrix; in that case, $H(\theta, \omega_1, \dots, \omega_m) = I(\theta, \omega_1, \dots, \omega_m)$, and the reference prior may be computed from the elements of Fisher matrix $I(\theta, \omega_1, \dots, \omega_m)$.

4.1. Reference Prior

The parameter vector (β, σ, s) is ordered and divided into three distinct groups, according to their inferential importance. We considered here the case in which β is the

parameter of interest and σ and s are the nuisance parameters. To obtain a joint reference prior for the parameters β, σ , and s , the following ordered parameterization was adopted:

$$\pi^R(\beta, \sigma, s) = \pi^R(s|\beta, \sigma)\pi^R(\sigma|\beta)\pi^R(\beta).$$

Consider the Fisher matrix in (9), the inverse Fisher matrix in (10), and Corollary 1. Let $H(\theta) = I(\theta)$; it follows that $h_{33}^{\frac{1}{2}}(\beta, \sigma, s) = \sqrt{\frac{n}{s^3} \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right)} = f_2(s)g_2(\beta, \sigma)$. Then,

$$\pi^R(s|\beta, \sigma) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}.$$

Let $H_2(\theta) = B_2^{-1}(\theta)$, where $B_2(\theta)$ is the upper left 2×2 submatrix of $B(\theta)$; it follows that $h_{22}^{\frac{1}{2}}(\beta, \sigma, s) = \frac{1}{\sigma} \sqrt{ns \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})}\right]} = f_1(\sigma)g_1(\beta, s)$. Then,

$$\pi^R(\sigma|\beta) \propto \frac{1}{\sigma}.$$

Finally, let $h_{11}(\beta, \sigma, s) = B_{11}^{-1}(\beta, \sigma, s)$; it follows that $h_{11}^{\frac{1}{2}}(\beta, \sigma, s) = \sqrt{\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top} \sqrt{\frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})}{\sigma^2}} = f_0(\beta)g_0(\sigma, s)$. Then,

$$\pi^R(\beta) \propto 1.$$

Therefore, a joint reference prior for the ordered parameter is given by:

$$\pi^R(\beta, \sigma, s) \propto \frac{1}{\sigma} \times s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}, \tag{11}$$

where $\beta \in \mathbb{R}^p, \sigma \in \mathbb{R}^+,$ and $s > 1$.

Using the likelihood function (4) and the joint reference prior (11), we obtain the joint posterior distribution for β, σ , and s ,

$$\pi^R(\beta, \sigma, s|\mathbf{y}) \propto \sigma^{-(n+1)} s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}. \tag{12}$$

The posterior conditional probability densities are given by,

$$\pi^R(\beta|\sigma, s, \mathbf{y}) \propto \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}, \tag{13}$$

$$\pi^R(\sigma|\beta, s, \mathbf{y}) \propto \sigma^{-(n+1)} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}, \tag{14}$$

$$\pi^R(s|\beta, \sigma, \mathbf{y}) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}. \tag{15}$$

The densities in (13) and (15) do not belong to any known parametric family, and the densities in (14) can be easily reduced to an inverse-gamma distribution form by the transformation $\lambda = \sigma^s$. The parameters of interest were obtained by Monte Carlo methods

in a Markov Chain (MCMC). Thus, the posterior densities were evaluated by applying the Metropolis–Hastings algorithm; see, e.g., Chen et al. [23].

4.2. A Problem with the Jeffreys Prior

The use of the Jeffreys prior in the multiparametric case is often controversial. Bernardo ([12], p. 41) argued that the use of the Jeffreys prior is not appropriate in many cases and can cause strong inconsistencies and marginalization paradoxes. This prior is obtained from the square root of the determinant of the Fisher information matrix of (9),

$$\pi^J(\boldsymbol{\beta}, \sigma, s) \propto \sqrt{\det(I(\theta))} = \sqrt{\det(I_{11})[I_{22} I_{33} - I_{23}^2]},$$

where:

$$\det(I_{11}) = \left[\frac{\Gamma\left(\frac{1}{s}\right)\Gamma\left(2 - \frac{1}{s}\right)}{\sigma^2} \right]^p \det\left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top\right),$$

and:

$$[I_{22} I_{33} - I_{23}^2] = \frac{n^2}{\sigma^2 s^2} \left[\left(1 + \frac{1}{s}\right) \Psi'\left(1 + \frac{1}{s}\right) - 1 \right],$$

is given by:

$$\pi^J(\boldsymbol{\beta}, \sigma, s) \propto \sigma^{-(p+1)} \left[\Gamma\left(\frac{1}{s}\right)\Gamma\left(2 - \frac{1}{s}\right) \right]^{\frac{p}{2}} s^{-1} \left[\left(1 + \frac{1}{s}\right) \Psi'\left(1 + \frac{1}{s}\right) - 1 \right]^{\frac{1}{2}}. \tag{16}$$

Such a prior was also presented in Salazar et al. [11]. Both priors are from the family of prior distributions represented as:

$$\pi(\boldsymbol{\beta}, \sigma, s) \propto \frac{\pi(s)}{\sigma^c}, \quad a \in \mathfrak{R}; \tag{17}$$

they are usually improper, and c is a hyperparameter, while $\pi(s)$ is the prior related to the shape parameter.

The Jeffreys prior and reference prior are of the form (17) with, respectively,

$$c = 1 \text{ and } \pi^R(s) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi'\left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}, \tag{18}$$

$$c = p + 1 \text{ and } \pi^J(s) \propto s^{-1} \left[\left(1 + \frac{1}{s}\right) \Psi'\left(1 + \frac{1}{s}\right) - 1 \right]^{\frac{1}{2}} \left[\Gamma\left(\frac{1}{s}\right)\Gamma\left(2 - \frac{1}{s}\right) \right]^{\frac{p}{2}}. \tag{19}$$

The posterior distribution associated with the prior in (17) is proper if:

$$\int_1^\infty L(s|\mathbf{y}) \pi(s) ds < \infty, \tag{20}$$

where $L(s|\mathbf{y})$ is the integrated likelihood for s , given by:

$$\int_{\mathfrak{R}^p} \int_0^\infty L(\boldsymbol{\beta}, \sigma, s|\mathbf{y}) \sigma^{-c} d\sigma d\boldsymbol{\beta}. \tag{21}$$

Corollary 2. *The marginal prior for s given in Equations (18) and (19) is a continuous function in $[1, \infty]$, and as $s \rightarrow \infty$, we have $\pi^R(s) = O(s^{-3/2})$ and $\pi^J(s) = O(s^{\frac{p-2}{2}})$.*

Proof. See Appendix A. \square

Corollary 3. For $n > p + 1 - c$, the likelihood function for the parameter s , $L(s|\mathbf{y})$, under the class of priors (17), is a continuous function bounded in $[1, \infty]$ and with complexity $O(1)$ when $s \rightarrow \infty$.

Proof. See Appendix A. \square

Proposition 3. The reference prior given in (11) yields a proper posterior distribution, and the Jeffreys prior given in (16) leads to an improper posterior distribution.

Proof. See Appendix A. \square

Therefore, the Jeffreys prior leads to an improper posterior distribution and cannot be used in a Bayesian analysis. Another objective prior known as the maximum data information prior could be considered [16,24–26]; however, such a prior is not invariant under one-to-one transformations, which limits its use. Additionally, the main aim was to consider objective priors. We avoided the use of normal or gamma priors due to the lack of invariance in the parameters. Moreover, the prior may depend on hyperparameters that are not easy to elicit for the GN distributions, and the posterior estimates may vary depending on the included information. Finally, Bernardo [12] pointed out that the use of our “flat” priors to assign “non-informative” priors should be strongly discouraged because they often result in the suppression of important inappropriate and unjustified assumptions that can easily have a strong influence on the analysis, or even make it invalid.

5. Metropolis–Hastings Algorithm

Here, the Metropolis–Hastings algorithm is considered to sample from μ, σ , and s . In this case, the following conditional distributions are considered: $\pi^R(\mu|\sigma, s, \mathbf{y})$, $\pi^R(\sigma|\mu, s, \mathbf{y})$, and $\pi^R(s|\mu, \sigma, \mathbf{y})$, respectively. On the other hand, $\mu \in \mathcal{R}$, $\sigma > 0$, $s > 1$, we considered the following change of variables (μ, σ, s) for $\omega = (\omega_1, \omega_2, \omega_3) = (\mu, \log(\sigma), \log(s - 1))$. This modification leads the parametric space to \mathcal{R}^3 , which allows us to sample from the posterior distribution in a more efficient way. Considering the Jacobian transformation, the posterior distribution is given by:

$$\pi^R(\omega|\mathbf{y}) \propto \exp(\omega_2)^{-n} \exp(\omega_3) \left[\left(1 + \frac{1}{(1 + \exp(\omega_3))} \right) \Psi' \left(1 + \frac{1}{(1 + \exp(\omega_3))} \right) \right]^{\frac{1}{2}} \\ \times (1 + \exp(\omega_3))^{-\frac{3}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{(1 + \exp(\omega_3))} \right) |y_i - \omega_1|}{\exp(\omega_2)} \right)^{(1 + \exp(\omega_3))} \right\}.$$

The construction of the Metropolis–Hastings algorithm is done using a random walk for the parameter ω_1 , for instance the transition is obtained using $q(\omega_1, \omega_1^*)$, where ω_1^* is generated from $\omega_1^* = \omega_1 + kz$ where $z \sim N(0, \eta^2)$ and k is a constant that controls the acceptance rate. We have that η^2 is the diagonal element of the covariance matrix from the joint posterior distribution of ω , obtained assuming the maximum a posterior estimate of $\pi^R(\omega|\mathbf{y})$.

The computational stability was improved considering the logarithm scale. The steps to sample from the posterior distribution are:

1. Set the values $\omega^{(0)} = (\omega_1^{(0)}, \omega_2^{(0)}, \omega_3^{(0)}) = (\mu^{(0)}, \log(\sigma^{(0)}), \log(s^{(0)} - 1))$.
2. Generate ω_1^* from the proposal distribution $q(\omega_1^{(0)}, \omega_1^*)$.
3. Sample u from a uniform distribution $U(0, 1)$.
4. If $u \leq \min \left\{ 1, \exp \left[\log \pi^R(\omega_1^*|\omega_2^{(0)}, \omega_3^{(0)}, \mathbf{y}) - \log \pi^R(\omega_1^{(0)}|\omega_2^{(0)}, \omega_3^{(0)}, \mathbf{y}) \right] \right\}$, then update $\omega_1^{(1)}$ from ω_1^* ; otherwise, use the value of $\omega_1^{(0)}$, i.e., $\omega_1^{(1)} = \omega_1^{(0)}$.
5. Repeat the same steps above for $\omega_2^{(1)}$ and $\omega_3^{(1)}$.

6. Repeat Steps 2–5 until we obtain the target sample size.

After we generated the values of ω , we computed $(\mu, \sigma = \exp(\omega_2), s = 1 + \exp(\omega_3))$. It was assumed that η has the same values for all steps. The value of k is defined as aiming for the acceptance rate to be between 20% and 80% [27]. To confirm the convergence of the chains, the Geweke diagnostic [28] was used, as well as graphical analysis.

6. Selection Criteria for Models

In the Bayesian context, there are a variety of criteria that can be adopted to select the best fit between a collection of models. This paper considered the following criteria: the Deviation Information Criterion (DIC) defined by Spiegelhalter et al. [29] and the Expected Bayesian Information Criterion (EBIC) proposed by Brooks [30]. These criteria are based on the posterior mean deviation, $E(D(\omega))$, which is the deviation evaluated at the posterior mean of ω ; thus, it is estimated with $\hat{D} = \frac{1}{Q} \sum_{q=1}^Q D(\omega^{(q)})$, where the index q indicates the q -th realization of a total of Q realizations and $D(\omega) = -2 \sum_{i=1}^n \log(f(y_i|\omega))$, where $f(\cdot)$ is the probability density of the GN distribution. The criteria DIC and EBIC are given by:

$$DIC = \bar{D} + p_D = 2\bar{D} - \hat{D} \quad \text{and} \quad EBIC = \bar{D} + 2b,$$

where b is the number of parameters in the model and p_D is the effective number of parameters, defined as $E[D(\omega)] - D[E(\omega)]$, where $D[E(\omega)]$ is the mean posterior deviation, which can be estimated as $\hat{D} = D\left(\frac{1}{Q} \sum_{q=1}^Q D(\omega^{(q)})\right)$. Smaller values for the DIC and EBIC indicate the preferred model.

Another widely accepted criterion for model selection is the Conditional Predictive Ordinate (CPO). A detailed description of this selection criterion and the CPO statistic, as well as the applicability of the method for selecting models can be found in Gelfand et al. [31,32]. Let \mathcal{D} denote the complete data set and $\mathcal{D}_{(-i)}$ denote the data with the i -th observation excluded. Consider $\pi(\omega|\mathcal{D}_{(-i)})$ for $i = 1, \dots, n$, the posterior density of ω given $\mathcal{D}_{(-i)}$. Thus, we can define the i -th observation of the CPO_{*i*} by:

$$CPO_i = \int_{\omega} f(y_i|\omega)\pi(\omega|\mathcal{D}_{(-i)})d\omega = \left\{ \int_{\omega} \frac{\pi(\omega|\mathcal{D}_{(-i)})}{f(y_i|\omega)} d\omega \right\}^{-1}, i = 1, \dots, n,$$

where $f(y_i|\omega)$ is the probability density function. High values of CPO_{*i*} indicate the best model. An estimate for CPO_{*i*} can be obtained using an MCMC sample of the posterior distribution of ω given \mathcal{D} , $\pi(\omega|\mathcal{D})$. For this, let $\omega_1, \dots, \omega_Q$ be a sample of the distribution $\pi(\omega|\mathcal{D})$, of size Q . A Monte Carlo approximation of the CPO_{*i*} [23] is given by:

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\theta_q)} \right\}^{-1}.$$

A summary statistic of the CPO_{*i*}'s is $\sum_{i=1}^n \log(\widehat{CPO}_i)$, wherein the higher the value of B , the better the fit of the model. To illustrate the proposed methodology, a comparison between the normal and GN models is presented in Section 8.

7. Bayesian Case Influence Diagnostics

One way to observe the influence of observations on a fit of the model is via the global diagnosis; for instance, one can remove some cases from the analysis and analyze the effect of the removal [33]. The diagnostics of the case influence in a Bayesian perspective are based on the Kullback–Leibler divergence (K-L). Let $K(\pi, \pi_{(-i)})$ denote the K-L divergence

between π and $\pi_{(-i)}$, where π denotes the posterior distribution of ω for all data and $\pi_{(-i)}$ denotes the posterior distribution of ω without the i -th case. Specifically,

$$K(\pi, \pi_{(-i)}) = \int \pi(\omega|D) \log \left\{ \frac{\pi(\omega|D)}{\pi(\omega|D_{(-i)})} \right\} d\omega, \quad (22)$$

and therefore, $K(\pi, \pi_{(-i)})$ measures the effect of deleting the i -th case from the full one on the joint posterior distribution of ω . The calibration can be obtained by solving the equation:

$$K(\pi, \pi_{(-i)}) = K(B(0.5), B(p_i)) = \frac{-\log\{4p_i(1-p_i)\}}{2},$$

where the Bernoulli distribution $B(p)$ is expressed by the parameter p with success probability [34] and p_i is a calibration measure of the K-L divergence. After some algebraic manipulation, the obtained expression is:

$$p_i = 0.5 \left(1 + \sqrt{1 - \exp\{-2K(\pi, \pi_{(-i)})\}} \right),$$

bounded by $0.5 \leq p_i \leq 1$. Therefore, if p_i is significantly higher than 0.5, then the i -th case is influential.

The posterior expectation (22) can also be written in the form:

$$\begin{aligned} K(\pi, \pi_{(-i)}) &= \log E_{\omega|D} \{ [f(y_i|\omega)]^{-1} \} + E_{\omega|D} \{ \log [f(y_i|\omega)] \} \\ &= -\log(\text{CPO}_i) + E_{\omega|D} \{ \log [f(y_i|\omega)] \}, \end{aligned} \quad (23)$$

where $E_{\omega|D}(\cdot)$ denotes the expectation from the posterior $\pi(\omega|D)$. Thus, (23) can be estimated by using the MCMC methods to achieve the sample from the posterior $\pi(\omega|D)$. Therefore, if $\omega_1, \dots, \omega_Q$ is a sample of size Q of $\pi(\omega|D)$, then:

$$K(\widehat{\pi}, \widehat{\pi}_{(-i)}) = -\log(\widehat{\text{CPO}}_i) + \frac{1}{Q} \sum_{q=1}^Q \log [f(y_i|\omega^{(q)})].$$

8. Applications

In this section, the proposed method is illustrated using artificial and real data.

8.1. Artificial Data

An artificial sample of size $n = 500$ was generated in accordance with (3) with $p = 2$, $\mathbf{x}_i^\top = (1, x_{i1})$, $x_{i1} \sim N(2.5, 1)$, $\boldsymbol{\beta} = (2, -1.5)^\top$, $\sigma = 1$, and $s = 2.5$. The posterior samples were generated by the Metropolis–Hastings technique through the MCMC implemented in the R software [35]. A single chain of dimensions 300,000 was considered for each parameter, and also, we discarded the first 150,000 iterations (burn-in), aiming to reduce correlation problems. A space with a size of 15 was used, resulting in a final sample size of 10,000. The convergence of the chain was verified by the criterion proposed by Geweke [28]. Table 1 shows the posterior summaries for the parameters of the GN linear regression model. It can be seen that the estimates were close to the true values, and the 95% HPD credibility intervals covered the true values of the parameters.

Table 1. Artificial data. Posterior mean, median, standard deviation (SD), and 95% HDP intervals for the parameters of the model.

Parameters	Mean	Median	SD	95% HDP
β_1	1.995	1.996	0.086	(1.826; 2.164)
β_2	−1.510	−1.510	0.032	(−1.572; −1.447)
σ	1.027	1.028	0.055	(0.922; 1.135)
s	2.657	2.631	0.320	(2.042; 3.293)

We used the same sample previously simulated to investigate the K-L divergence measure in detecting the GN linear regression model's influential observations. We selected Cases 50 and 250 for perturbation. For each of these two cases and also considering both cases simultaneously, the response variable was disturbed as follows: $\tilde{y}_i = y_i + 5S_y$, where S_y is the standard deviation of y_i . The MCMC estimates were done similarly to those in the last section. Note that, due to the invariance property, μ can be computed for the standard GN distribution using the Bayes estimates of β_1 and β_2 , that is $\mu = x_i^\top \beta$.

To reveal the impact of the influential observations in the estimates of β_1, β_2, σ , and s , we calculated the measure of Relative Variation (RV), which was obtained by $RV = \left| \frac{\hat{\theta} - \hat{\theta}_0}{\hat{\theta}} \right| \times 100\%$, where $\hat{\theta}$ and $\hat{\theta}_0$ are the posterior averages of the model parameters considering the original data and perturbed data, respectively.

Table 2 shows the posterior estimates for the artificial data and the RVs of the estimates of the parameters concerning the original simulated data. The data set denoted by (a) consisted of the original simulated data set without perturbation, and the data sets denoted by (b) to (d) consisted of data sets resulting from perturbations in the original simulated data set. Higher values of the relative variations for the parameters σ and s showed the presence of influential points in the data set. However, the estimate of s did not differ so much from the perturbed Cases (c) and (d).

Table 2. Artificial data. Posterior mean, RV(%), and 95% HDP intervals for the parameters of the model.

Data Names	Perturbed Case	Parameters	Mean	RV (%)	95% HDP
a	None	β_1	1.995	—	(1.826; 2.164)
		β_2	−1.510	—	(−1.572; −1.447)
		σ	1.027	—	(0.922; 1.135)
		s	2.657	—	(2.041; 3.293)
b	50	β_1	2.026	1.554	(1.841; 2.211)
		β_2	−1.517	0.464	(−1.585; −1.449)
		σ	0.970	5.550	(0.866; 1.073)
		s	2.188	17.651	(1.763; 2.613)
c	250	β_1	1.966	1.454	(1.775; 2.154)
		β_2	−1.495	0.993	(−1.564; −1.423)
		σ	0.916	10.808	(0.816; 1.016)
		s	1.867	29.733	(1.565; 2.165)
d	{50, 250}	β_1	2.002	1.332	(1.807; 2.198)
		β_2	−1.506	0.265	(−1.578; −1.436)
		σ	0.902	12.171	(0.801; 1.003)
		s	1.773	33.271	(1.496; 2.052)

Considering the samples generated from the posterior distribution of the GN linear regression model parameters, we estimated the measures of the K-L divergence and their respective calibrations for each of the cases considered (a–d), as described in Section 7. The results in Table 3 show that for the data without perturbation (a), the selected cases

were not influential because they had small values for $K(\pi, \pi_{(-i)})$ and calibration close to 0.577. However, when the data were perturbed (b,d), the values of $K(\pi, \pi_{(-i)})$ were more extensive, and their calibrations were close or equal to one, indicating that these data were influential.

Table 3. Diagnostic measures for the artificial data.

Data Names	Case Number	$K(\pi, \pi_{(-i)})$	Calibration
a	50	0.0121	0.5774
	250	0.0014	0.5262
b	50	16.1593	1.0000
c	250	19.2236	1.0000
d	50	2.8796	0.9992
	250	18.2292	1.0000

8.2. Real Data Set

In order to illustrate the proposed methodology, recall the Brazilian Eucalyptus clones data set on the height (in meters) and the diameter (in centimeters) of Eucalyptus clones. The data belong to a large pulp and paper company from Brazil. As a strategy for the rising rentability of the forestry enterprise and keeping pulp and paper production under control, the company needs to keep an intensive Eucalyptus clone silviculture. The height of the trees is an important measure for selecting different clone species. Moreover, it is desirable to have trees of similar heights, possibly with a slight variation, and consequently with a distribution function with lighter tails.

The objective is to relate the tree's diameter (explanatory variable) with its height (response variable). The GN and normal linear regression models were fit to the data via the Bayesian reference process. The posterior samples were generated by the Metropolis–Hastings technique, similar to the simulation study, in which we considered a single chain of dimensions 300,000 for each parameter, with a burn-in of 150,000 iterations; additionally, jumps with a size of 15 were used, resulting in a final sample size of 10,000. The Geweke criteria verified the convergence of the chain.

Table 4 shows the posterior summaries for the parameters of both distributions and the model selection criteria. The GN linear regression model was the most suitable to represent the data as it performed better than the normal linear regression model for all the criteria used. For the fit regression model chosen, note that β_2 was significant, and then, for every one-unit increase in the diameter, the average height of Eucalyptus 0.95 meters increased. In particular, the analysis of the shape parameter ($s > 2$) provided strong evidence of a platykurtic distribution for the errors, and this favored the GN linear regression model. This was further confirmed by graphical analyses of the quantile residuals of the GN linear regression model presented in Figure 2d.

Table 4. Real data. Posterior mean and 95% HPD intervals for the parameters of the model and Bayesian comparison criteria.

Model	Parameters	Mean	95% HDP	DIC	EBIC	B
Generalized Normal	β_1	7.066	(6.398; 7.724)	5829.180	5847.96	−2914.595
	β_2	0.948	(0.907; 0.990)			
	σ	3.237	(3.034; 3.440)			
	s	2.865	(2.436; 3.305)			
Normal	β_1	7.002	(6.331; 7.671)	6248.652	6262.98	−3126.614
	β_2	0.949	(0.907; 0.991)			
	σ	1.993	(1.916; 2.068)			

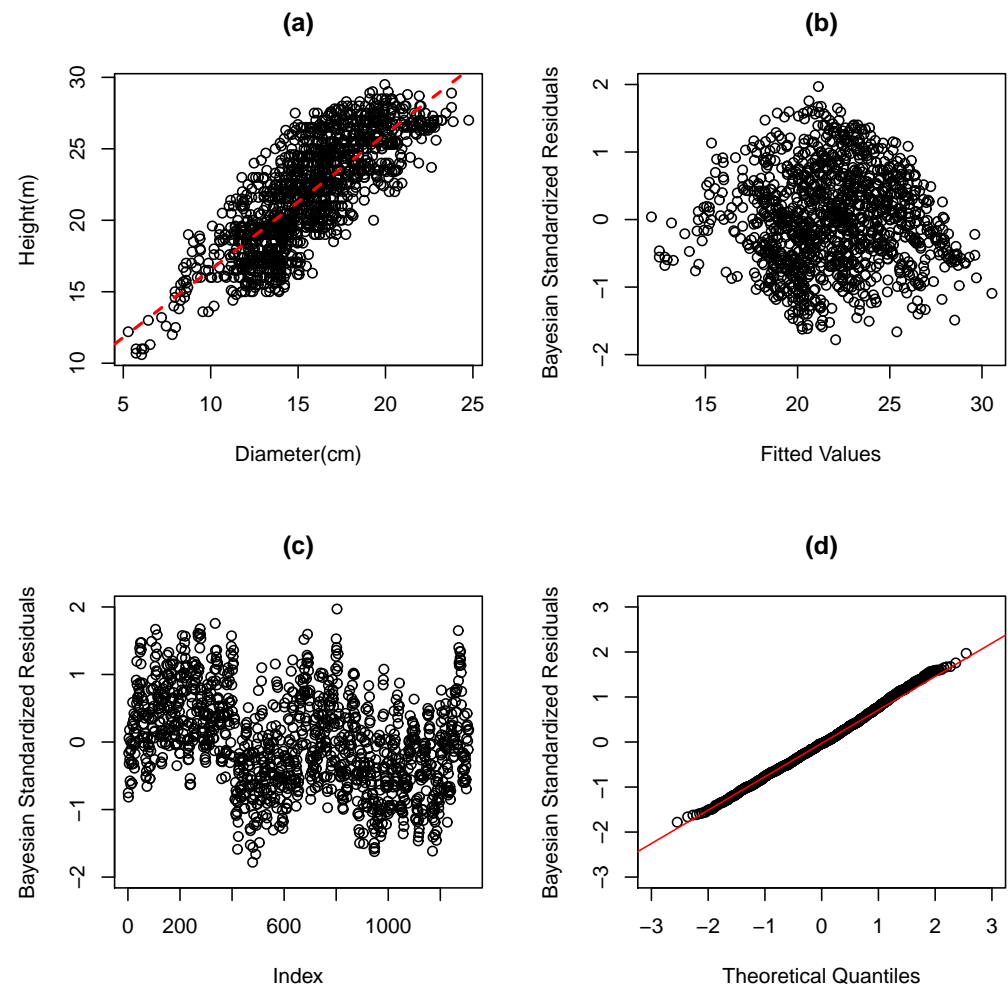


Figure 2. Real data. Generalized normal linear regression model with $s = 2.865$. (a) Scatterplot of the data and fit generalized normal linear regression model. (b) Quantile residuals versus fit values. (c) Quantile residuals versus the index. (d) Graph of the quantiles of the GN distribution for the residuals of the model.

Figures 2a and 3a show the scatter plot of the data and the adjusted normal and GN linear regression models. It was observed that, on average, the estimated heights were close to those observed, indicating that the models considered had a good fit. The residuals graph by the fit values and the residuals graph by the observations were also quite similar for both models. The presence of heteroscedasticity (see Figures 2b and 3b) was noted, as well as the quadratic trend (see Figures 2c and 3c) for the height-to-diameter ratio of the Eucalyptus. The graphs of the quantiles of the GN distribution and the normal distribution for the residuals of the models are presented in Figures 2d and 3d, respectively. It can be seen that in the setting of the normal linear regression model, many points were far from their tails, indicating an inadequate specification of the error distribution for the model. On the other hand, using our proposed approach in the GN distribution, we observed a good fit as points were following the line, indicating that the theoretical residuals were close to the observed residuals. Therefore, there was evidence that the model chosen outperformed the normal linear regression model in fitting the data.

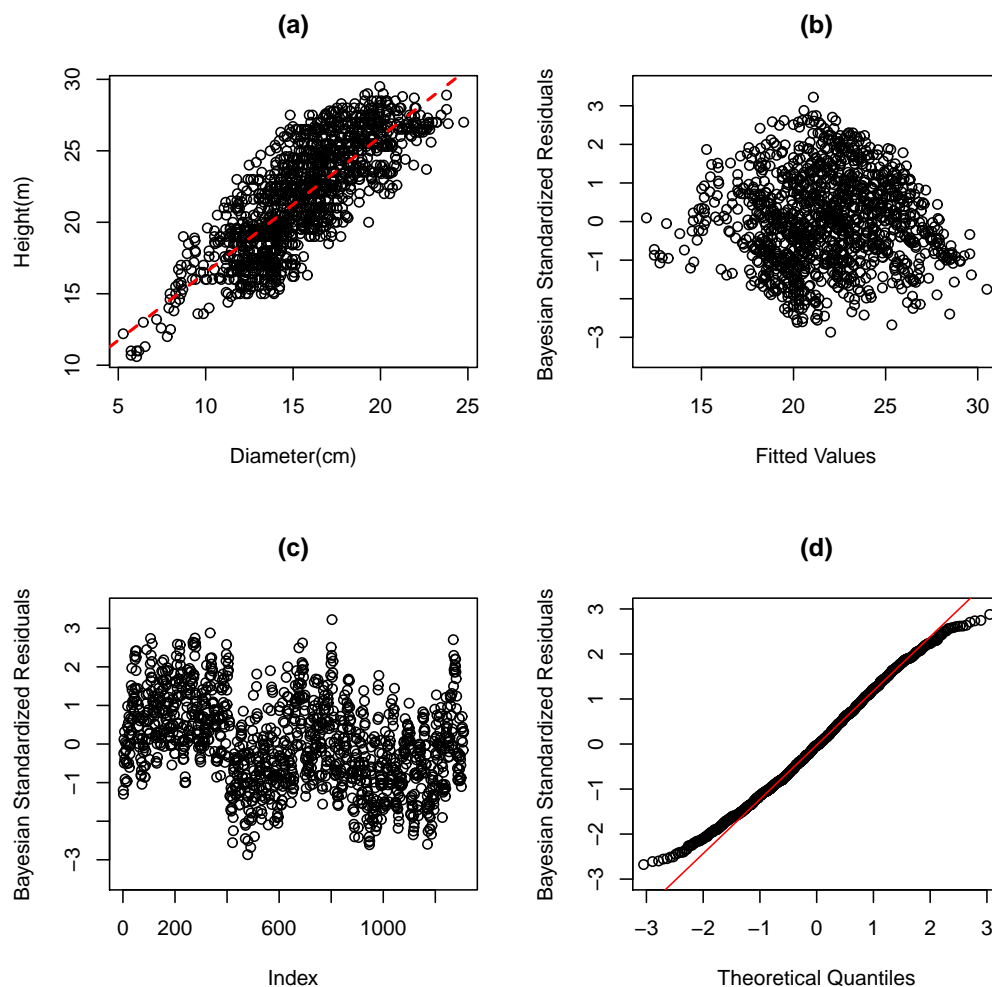


Figure 3. Real data. Normal linear regression model. (a) Scatterplot of the data and fit normal linear regression model. (b) Quantile residuals versus fit values. (c) Quantile residuals versus the index. (d) Graph of the quantiles of the normal distribution for the residuals of the model.

To investigate the influence of height and diameter data from Eucalyptus on the fit of the generalized normal linear regression model chosen, we calculated the K-L divergence measures and their respective calibrations. Figure 4 shows the K-L measurements for each observation. Note that Cases 335, 479, and 803 exhibited higher values of the K-L divergence when compared with other observations. The K-L divergences and calibrations concerning three observations that showed the highest calibration values are presented in Table 5. It can be seen that Observation 803 was possibly an influential case, which was also shown to be an outlier from the visual analysis. To assess whether this observation altered the parameter estimates of the GN linear regression model, we carried out a sensitivity analysis.

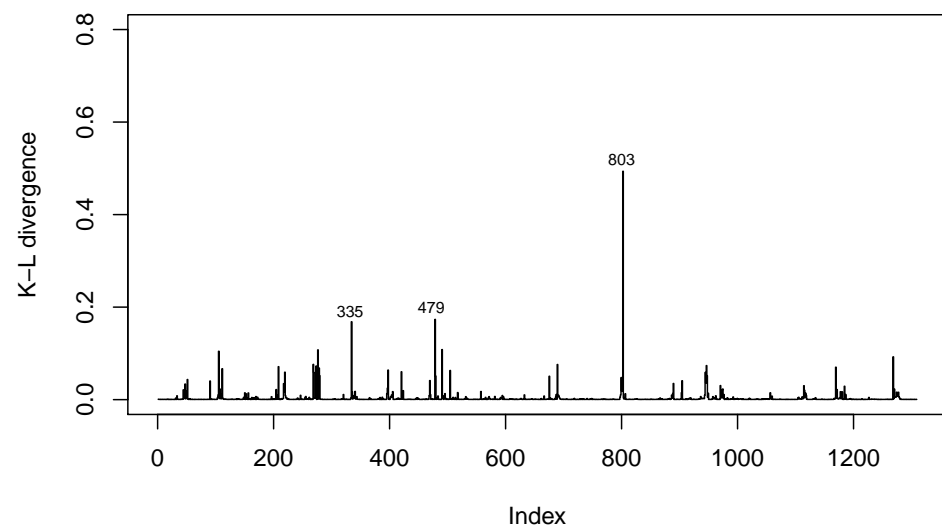


Figure 4. Index plot of $K(\pi, \pi_{(-i)})$ for the height and diameter data of Eucalyptus.

Table 5. Diagnostic measures for the height and diameter data of Eucalyptus.

Case Number	$K(\pi, \pi_{(-i)})$	Calibration
335	0.1683	0.7673
479	0.1734	0.7707
803	0.4936	0.8960

Table 6 shows the new estimates of the model parameters after excluding the case with the greatest calibration value and the relative variations (RV) for these estimates regarding the Eucalyptus data. Here, the relative variations were obtained by $RV = \left| \frac{\hat{\theta} - \hat{\theta}_0}{\hat{\theta}} \right| \times 100\%$, where $\hat{\theta}$ and $\hat{\theta}_0$ are the posterior averages of the model parameters obtained from the original data and from the data without influential observation, respectively. We noted a slight change in the RV of the s parameter when we excluded influential observations. However, such a change was insignificant. This indicated that the GN linear regression model was not affected by the compelling cases.

Table 6. Posterior estimates and $RV(\%)$ for Eucalyptus height and diameter data following the removal of the influential case.

Cases Deletions	Parameter	Mean	$RV(\%)$	95% HDP
803	β_1	7.067	0.014	(6.375; 7.753)
	β_2	0.948	-	(0.905; 0.991)
	σ	3.252	0.463	(3.054; 3.454)
	s	2.936	2.478	(2.486; 3.390)

Overall, regardless of the unit measurement/scale in both cases (synthetic and real data sets), the visual representation corroborated the explainability/adjustment of the GN model, by the quantile-quantile plot, residuals versus fits plot, and standardized errors showing no pattern left to explain equal error variances, as well as no outliers.

9. Discussion

In this paper, we presented the generalized normal linear regression model from objective Bayesian analysis. The Jeffreys prior and reference prior for the generalized normal model were discussed in detail. We proved that the Jeffreys prior leads to an improper posterior distribution and cannot be used in a Bayesian analysis. On the other hand, the reference prior leads to a proper posterior distribution.

The parameter estimates were based on a Bayesian reference analysis procedure via MCMC. Diagnostic techniques based on the Kullback–Leibler divergence were built for the generalized typical linear regression model. Studies with artificial and real data were performed to verify the adequacy of the proposed inferential method.

The result of the application to a set of actual data showed that the generalized normal linear regression model outperformed the normal linear regression model, regardless of the model selection criteria. Furthermore, through a study of artificial data and real data, the Kullback–Leibler divergence effectively detected the points that were influential in the fit of the generalized normal linear regression model. The withdrawal of such influential points from the set of real data showed that the generalized normal model was not affected by influential observations. This result was corroborated by the fact that the generalized normal distribution was considered a tool for reducing outliers and achieving robust estimates. The proposed methodology showed consistent marginalization and sampling properties and thus eliminated the problem of estimating the parameters of this important regression model. Moreover, adopting the reference (objective) prior, we obtained one-to-one consistent results, under the Bayesian paradigm, enabling the GN distribution in a practical form.

Further works can explore a great number of extensions using this study. For instance, the method developed in this article may be applied to other regression models such as the Student *t* regression model and the Birnbaum–Saunders regression model, among others. Additionally, other generalizations of the normal distribution should be considered [36–38].

Author Contributions: Conceptualization, V.L.D.T. and S.R.d.J.; methodology, S.R.d.J. and P.L.R.; software, S.R.d.J.; validation, F.L., F.A.R. and D.C.d.N.; writing—original draft preparation, S.R.d.J., P.L.R., F.L., F.A.R., A.B.G. and D.C.d.N.; writing—review and editing, A.B.G., F.A.R., F.L., S.N.; supervision, F.L., S.N.; project administration, V.L.D.T.; funding acquisition, P.L.R. and D.C.d.N. All authors have read and agreed to the published version of the manuscript.

Funding: Francisco Louzada is supported by the Brazilian agencies CNPq (grant number 301976/2017-1) and FAPESP (grant number 2013/07375-0). Francisco Rodrigues acknowledges financial support from CNPq (grant number 309266/2019-0). Diego Nascimento acknowledges the support from the São Paulo State Research Foundation (FAPESP process 2020/09174-5). Pedro L. Ramos acknowledges support from São Paulo State Research Foundation (FAPESP Proc. 2017/25971-0).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Appendix A

Proof of Corollary 2. Equations (18) and (19) are continuous functions in $[1, \infty]$. When $s \rightarrow \infty$, we have that $\Gamma\left(\frac{1}{s}\right) = O(s)$ and $\Psi'\left(1 + \frac{1}{s}\right) \rightarrow 1.6449$. \square

Proof of Corollary 3. It is known that σ can be obtained analytically. Integrating σ , we obtain the integrated likelihood function for (β, s) ,

$$L(\beta, s|\mathbf{y}) = \int_0^\infty L(\beta, \sigma, s|\mathbf{y})\pi(\sigma)d\sigma \tag{A1}$$

$$\propto s^{-1}\Gamma\left(\frac{n+c-1}{s}\right)\left[\sum_{i=1}^n\left(\Gamma\left(1+\frac{1}{s}\right)|y_i-\mathbf{x}_i^\top\beta|\right)^s\right]^{-\frac{(n+c-1)}{s}}.$$

Considering the likelihood $L(\beta, s|\mathbf{y})$, we integrate β , obtaining:

$$L(s|\mathbf{y}) = \int_{\mathbb{R}^p} L(\beta, s|\mathbf{y})\pi(\beta)d\beta \tag{A2}$$

$$\propto s^{-1}\Gamma\left(\frac{n+c-1}{s}\right)\Gamma\left(1+\frac{1}{s}\right)^{-(n+c-1)}\int_{\mathbb{R}^p}\left[\sum_{i=1}^n|y_i-\mathbf{x}_i^\top\beta|^s\right]^{-\frac{(n+c-1)}{s}}d\beta.$$

Moreover, by Lemma 3.2 in [11], $\int_{\mathbb{R}^p}\left[\sum_{i=1}^n|y_i-\mathbf{x}_i^\top\beta|^s\right]^{-\frac{(n+c-1)}{s}}d\beta$ is limited and of order $n^{-\frac{(n+c-1)}{s}}$ for $n > p + 1 - c$.

Therefore,

$$L(s|\mathbf{y}) \propto s^{-1}\Gamma\left(\frac{n+c-1}{s}\right)\Gamma\left(1+\frac{1}{s}\right)^{-(n+c-1)}O\left(n^{-\frac{(n+c-1)}{s}}\right). \tag{A3}$$

In order to understand the behavior of the integrated likelihood of s , we consider the following result: the expansion of the series $\frac{1}{\Gamma(z)} \approx z$ for z approaching zero [39]. Therefore, if $s \rightarrow \infty$, we have $\Gamma\left(\frac{1}{s}\right) \approx s$. Moreover, considering the expansion of the first-order Taylor series of $\log \Gamma(1+z)$ for values near $z = 0$, it follows that $\log \Gamma(1+z) \approx z\Psi(1)$, where $\Psi(1) \approx -0.57721$. Thus, $\Gamma\left(1+\frac{1}{s}\right) \approx e^{\frac{\Psi(1)}{s}}$ for large values of s . Therefore, for $s \rightarrow \infty$, we have $\Gamma\left(\frac{n+c-1}{s}\right) \approx \frac{s}{n+c-1}$. Therefore:

$$L(s|\mathbf{y}) \approx s^{-1}\frac{s}{n+c-1}\left(e^{\frac{\Psi(1)}{s}}\right)^{-(n+c-1)}O\left(n^{-\frac{(n+c-1)}{s}}\right) \tag{A4}$$

$$\approx e^{-\frac{\Psi(1)(n+c-1)}{s}}O\left(n^{-\frac{(n+c-1)}{s}}\right)$$

$$= O\left(e^{-\frac{(n+c-1)}{s}\{\Psi(1)+\log n\}}\right)$$

$$= O(1).$$

Completing the proof. \square

Proof of Proposition 3. Considering the reference prior given in Corollary 2, the result of Corollary 3, and Condition (20), it follows that the posterior reference distribution is proper if:

$$\int_1^\infty O(1)O\left(s^{-\frac{3}{2}}\right)ds < \infty.$$

Thus,

$$\int_1^\infty O(1)O\left(s^{-\frac{3}{2}}\right)ds = \int_1^\infty s^{-\frac{3}{2}}ds = -2 < \infty. \tag{A5}$$

Therefore, the reference prior leads to a proper posterior distribution.

Considering the Jeffreys prior given in Corollary 2, the result of Corollary 3, and Condition (20), it follows that the Jeffreys posterior distribution is proper if:

$$\int_1^{\infty} O(1)O\left(s^{\frac{p-2}{2}}\right)ds < \infty.$$

Thus,

$$\int_1^{\infty} O(1)O\left(s^{\frac{p-2}{2}}\right)ds = \int_1^{\infty} s^{\frac{p-2}{2}} ds = \infty. \quad (\text{A6})$$

Therefore, the Jeffreys prior returned a posterior that is improper, completing the proof. \square

References

- Subbotin, M.T. On the law of frequency of errors. *Math. Sb.* **1923**, *31*, 296–301.
- Belsley, D.A.; Kuh, E.; Welsch, R.E. *Regression Diagnostic: Identifying Influential Data and Sources of Collinearity*; John Wiley: New York, NY, USA, 1980.
- Barnett, V.; Lewis, T. *Outliers in Statistical Data*; John Wiley: New York, NY, USA, 1994.
- Box, G.E.P.; Tiao, G.C. *Bayesian Inference in Statistical Analysis*; John Wiley & Sons: New York, NY, USA, 1973.
- Walker, S.G.; Gutiérrez-Peña, E. Robustifying Bayesian procedures. In *Bayesian Statistics 6*; Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., Eds.; Oxford University Press: Oxford, UK, 1999; pp. 685–710.
- Liang, F.; Liu, C.; Wang, N. A robust sequential Bayesian method for identification of differentially expressed genes. *Stat. Sin.* **2007**, *17*, 571–595.
- Varanasi, M.k.; Aazhang, B. Parametric generalized Gaussian density estimation. *J. Acoust. Soc. Am.* **1989**, *86*, 1404–1415. [[CrossRef](#)]
- Agro, G. Maximum likelihood estimation for the exponential power function parameters. *Commun. Stat.-Simul. Comput.* **1995**, *24*, 523–536. [[CrossRef](#)]
- West, M. On scale mixtures of normal distributions. *Biometrika* **1987**, *79*, 646–648. [[CrossRef](#)]
- Choy, S.T.B.; Smith, F.M. On Robust Analysis of a Normal Location Parameter. *J. R. Stat. Soc. Ser. B* **1997**, *59*, 463–474. [[CrossRef](#)]
- Salazar, E.; Ferreira, M.A.; Migon, H.S. Objective Bayesian analysis for exponential power regression models. *Sankhya Ser. B* **2012**, *74*, 107–125. [[CrossRef](#)]
- Bernardo, J.M. Reference Analysis. In *Handbook of Statistics 25*; Dey, D., Rao, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2005; Volume 25, pp. 17–90.
- Bernardo, J.M. Reference Posterior Distributions for Bayesian-Inference. *J. R. Stat. Soc. Ser. B-Methodol.* **1979**, *41*, 113–147. [[CrossRef](#)]
- Berger, J.O.; Bernardo, J.M. On the development of reference priors. In *Bayesian Statistics 4*; Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., Eds.; Oxford University Press: Oxford, UK, 1992; pp. 35–60.
- Fonseca, T.C.; Ferreira, M.A.; Migon, H.S. Objective Bayesian analysis for the Student-t regression model. *Biometrika* **2008**, *95*, 325–333. [[CrossRef](#)]
- Northrop, P.J.; Attalides, N. Posterior propriety in Bayesian extreme value analyses using reference priors. *Stat. Sin.* **2016**, *26*, 721–743. [[CrossRef](#)]
- Ramos, P.L.; Louzada, F.; Ramos, E. Posterior Properties of the Nakagami-m Distribution Using Noninformative Priors and Applications in Reliability. *IEEE Trans. Reliab.* **2018**, *67*, 105–117. [[CrossRef](#)]
- Ramos, E.; Ramos, P.L.; Louzada, F. Posterior properties of the Weibull distribution for censored data. *Stat. Probab. Lett.* **2020**, *166*, 108873. [[CrossRef](#)]
- Tomazella, V.L.; de Jesus, S.R.; Louzada, F.; Nadarajah, S.; Ramos, P.L. Reference Bayesian analysis for the generalized lognormal distribution with application to survival data. *Stat. Its Interface* **2020**, *13*, 139–149. [[CrossRef](#)]
- Ramos, P.L.; Mota, A.L.; Ferreira, P.H.; Ramos, E.; Tomazella, V.L.; Louzada, F. Bayesian analysis of the inverse generalized gamma distribution using objective priors. *J. Stat. Comput. Simul.* **2021**, *91*, 786–816. [[CrossRef](#)]
- Nadarajah, S. A generalized normal distribution. *J. Appl. Stat.* **2005**, *32*, 685–694. [[CrossRef](#)]
- Zhu, D.; Zinde-Walsh, V. Properties and estimation of asymmetric exponential power distribution. *J. Econom.* **2009**, *148*, 86–99. [[CrossRef](#)]
- Chen, M.H.; Shao, Q.M.; Ibrahim, J.G. *Monte Carlo methods in Bayesian Computation*; Springer: New York, NY, USA, 2000.
- Zellner, A. Maximal Data Information Prior Distributions. *Basic Issues Econom.* **1984**, 211–232.
- Moala, F.A.; Dey, S. Objective and subjective prior distributions for the Gompertz distribution. *Anais Acad. Bras. Ciências* **2018**, *90*, 2643–2661. [[CrossRef](#)]
- Moala, F.A.; Rodrigues, J.; Tomazella, V.L.D. A note on the prior distributions of weibull parameters for the reliability function. *Commun. Stat. Theory Methods* **2009**, *38*, 1041–1054. [[CrossRef](#)]

27. Muller, P. *A Generic Approach to Posterior Integration and Gibbs Sampling*; Technical Report; Department of Statistics, Purdue University: West Lafayette, IN, USA, 1991.
28. Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*; Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., Eds.; Oxford University Press: Oxford, UK, 1992; pp. 625–631.
29. Spiegelhalter, D.; Best, N.; Carlin, B.; van der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 583–639. [[CrossRef](#)]
30. Brooks, S.P. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002). *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **2002**, *64*, 616–618.
31. Gelfand, A.E.; Dey, D.K.; Chang, H. Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *Bayesian Statistics*; Kotz, Johnson, N.L., Eds.; Oxford University Press: New York, NY, USA, 1992; Volume 4, pp. 147–167.
32. Gelfand, A.E.; Dey, K.D. Bayesian Model Choice: Asymptotics and Exact Calculations. *J. R. Stat. Soc. Ser. B* **1994**, *56*, 501–514. [[CrossRef](#)]
33. Cook, R.D.; Weisberg, S. *Residuals and Influence in Regression*; Chapman and Hall: Boca Raton, FL, USA, 1982.
34. Peng, F.; Dey, D. Bayesian analysis of outlier problems using divergence measures. *Can. J. Stat.* **1995**, *23*, 199–213. [[CrossRef](#)]
35. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
36. Alzaatreh, A.; Aljarrah, M.; Almagambetova, A.; Zakiyeva, N. On the Regression Model for Generalized Normal Distributions. *Entropy* **2021**, *23*, 173. [[CrossRef](#)]
37. García, V.J.; Gómez-Déniz, E.; Vázquez-Polo, F.J. A new skew generalization of the normal distribution: Properties and applications. *Comput. Stat. Data Anal.* **2010**, *54*, 2021–2034. [[CrossRef](#)]
38. Nascimento, D.C.; Ramos, P.L.; Elal-Olivero, D.; Cortes-Araya, M.; Louzada, F. Generalizing the normality: a novel towards different estimation methods for skewed information. *arXiv* **2021**, arXiv:2105.00031.
39. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions*; Dover: New York, NY, USA, 1972; p. 256.