

Article

A New Quantile Regression Model and Its Diagnostic Analytics for a Weibull Distributed Response with Applications

Luis Sánchez ¹, Víctor Leiva ^{2,*}, Helton Saulo ³, Carolina Marchant ^{4,5} and José M. Sarabia ⁶¹ Institute of Statistics, Universidad Austral de Chile, Valdivia 5091000, Chile; luis.sanchez@uach.cl² School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile³ Department of Statistics, Universidade de Brasília, Brasília 70910-900, Brazil; heltonsauro@unb.br⁴ Faculty of Basic Sciences, Universidad Católica del Maule, Talca 3480112, Chile; cmarchant@ucm.cl⁵ ANID-Millennium Science Initiative Program-Millennium Nucleus Center for the Discovery of Structures in Complex Data, Santiago 7820244, Chile⁶ Department of Quantitative Methods, Universidad CUNEF, 28040 Madrid, Spain; josemaria.sarabia@cunef.edu

* Correspondence: victor.leiva@pucv.cl or victorleivasanchez@gmail.com

Abstract: Standard regression models focus on the mean response based on covariates. Quantile regression describes the quantile for a response conditioned to values of covariates. The relevance of quantile regression is even greater when the response follows an asymmetrical distribution. This relevance is because the mean is not a good centrality measure to resume asymmetrically distributed data. In such a scenario, the median is a better measure of the central tendency. Quantile regression, which includes median modeling, is a better alternative to describe asymmetrically distributed data. The Weibull distribution is asymmetrical, has positive support, and has been extensively studied. In this work, we propose a new approach to quantile regression based on the Weibull distribution parameterized by its quantiles. We estimate the model parameters using the maximum likelihood method, discuss their asymptotic properties, and develop hypothesis tests. Two types of residuals are presented to evaluate the model fitting to data. We conduct Monte Carlo simulations to assess the performance of the maximum likelihood estimators and residuals. Local influence techniques are also derived to analyze the impact of perturbations on the estimated parameters, allowing us to detect potentially influential observations. We apply the obtained results to a real-world data set to show how helpful this type of quantile regression model is.

Keywords: likelihood methods; local influence diagnostics; Monte Carlo simulation; R software



Citation: Sánchez, L.; Leiva, V.; Saulo, H.; Marchant, C.; Sarabia, J.M. A New Quantile Regression Model and Its Diagnostic Analytics for a Weibull Distributed Response with Applications. *Mathematics* **2021**, *9*, 2768. <https://doi.org/10.3390/math9212768>

Academic Editor: Stelios Psarakis

Received: 14 October 2021

Accepted: 27 October 2021

Published: 1 November 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/>).

1. Introduction, Motivations, and Outline

1.1. Bibliographical Review

In the context of usual regression, it is common to model the relationship between a response variable and covariates by employing the mean response conditioned to such covariates. In this usual modeling, the normal distribution is often considered. However, there are many real-world phenomena in which the data follow an asymmetrical distribution. In this case, the relation between the response and covariates utilizing the mean is not suitable since it is strongly affected by asymmetry and atypical observations. Another limitation of the usual regression approach is when we are interested in studying other parameters in addition to the mean; see [1,2].

Observations that follow an asymmetrical behavior can come from different models, with the Weibull distribution frequently being considered. This distribution is skewed, has positive support, and possesses two parameters which modify its shape and scale; for a detailed description of its main properties and associated inference, see [3] (pp. 629–666) and [4]. Estimation and testing methodologies based on several data configurations and situations may be found in [5]. In its origins, the Weibull distribution was used to study the

breaking strength of materials; see [6]. Simple parsimonious Weibull models were derived in [7] and applied to fatigue life data of longitudinal elements by considering functional equations, proportional hazards techniques, and subsequent likelihood analysis. Other applications include different areas in problems related to health sciences, lumber industry, microscopic degradation, migratory systems, quality control, rainfall and flood, reliability, and wind speed. Details of these and other applications of the Weibull distribution, as well as data sets described with this distribution, can be found in chapter 7 of [8] (pp. 275–310) and references therein.

Unlike the mean, which can be challenging to interpret when the distribution of the response variable is asymmetrical, the median remains highly informative in that case. Thus, under this scenario, the modeling of the median response based on values of covariates is more appropriate. The first idea of median regression was presented in [9]. However, quantile regression models have the median regression as a particular case (50th percentile) and can describe other locations (non-central) of the distribution. In [10], the authors introduced quantile regression models, and from then, different versions and applications of these models have been developed; see [11–13]. Therefore, to describe the relationship between a response variable that follows an asymmetrical distribution and the covariates, quantile regression is a better alternative to the usual regression.

The standard approach in parametric quantile regression considers a functional equation that relates the response (say Y), a parametric component (say $x^\top \beta$, which corresponds also to the modeled quantile of Y), and an error component (say ε) with its associated assumptions; see [11] (p. 29). The traditional procedure for estimating the model parameters in this approach does not make a distributional assumption for the error component. However, if we add this assumption, it is natural to incorporate it in the response variable rather than in the error component. In addition, the maximum likelihood method is often chosen to estimate parameters because of the good properties of the obtained estimators; see [14] (pp. 94–125). Based on these two previous considerations, a similar approach to generalized linear models (GLM) can be used for quantile regression; see [15–17]. In GLM, the mean is modeled, which is besides one of the parameters of the assumed distribution. In our approach, the modeled quantile is a parameter of the distribution as well. When using a parametric distribution, we can develop statistical analysis employing the likelihood function to perform estimation, hypothesis tests, and local influence analysis.

Diagnostic analytics plays a relevant role in statistical modeling, including global, local influence methods and goodness of fit. Goodness-of-fit techniques for a determined model permit us to evaluate the adequacy of the model to the data; see [18]. The pseudo- R^2 proposed by [19]—from now denoted as R_M^2 —and randomized quantile (RQ) and generalized Cox–Snell (GCS) residuals are helpful tools for evaluating goodness of fit; see [20,21]. Local influence assesses the effect of small perturbations in the data and/or model assumptions on parameter estimates; see [22]. Different scenarios of perturbation are considered to detect potentially influential cases. Local influence techniques have been developed for different non-Gaussian and asymmetrical models; see, for example, refs. [16,17,23–25]. As a motivation to develop our work, next, we show the inadequacy of the usual mean regression when analyzing real-world data with an asymmetrical distribution.

1.2. Limitations of the Usual Regression Model

The usual regression model can be formulated as

$$Y_i = x_i^\top \beta + \varepsilon_i, \quad i \in \{1, \dots, n\}, \quad (1)$$

where Y_i and x_i are the response variable and the vector that contains the values of covariates X (with the first component equal to one), respectively, for the i th observation, and β is a vector of the unknown regression coefficients to be estimated. The errors $\varepsilon_1, \dots, \varepsilon_n$ satisfy (i) $E[\varepsilon_i] = 0$ and $\text{Var}[\varepsilon_i] = \sigma^2$, for all $i \in \{1, \dots, n\}$; and (ii) $\text{Cov}[\varepsilon_j, \varepsilon_l] = 0$, for $j \neq l$. Observe that the structural component formulated in (1) describes the mean $E[Y | X = x] = x^\top \beta$.

When the data follow a skew distribution, the mean model is not appropriate. To demonstrate this fact, consider a data set with $n = 41$ observations regarding the time (in hours) to electrical breakdown of an insulating fluid (response variable Y) and the test voltage in kV (covariate X). This data set is taken from [26] and is available in the R software by the package survival; see [27,28]. The characteristics of the insulating fluid defined in various standards can be broadly classified into chemical, electrical, and physical features. For example, the electrical characteristics (breakdown voltages) of the insulating fluid are affected by elements such as water content and electrostatic charges, but also possibly affected by trace components in this fluid.

A descriptive summary of the times to electrical breakdown is presented in Table 1, including the median, mean, standard deviation (SD), coefficients of variation (CV), skewness (CS), and kurtosis (CK), besides minimum ($y_{(1)}$) and maximum ($y_{(n)}$) values. Figure 1a presents a histogram for Y , and Figure 1b shows the corresponding adjusted and usual boxplots. An adjusted boxplot is used when the data present an asymmetrical distribution; see details in [29]. In this case, the adjusted boxplot gives a better description to detect atypical cases.

Table 1. Descriptive statistics for the data of times to electrical breakdown (in hours).

| Median | Mean | SD | CV | CS | CK | $y_{(1)}$ | $y_{(n)}$ | n |
|--------|--------|--------|------|------|-------|-----------|-----------|-----|
| 7.7400 | 122.51 | 430.24 | 3.51 | 4.36 | 20.93 | 0.09 | 2323.70 | 41 |

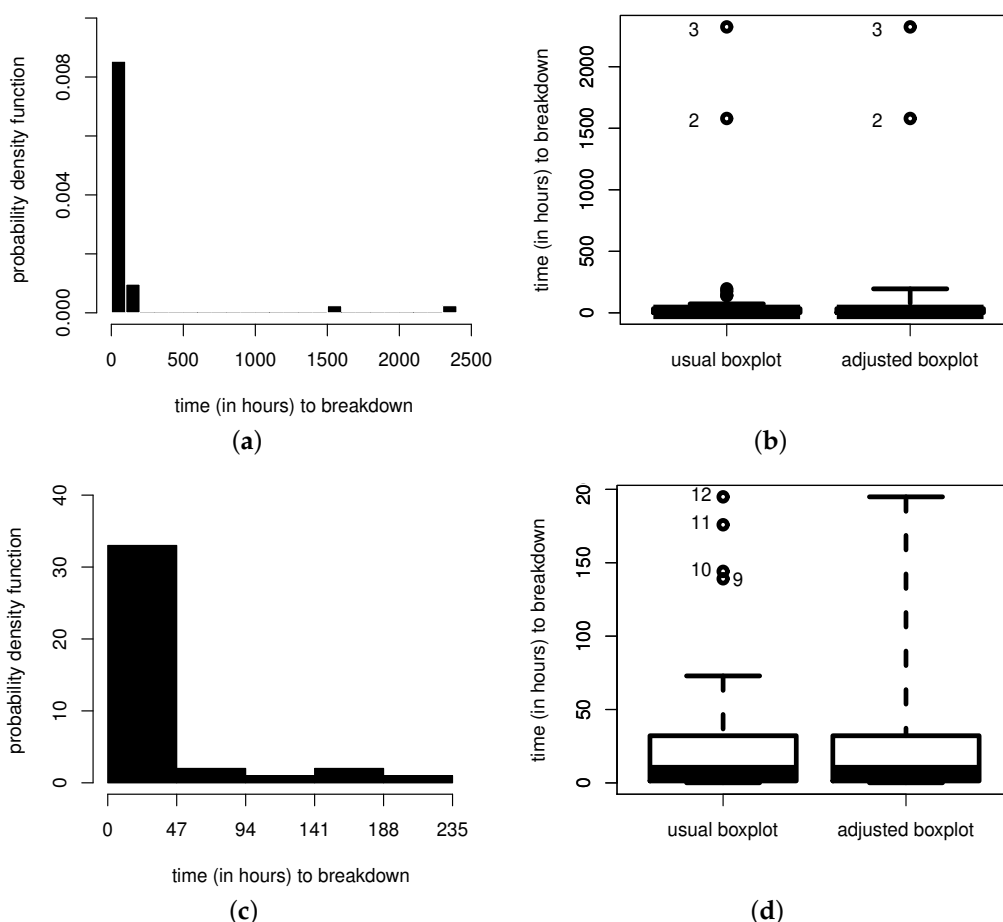


Figure 1. Histogram (a) and boxplots (b) for the data of times to electrical breakdown with the full data set, and histogram (c) and boxplots (d) for the data set without cases #2 and #3.

From Table 1, note that the median is noticeably smaller than the mean, whereas Figure 1a allows us to observe that the empirical distribution of the times to electrical breakdown is unimodal and positively skewed. Therefore, the assumption of an asymmetric distribution for the response variable seems to be adequate. This asymmetry is also evidenced by the values of the CS, which is positive. Furthermore, in Figure 1b, we highlight two atypical cases (#2 and #3), which can correspond to potentially influential cases. The possible potential influence of these and other cases is analyzed by using local influence in Section 6.2. In Figure 1c, we observe the empirical distribution of Y without cases #2 and #3, whereas the boxplots associated are displayed in Figure 1d. Note that the asymmetrical behavior of the data is kept. However, now the adjusted boxplot does not present atypical cases despite the usual boxplots identify some of them.

Next, the model stated in (1) is adjusted to this data set employing the ordinary least squares method. Then, we obtain the predictive model $\hat{y}_i = \hat{E}(Y_i | X = x_i) = 2274.12 - 64.96 x_i$, for $i \in \{1, \dots, 41\}$. The fit of the model is evaluated by the usual standardized Pearson residual, presented in the theoretical quantile versus empirical quantile (QQ) plot with envelopes in Figure 2. Note that the points follow an irregular behavior around the straight line, and many observations are outside the bands. Hence, it is not clear that the usual regression is appropriate for modeling this data set due to the asymmetry of the response distribution.

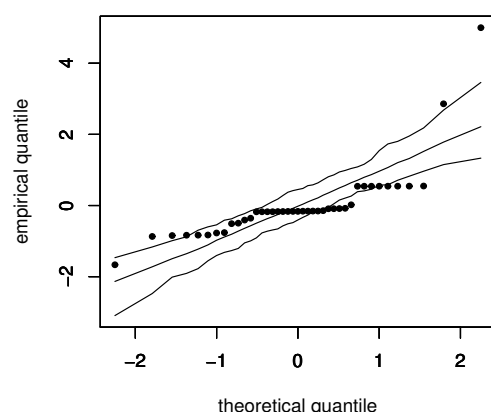


Figure 2. QQ plot with envelopes of the Pearson residual for normal regression with the data of times to electrical breakdown.

For these data (full set and set without cases #2 and #3), we may assume an asymmetrically distributed response. In addition, in this case, the modeling of the conditional median is a better alternative for describing the relation of the response with the covariates (as we show below) because the median is a robust measure in the presence of atypical observations. However, the median is a quantile and, in consequence, a description of the full range of the response based on covariates can be performed by using quantile regression.

1.3. Objective and Outline

The main objective of this work is to propose a new quantile regression model based on a parameterization of the Weibull distribution, following the approach of [16]; see [30,31] for similar but not identical models. Our approach intends to be an alternative to the existing quantile models in the literature. Some characteristics of the proposed Weibull quantile regression are as follows: (i) flexibility for modeling different types of data, since the Weibull distribution, as mentioned, has been successfully applied in several areas; and (ii) easy computational implementation, since the Weibull distribution has a simple closed-form inverse cumulative distribution function, which facilitates its utilization when modeling data by a parametric quantile regression with distributional assumption for the response.

The maximum likelihood method is used for model parameter estimation. Our study includes the evaluation of the adequacy of the models to the data by Akaike (AIC), Bayesian (BIC), and corrected Akaike (CAIC) information criteria; see [1] for details. In addition, R_M^2 as well as RQ and GCS residuals are considered in this evaluation. We identify potentially influential observations under different scenarios of perturbation employing local influence techniques; see [22]. Moreover, an application to real-world data is discussed to illustrate the proposed methodology and show how helpful this type of quantile regression model is in practice.

The rest of this paper proceeds as follows. The new parametric quantile regression model based on the Weibull distribution is formulated in Section 2. In contrast, in Section 3, we describe the parameter estimation method, associated inference, and the related RQ and GCS residuals to evaluate the fit of the model to the data. In Section 4, two Monte Carlo simulation studies are conducted to evaluate the statistical performance of the maximum likelihood estimators and the empirical distribution of residuals. In Section 5, we propose techniques to study potentially influential cases by using local influence and four perturbation schemes. In Section 6, an illustration of the proposed Weibull quantile regression models is carried out for the same real-world data set presented in Section 1. Finally, in Section 7, we present some concluding remarks.

2. A New Weibull Quantile Regression Model

2.1. A Reparameterized Weibull Distribution

The probability density function of a random variable Y that follows a Weibull distribution with shape and scale parameters $k > 0$ and $\lambda > 0$, respectively, is given by

$$f(y; \lambda, k) = \frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{y}{\lambda}\right)^k\right), \quad y > 0. \quad (2)$$

It is possible to prove that, if $q \in (0, 1)$ is a fixed number, the q th quantile of Y corresponds to

$$Q = \lambda (-\log(1 - q))^{1/k},$$

from which we obtain

$$\lambda = Q (-\log(1 - q))^{-1/k}. \quad (3)$$

For more details about properties of the Weibull distribution, see [3] (pp. 629–666) and [8]. Replacing the formula stated in (3) for λ in the expression given in (2), we have a new parameterization of the Weibull distribution based on its quantiles, which is denoted by $\text{Wei}(Q, k)$, and its probability density and cumulative distribution functions are formulated, respectively, as

$$f(y; Q, k) = -k y^{k-1} Q^{-k} \log(1 - q) \exp\left(y^k Q^{-k} \log(1 - q)\right), \quad y > 0, \quad (4)$$

and

$$F(y; Q, k) = 1 - (1 - q)^{y^k Q^{-k}}, \quad y > 0. \quad (5)$$

2.2. Shape Analysis

Figure 3 shows the behavior of the reparameterized Weibull probability density function defined in (4) under different values of the parameters. Note that, as Q decreases, the kurtosis of the distribution increases; see Figure 3d–i. Thus, when Q increases, the tails are heavier. Moreover, observe in Figure 3a–c that, when k takes values less or equal to one, the distribution mode is zero, while if it takes values greater than one, this mode is positive.

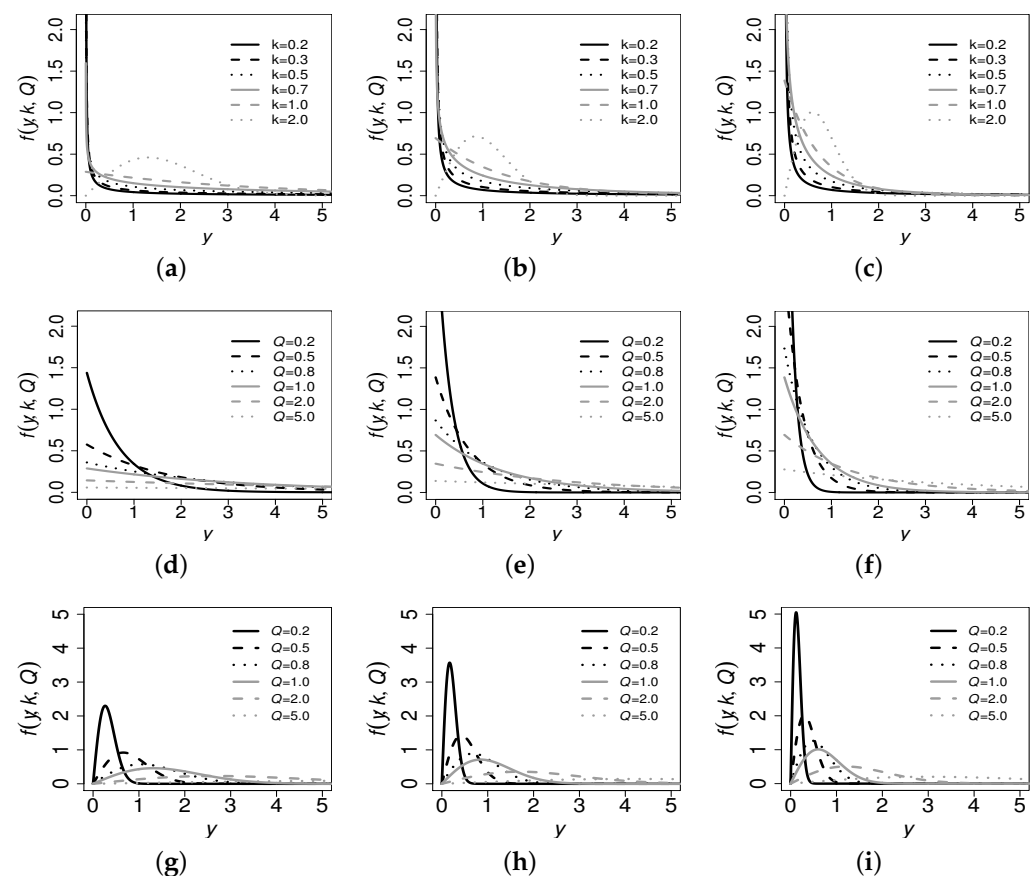


Figure 3. Plots of the $\text{Wei}(Q, k)$ probability density function for $q = 0.25$ (left), $q = 0.5$ (center) and $q = 0.75$ (right), with $Q = 1.0$ (a–c), $k = 1.0$ (d–f) and $k = 2.0$ (g–i).

2.3. The Weibull Quantile Regression Model

Let Y_1, \dots, Y_n be independent random variables with $Y_i \sim \text{Wei}(Q_i, k)$, for $i \in \{1, \dots, n\}$. Suppose that the quantile parameter Q_i can be modeled by

$$h(Q_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i \in \{1, \dots, n\}, \quad (6)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^\top$, for $p < n$, is a vector of unknown regression parameters and $\mathbf{x}_i^\top = (1, x_{i1}, \dots, x_{i(p-1)})$ represents the values of p covariates. Note that the link function h is invertible, at least twice differentiable, and has positive support. The last condition of h guarantees that the quantile is modeled for a positive expression. Link functions that may be considered are, for example, $h(u) = \log_k(u)$ and $h(u) = \sqrt[a]{u}$, with $a \geq 2$ and k being a positive integer number.

Note that the reparametrization of the Weibull distribution by quantiles is necessary to formulate the Weibull quantile regression defined in (6), which allows us to model any quantile value of the distribution. Furthermore, this reparameterization makes it possible to incorporate directly the regression structure given in (6) into the corresponding likelihood function. Note that this structure is different from the traditional quantile regression model with an error component; see [11] (p. 29). Doing that, as mentioned, the distributional assumption is directly related to the response variable, permitting statistical tools based on the associated likelihood function to be obtained in a similar form to GLM.

3. Estimation, Inference and Goodness of Fit

3.1. Parameter Estimation

Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ be an observation of $(Y_1, \dots, Y_n)^\top$, with $Y_i \sim \text{Wei}(Q_i, k)$, for $i \in \{1, \dots, n\}$. The log-likelihood function of the model given in (6) for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, k)^\top$ based on \mathbf{y} can be written as

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n \ell_i(Q_i, k), \quad (7)$$

where $\ell_i(Q_i, k)$ stated in (7) is formulated as

$$\ell_i(Q_i, k) = \log(-\log(1-q)) + \log(k) + (k-1)\log(y_i) - k\log(Q_i) + y_i^k Q_i^{-k} \log(1-q).$$

Therefore, the score vector has as components $\dot{\ell}_{\beta_j}$, for $j \in \{0, 1, \dots, p-1\}$, and $\dot{\ell}_k$, expressed as

$$\dot{\ell}_{\beta_j} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^n z_i a_i x_{ij}, \quad \dot{\ell}_k = \frac{\partial \ell(\boldsymbol{\theta})}{\partial k} = \sum_{i=1}^n b_i, \quad (8)$$

where

$$\begin{aligned} z_i &= -\frac{k}{Q_i} - k Q_i^{-k-1} y_i^k \log(1-q), \\ a_i &= \frac{1}{h'(Q_i)}, \quad h'(Q_i) = \frac{dh}{dQ_i}, \\ b_i &= \log(y_i) - \log(Q_i) + \frac{1}{k} + \left(\frac{y_i}{Q_i}\right)^k \log\left(\frac{y_i}{Q_i}\right) \log(1-q). \end{aligned}$$

The elements of the associated Hessian matrix are written as

$$\begin{aligned} \ddot{\ell}_{\beta_i \beta_j} &= \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_i \partial \beta_j} = \sum_{i=1}^n c_i x_{ij} x_{il}, \\ \ddot{\ell}_{\beta_j k} &= \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial k} = \sum_{i=1}^n m_i a_i x_{ij}, \\ \ddot{\ell}_{kk} &= \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial k^2} = \sum_{i=1}^n d_i, \end{aligned} \quad (9)$$

where

$$\begin{aligned} c_i &= \left(\frac{k}{Q_i^2} + k(k+1) Q_i^{-k-2} y_i^k \log(1-q) \right) a_i^2 - z_i a_i \frac{h''(Q_i)}{(h'(Q_i))^2}, \quad h''(Q_i) = \frac{d^2 h}{dQ_i^2}, \\ m_i &= -\frac{1}{Q_i} - y_i^k Q_i^{-k-1} \log(1-q) \left(1 - k \log\left(\frac{Q_i}{y_i}\right) \right), \\ d_i &= -\frac{1}{k^2} + \log(1-q) \left(\frac{y_i}{Q_i} \right)^k \log^2\left(\frac{y_i}{Q_i}\right). \end{aligned}$$

To estimate the vector $\boldsymbol{\theta}$ of parameters with the maximum likelihood method, we often solve the equation $\dot{\ell}_{\boldsymbol{\theta}} = \mathbf{0}_{p+1}$, where $\mathbf{0}_{p+1}$ is the $p+1$ null vector. However, no closed-form expressions for the maximum likelihood estimates can be obtained, and therefore numeric procedures must be used to calculate the estimate of $\boldsymbol{\theta}$. For example, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method or other quasi-Newton algorithms may be considered; see [32]. Different algorithms are implemented in the R software, including the BFGS approach for constrained and unconstrained maximization; see [27].

3.2. Inference and Hypothesis Testing

Under some regularity conditions [14] (pp. 118–119), it is possible to establish that

$$\hat{\theta} \sim N_{p+1}(\theta, (I(\theta))^{-1}), \quad (10)$$

where $I(\theta)$ is the expected Fisher information matrix, which may be computed by

$$I(\theta) = E \left[-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\top} \right]. \quad (11)$$

We can obtain approximate confidence intervals using the results provided in (10), whereas for approximating the information matrix defined in (11), we may employ the observed Fisher information matrix stated as

$$J(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\top}, \quad (12)$$

whose elements established in (12) may be calculated from (9), evaluated at $\theta = \hat{\theta}$.

Note that if we want to test the hypothesis $H_0: \theta = \theta_0$ versus the alternative hypothesis $H_1: \theta \neq \theta_0$, where, as mentioned, $\theta = (\beta^\top, k)^\top$, then we can use the Wald and likelihood ratio tests. The Wald [33] and likelihood ratio statistics based on the observed Fisher information matrix [34] are, respectively, given by

$$W = (\hat{\theta} - \theta_0)^\top J(\hat{\theta})(\hat{\theta} - \theta_0), \quad (13)$$

$$L = -2(\ell(\theta_0) - \ell(\hat{\theta})). \quad (14)$$

When $n \rightarrow \infty$, both statistics converge to a random variable that follows a χ^2 distribution with r degrees of freedom, χ_r^2 in short, where r is the number of parameters under H_0 , which is rejected, at a nominal level of significance α , if the statistic computed according to (13) or (14) is greater than $\chi_{r,1-\alpha}^2$, which denotes the $100(1 - \alpha)$ th χ_r^2 quantile.

3.3. Residuals

To evaluate the model adequacy—that is, to assess the fit of our model to a data set—we can employ the RQ and GCS residuals. For our reparameterized Weibull model, these residuals are given, respectively, by

$$r_i^{\text{RQ}} = \Phi^{-1}(F(y_i; \hat{Q}_i; \hat{k})), \quad r_i^{\text{GCS}} = -\log(S(y_i; \hat{Q}_i; \hat{k})), \quad (15)$$

where Φ is the standard normal cumulative distribution function; F is given by (5); \hat{Q}_i and \hat{k} are the maximum likelihood estimates of Q_i and k , respectively; and $S = 1 - F$ is the corresponding survival function. The RQ residual is approximately standard normal distributed, whereas the GCS residual follows a standard exponential asymptotic distribution when the model is correctly specified, whatever its specification is.

4. Monte Carlo Simulation

4.1. Setting

We present the results of two Monte Carlo simulation studies for the Weibull quantile regression model. The first scenario considers the evaluation of the statistical performance of the maximum likelihood estimators, while the second scenario assesses the empirical distribution of the residuals. Both simulation scenarios consider the following setting: sample size $n \in \{50, 200, 600\}$, and combinations of the vector of true parameters stated as $(\beta_0, \beta_1, k) = (0.50, 1.00, 0.50)$, $(\beta_0, \beta_1, k) = (1.00, 0.50, 1.00)$, $(\beta_0, \beta_1, k) = (1.00, 0.50, 2.00)$, $(\beta_0, \beta_1, k) = (2.50, 1.00, 0.50)$, $(\beta_0, \beta_1, k) = (2.50, 1.00, 1.00)$, $(\beta_0, \beta_1, k) = (2.50, 1.00, 2.00)$, including different degrees of asymmetry; and $q \in \{0.10, 0.50, 0.90\}$, with 1000 Monte Carlo

replications for each n . The Weibull quantile regression samples are generated using the inverse transformation method applied to the expression formulated in (5), which gives

$$Y_i = (\log(1 - U_i)/\log(1 - q)Q_i)^{1/k}, \quad i \in \{1, \dots, n\}, \quad (16)$$

where Q_i and U_i defined in (16) are specified as $Q_i = \exp(\beta_0 + \beta_1 x_i)$ and $U_i \sim \text{Uniform}(0, 1)$, with x_i being the value of a covariate obtained from a standard normal distribution.

4.2. Scenario 1: Maximum Likelihood Estimation

We employ the R software and its `maxBFGS` function, which implements the BFGS algorithm with constraints for maximization and requires initial values for estimating $\beta = (\beta_0, \beta_1)^\top$ and k . We utilize the least square estimator of β assuming a usual linear regression and the maximum likelihood estimate of k based on the observations y_1, \dots, y_n without considering covariates. The maximum likelihood estimates are presented in Tables 2–7 wherein the empirical mean, bias, variance, root mean squared error (RMSE), CS, and CK are all reported. A look at the results in Tables 2–7 allows us to conclude that, in general, as the sample size increases, the bias, variance, and RMSE of the estimators decrease, as expected. Moreover, $\hat{\beta}_0$, $\hat{\beta}_1$, and \hat{k} seem all to be consistent and asymptotically normal distributed. Our study was conducted on a Dell Inspiron 5748 personal computer with an Intel core i7-4510U CPU, 2.00 GHz \times 4, and 8 GB of RAM.

4.3. Scenario 2: Empirical Distribution of the Residuals

Now, we report a second Monte Carlo simulation study to evaluate the performance of the GCS and RQ residuals defined in (15). Tables 8 and 9 present the empirical mean, SD, CS, and CK for $\beta_0 = 0.5$ and $\beta_1 = 1.0$, whose values are expected to be 0, 1, 0, and 3, respectively, for r^{RQ} , and 1, 1, 2, and 9, respectively, for r^{GCS} . From Tables 8 and 9, we observe that, in general, the considered residuals conform well with the reference distributions. The same conclusions are obtained for the other values of β_0 and β_1 .

Table 2. Statistics from simulated Weibull regression data ($q = 0.10, \beta_0 = 0.50, \beta_1 = 1.00$).

| Statistic | $k = 0.5$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |
| Mean | 0.6011 | 0.5491 | 0.5138 | 0.5506 | 0.5244 | 0.5069 | 0.5253 | 0.5122 | 0.5034 |
| Bias | 0.1011 | 0.0491 | 0.0138 | 0.0506 | 0.0244 | 0.0069 | 0.0253 | 0.0122 | 0.0034 |
| Variance | 0.6747 | 0.1746 | 0.0537 | 0.1687 | 0.0437 | 0.0134 | 0.0422 | 0.0109 | 0.0034 |
| RMSE | 0.8276 | 0.4207 | 0.2322 | 0.4138 | 0.2104 | 0.1161 | 0.2069 | 0.1052 | 0.0581 |
| CS | −0.1288 | −0.1332 | −0.1183 | −0.1287 | −0.1331 | −0.1179 | −0.1286 | −0.1327 | −0.1180 |
| CK | 3.1481 | 3.0124 | 2.9364 | 3.1475 | 3.0105 | 2.9359 | 3.1476 | 3.0094 | 2.9360 |
| | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | |
| True value | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 1.0193 | 0.9831 | 0.9954 | 1.0097 | 0.9915 | 0.9977 | 1.0048 | 0.9958 | 0.9988 |
| Bias | 0.0193 | −0.0169 | −0.0046 | 0.0097 | −0.0085 | −0.0023 | 0.0048 | −0.0042 | −0.0012 |
| Variance | 0.8966 | 0.2356 | 0.0745 | 0.2241 | 0.0589 | 0.0186 | 0.0560 | 0.0147 | 0.0047 |
| RMSE | 0.9471 | 0.4856 | 0.2730 | 0.4735 | 0.2428 | 0.1365 | 0.2368 | 0.1214 | 0.0682 |
| CS | 0.0619 | −0.0344 | 0.1067 | 0.0621 | −0.0347 | 0.1068 | 0.0621 | −0.0348 | 0.1067 |
| CK | 2.8443 | 3.0606 | 3.0311 | 2.8454 | 3.0607 | 3.0311 | 2.8440 | 3.0633 | 3.0311 |
| | \hat{k} | | | \hat{k} | | | \hat{k} | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 1.0000 | 1.0000 | 1.0000 | 2.0000 | 2.0000 | 2.0000 |
| Mean | 0.5210 | 0.5061 | 0.5021 | 1.0419 | 1.0122 | 1.0043 | 2.0838 | 2.0244 | 2.0086 |
| Bias | 0.0210 | 0.0061 | 0.0021 | 0.0419 | 0.0122 | 0.0043 | 0.0838 | 0.0244 | 0.0086 |
| Variance | 0.0036 | 0.0008 | 0.0003 | 0.0144 | 0.0033 | 0.0010 | 0.0576 | 0.0130 | 0.0041 |
| RMSE | 0.0636 | 0.0292 | 0.0162 | 0.1271 | 0.0584 | 0.0324 | 0.2543 | 0.1168 | 0.0648 |
| CS | 0.5824 | 0.2446 | 0.0840 | 0.5826 | 0.2450 | 0.0840 | 0.5831 | 0.2447 | 0.0841 |
| CK | 3.7567 | 2.9277 | 2.6255 | 3.7563 | 2.9247 | 2.6253 | 3.7577 | 2.9246 | 2.6250 |

Table 3. Statistics from simulated Weibull regression data ($q = 0.50, \beta_0 = 0.50, \beta_1 = 1.00$).

| Statistic | $k = 0.5$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |
| Mean | 0.4963 | 0.5152 | 0.5015 | 0.4982 | 0.5076 | 0.5008 | 0.4991 | 0.5038 | 0.5004 |
| Bias | −0.0037 | 0.0152 | 0.0015 | −0.0018 | 0.0076 | 0.0008 | −0.0009 | 0.0038 | 0.0004 |
| Variance | 0.3423 | 0.0885 | 0.0261 | 0.0856 | 0.0221 | 0.0065 | 0.0214 | 0.0055 | 0.0016 |
| RMSE | 0.5851 | 0.2978 | 0.1615 | 0.2925 | 0.1489 | 0.0808 | 0.1463 | 0.0745 | 0.0404 |
| CS | −0.2084 | −0.1717 | −0.1039 | −0.2085 | −0.1716 | −0.1039 | −0.2084 | −0.1715 | −0.1038 |
| CK | 3.0076 | 3.1321 | 2.9602 | 3.0078 | 3.1320 | 2.9601 | 3.0075 | 3.1319 | 2.9600 |
| | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | |
| True value | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 1.0194 | 0.9831 | 0.9954 | 1.0097 | 0.9916 | 0.9977 | 1.0049 | 0.9958 | 0.9988 |
| Bias | 0.0194 | −0.0169 | −0.0046 | 0.0097 | −0.0084 | −0.0023 | 0.0049 | −0.0042 | −0.0012 |
| Variance | 0.8965 | 0.2355 | 0.0745 | 0.2241 | 0.0589 | 0.0186 | 0.0560 | 0.0147 | 0.0047 |
| RMSE | 0.9470 | 0.4856 | 0.2730 | 0.4735 | 0.2428 | 0.1365 | 0.2368 | 0.1214 | 0.0683 |
| CS | 0.0619 | −0.0343 | 0.1067 | 0.0619 | −0.0343 | 0.1068 | 0.0620 | −0.0344 | 0.1067 |
| CK | 2.8447 | 3.0612 | 3.0311 | 2.8448 | 3.0612 | 3.0309 | 2.8450 | 3.0609 | 3.0310 |
| | \hat{k} | | | \hat{k} | | | \hat{k} | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 1.0000 | 1.0000 | 1.0000 | 2.0000 | 2.0000 | 2.0000 |
| Mean | 0.5210 | 0.5061 | 0.5021 | 1.0419 | 1.0122 | 1.0043 | 2.0838 | 2.0243 | 2.0086 |
| Bias | 0.0210 | 0.0061 | 0.0021 | 0.0419 | 0.0122 | 0.0043 | 0.0838 | 0.0243 | 0.0086 |
| Variance | 0.0036 | 0.0008 | 0.0003 | 0.0144 | 0.0033 | 0.0010 | 0.0576 | 0.0130 | 0.0041 |
| RMSE | 0.0636 | 0.0292 | 0.0162 | 0.1271 | 0.0584 | 0.0324 | 0.2543 | 0.1168 | 0.0648 |
| CS | 0.5826 | 0.2448 | 0.0841 | 0.5825 | 0.2448 | 0.0840 | 0.5824 | 0.2448 | 0.0840 |
| CK | 3.7568 | 2.9256 | 2.6256 | 3.7565 | 2.9256 | 2.6254 | 3.7559 | 2.9256 | 2.6255 |

Table 4. Statistics from simulated Weibull regression data ($q = 0.90, \beta_0 = 0.50, \beta_1 = 1.00$).

| Statistic | $k = 0.5$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |
| Mean | 0.4295 | 0.4939 | 0.4937 | 0.4648 | 0.4969 | 0.4969 | 0.4824 | 0.4985 | 0.4984 |
| Bias | −0.0705 | −0.0061 | −0.0063 | −0.0352 | −0.0031 | −0.0031 | −0.0176 | −0.0015 | −0.0016 |
| Variance | 0.3075 | 0.0794 | 0.0235 | 0.0769 | 0.0198 | 0.0059 | 0.0192 | 0.0050 | 0.0015 |
| RMSE | 0.5590 | 0.2818 | 0.1534 | 0.2795 | 0.1409 | 0.0767 | 0.1397 | 0.0704 | 0.0384 |
| CS | −0.1501 | −0.1109 | −0.1234 | −0.1505 | −0.1108 | −0.1234 | −0.1504 | −0.1109 | −0.1234 |
| CK | 2.9205 | 3.1507 | 2.9856 | 2.9191 | 3.1504 | 2.9857 | 2.9190 | 3.1504 | 2.9857 |
| | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | |
| True value | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 1.0194 | 0.9831 | 0.9953 | 1.0097 | 0.9915 | 0.9977 | 1.0048 | 0.9958 | 0.9988 |
| Bias | 0.0194 | −0.0169 | −0.0047 | 0.0097 | −0.0085 | −0.0023 | 0.0048 | −0.0042 | −0.0012 |
| Variance | 0.8965 | 0.2355 | 0.0745 | 0.2241 | 0.0589 | 0.0186 | 0.0560 | 0.0147 | 0.0047 |
| RMSE | 0.9470 | 0.4856 | 0.2730 | 0.4735 | 0.2428 | 0.1365 | 0.2368 | 0.1214 | 0.0682 |
| CS | 0.0617 | −0.0343 | 0.1067 | 0.0620 | −0.0343 | 0.1067 | 0.0619 | −0.0342 | 0.1067 |
| CK | 2.8453 | 3.0612 | 3.0310 | 2.8448 | 3.0612 | 3.0311 | 2.8447 | 3.0609 | 3.0311 |
| | \hat{k} | | | \hat{k} | | | \hat{k} | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 1.0000 | 1.0000 | 1.0000 | 2.0000 | 2.0000 | 2.0000 |
| Mean | 0.5210 | 0.5061 | 0.5021 | 1.0419 | 1.0122 | 1.0043 | 2.0838 | 2.0243 | 2.0086 |
| Bias | 0.0210 | 0.0061 | 0.0021 | 0.0419 | 0.0122 | 0.0043 | 0.0838 | 0.0243 | 0.0086 |
| Variance | 0.0036 | 0.0008 | 0.0003 | 0.0144 | 0.0033 | 0.0010 | 0.0576 | 0.0130 | 0.0041 |
| RMSE | 0.0636 | 0.0292 | 0.0162 | 0.1271 | 0.0584 | 0.0324 | 0.2543 | 0.1168 | 0.0648 |
| CS | 0.5825 | 0.2448 | 0.0840 | 0.5825 | 0.2449 | 0.0840 | 0.5825 | 0.2447 | 0.0840 |
| CK | 3.7567 | 2.9257 | 2.6254 | 3.7567 | 2.9258 | 2.6255 | 3.7566 | 2.9257 | 2.6254 |

Table 5. Statistics from simulated Weibull regression data ($q = 0.10, \beta_0 = 1.00, \beta_1 = 2.50$).

| Statistic | $k = 0.5$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | |
| True value | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 1.1013 | 1.0493 | 1.0138 | 1.0506 | 1.0244 | 1.0069 | 1.0253 | 1.0122 | 1.0034 |
| Bias | 0.1013 | 0.0493 | 0.0138 | 0.0506 | 0.0244 | 0.0069 | 0.0253 | 0.0122 | 0.0034 |
| Variance | 0.6747 | 0.1747 | 0.0537 | 0.1687 | 0.0437 | 0.0134 | 0.0422 | 0.0109 | 0.0034 |
| RMSE | 0.8276 | 0.4208 | 0.2322 | 0.4138 | 0.2104 | 0.1161 | 0.2069 | 0.1052 | 0.0581 |
| CS | −0.1295 | −0.1353 | −0.1183 | −0.1290 | −0.1330 | −0.1180 | −0.1292 | −0.1332 | −0.1185 |
| CK | 3.1489 | 3.0134 | 2.9364 | 3.1480 | 3.0104 | 2.9358 | 3.1485 | 3.0116 | 2.9364 |
| | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | |
| True value | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 |
| Mean | 2.5193 | 2.4830 | 2.4954 | 2.5097 | 2.4915 | 2.4977 | 2.5048 | 2.4958 | 2.4989 |
| Bias | 0.0193 | −0.0170 | −0.0046 | 0.0097 | −0.0085 | −0.0023 | 0.0048 | −0.0042 | −0.0011 |
| Variance | 0.8962 | 0.2358 | 0.0745 | 0.2241 | 0.0589 | 0.0186 | 0.0560 | 0.0147 | 0.0047 |
| RMSE | 0.9469 | 0.4859 | 0.2730 | 0.4735 | 0.2428 | 0.13657 | 0.2368 | 0.1214 | 0.0682 |
| CS | 0.0622 | −0.0376 | 0.1068 | 0.0619 | −0.0346 | 0.1066 | 0.0621 | −0.0336 | 0.1070 |
| CK | 2.8454 | 3.0685 | 3.0310 | 2.8447 | 3.0617 | 3.0290 | 2.8452 | 3.0608 | 3.0310 |
| | \hat{k} | | | \hat{k} | | | \hat{k} | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 1.0000 | 1.0000 | 1.0000 | 2.0000 | 2.0000 | 2.0000 |
| Mean | 0.5210 | 0.5061 | 0.5021 | 1.0419 | 1.0122 | 1.0043 | 2.0838 | 2.0243 | 2.0086 |
| Bias | 0.0210 | 0.0061 | 0.0021 | 0.0419 | 0.0122 | 0.0043 | 0.0838 | 0.0243 | 0.0086 |
| Variance | 0.0036 | 0.0008 | 0.0003 | 0.0144 | 0.0033 | 0.0010 | 0.0576 | 0.0130 | 0.0041 |
| RMSE | 0.0636 | 0.0292 | 0.0162 | 0.1271 | 0.0584 | 0.0324 | 0.2543 | 0.1168 | 0.0648 |
| CS | 0.5824 | 0.2461 | 0.0840 | 0.5826 | 0.2453 | 0.0829 | 0.5824 | 0.2456 | 0.0836 |
| CK | 3.7574 | 2.9265 | 2.6256 | 3.7571 | 2.9272 | 2.6223 | 3.7563 | 2.9261 | 2.6256 |

Table 6. Statistics from simulated Weibull regression data ($q = 0.50, \beta_0 = 1.00, \beta_1 = 2.50$).

| Statistic | $k = 0.5$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | |
| True value | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 0.9963 | 1.0152 | 1.0015 | 0.9982 | 1.0076 | 1.0008 | 0.9991 | 1.0038 | 1.0004 |
| Bias | −0.0037 | 0.0152 | 0.0015 | −0.0018 | 0.0076 | 0.0008 | −0.0009 | 0.0038 | 0.0004 |
| Variance | 0.3423 | 0.0885 | 0.0261 | 0.0856 | 0.0221 | 0.0065 | 0.0214 | 0.0055 | 0.0016 |
| RMSE | 0.5851 | 0.2978 | 0.1615 | 0.2925 | 0.1489 | 0.0807 | 0.1463 | 0.0745 | 0.0404 |
| CS | −0.2084 | −0.1718 | −0.1039 | −0.2083 | −0.1716 | −0.1038 | −0.2084 | −0.1715 | −0.1044 |
| CK | 3.0076 | 3.1324 | 2.9603 | 3.0073 | 3.1323 | 2.9601 | 3.0069 | 3.1312 | 2.9588 |
| | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | |
| True value | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 |
| Mean | 2.5194 | 2.4831 | 2.4954 | 2.5097 | 2.4916 | 2.4977 | 2.5049 | 2.4958 | 2.4989 |
| Bias | 0.0194 | −0.0169 | −0.0046 | 0.0097 | −0.0084 | −0.0023 | 0.0049 | −0.0042 | −0.0011 |
| Variance | 0.8964 | 0.2355 | 0.0745 | 0.2241 | 0.0589 | 0.0186 | 0.0560 | 0.0147 | 0.0047 |
| RMSE | 0.9470 | 0.4856 | 0.2730 | 0.4735 | 0.2428 | 0.1365 | 0.2368 | 0.1214 | 0.0682 |
| CS | 0.0619 | −0.0343 | 0.1067 | 0.0618 | −0.0343 | 0.1066 | 0.0618 | −0.0342 | 0.1060 |
| CK | 2.8447 | 3.0612 | 3.0310 | 2.8448 | 3.0615 | 3.0307 | 2.8453 | 3.0600 | 3.0309 |
| | \hat{k} | | | \hat{k} | | | \hat{k} | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 1.0000 | 1.0000 | 1.0000 | 2.0000 | 2.0000 | 2.0000 |
| Mean | 0.5210 | 0.5061 | 0.5021 | 1.0419 | 1.0122 | 1.0043 | 2.0838 | 2.0243 | 2.0087 |
| Bias | 0.0210 | 0.0061 | 0.0021 | 0.0419 | 0.0122 | 0.0043 | 0.0838 | 0.0243 | 0.0087 |
| Variance | 0.0036 | 0.0008 | 0.0003 | 0.0144 | 0.0033 | 0.0010 | 0.0576 | 0.0130 | 0.0041 |
| RMSE | 0.0636 | 0.0292 | 0.0162 | 0.1271 | 0.0584 | 0.0324 | 0.2543 | 0.1168 | 0.0648 |
| CS | 0.5825 | 0.2447 | 0.0838 | 0.5824 | 0.2446 | 0.0838 | 0.5825 | 0.2446 | 0.0820 |
| CK | 3.7565 | 2.9255 | 2.6254 | 3.7563 | 2.9257 | 2.6253 | 3.7571 | 2.9255 | 2.6247 |

Table 7. Statistics from simulated Weibull regression data ($q = 0.90, \beta_0 = 1.00, \beta_1 = 2.50$).

| Statistic | $k = 0.5$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | | $\hat{\beta}_0$ | | |
| True value | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Mean | 0.9295 | 0.9938 | 0.9937 | 0.9648 | 0.9969 | 0.9969 | 0.9824 | 0.9985 | 0.9984 |
| Bias | −0.0705 | −0.0062 | −0.0063 | −0.0352 | −0.0031 | −0.0031 | −0.0176 | −0.0015 | −0.0016 |
| Variance | 0.3074 | 0.0794 | 0.0235 | 0.0769 | 0.0198 | 0.0059 | 0.0192 | 0.0050 | 0.0015 |
| RMSE | 0.5589 | 0.2818 | 0.1534 | 0.2795 | 0.1409 | 0.0767 | 0.1397 | 0.0705 | 0.0384 |
| CS | −0.1500 | −0.1104 | −0.1234 | −0.1505 | −0.1111 | −0.1234 | −0.1504 | −0.1107 | −0.1230 |
| CK | 2.9199 | 3.1514 | 2.9857 | 2.9190 | 3.1513 | 2.9857 | 2.9190 | 3.1501 | 2.9853 |
| | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | | $\hat{\beta}_1$ | | |
| True value | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 | 2.5000 |
| Mean | 2.5194 | 2.4832 | 2.4954 | 2.5097 | 2.4916 | 2.4977 | 2.5048 | 2.4958 | 2.4988 |
| Bias | 0.0194 | −0.0168 | −0.0046 | 0.0097 | −0.0084 | −0.0023 | 0.0048 | −0.0042 | −0.0012 |
| Variance | 0.8963 | 0.2355 | 0.0745 | 0.2241 | 0.0589 | 0.0186 | 0.0560 | 0.0147 | 0.0047 |
| RMSE | 0.9469 | 0.4856 | 0.2730 | 0.4735 | 0.2428 | 0.1365 | 0.2368 | 0.1214 | 0.0683 |
| CS | 0.0615 | −0.0349 | 0.1067 | 0.0620 | −0.0343 | 0.1068 | 0.0619 | −0.0341 | 0.1064 |
| CK | 2.8454 | 3.0617 | 3.0310 | 2.8448 | 3.0610 | 3.0311 | 2.8448 | 3.0609 | 3.0310 |
| | \hat{k} | | | \hat{k} | | | \hat{k} | | |
| True value | 0.5000 | 0.5000 | 0.5000 | 1.0000 | 1.0000 | 1.0000 | 2.0000 | 2.0000 | 2.0000 |
| Mean | 0.5210 | 0.5061 | 0.5021 | 1.0419 | 1.0122 | 1.0043 | 2.0838 | 2.0243 | 2.0086 |
| Bias | 0.0210 | 0.0061 | 0.0021 | 0.0419 | 0.0122 | 0.0043 | 0.0838 | 0.0243 | 0.0086 |
| Variance | 0.0036 | 0.0008 | 0.0003 | 0.0144 | 0.0033 | 0.0010 | 0.0576 | 0.0130 | 0.0041 |
| RMSE | 0.0636 | 0.0292 | 0.0162 | 0.1271 | 0.0584 | 0.0324 | 0.2543 | 0.1168 | 0.0648 |
| CS | 0.5825 | 0.2448 | 0.0840 | 0.5825 | 0.2449 | 0.0840 | 0.5825 | 0.2448 | 0.0838 |
| CK | 3.7566 | 2.9255 | 2.6255 | 3.7567 | 2.9257 | 2.6254 | 3.7566 | 2.9256 | 2.6257 |

Table 8. Summary statistics of the GCS residuals ($\beta_0 = 0.5; \beta_1 = 1.0$).

| Statistic | $k = 0.50$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|------------|-----------|-----------|------------|-----------|-----------|------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| $q = 0.10$ | | | | | | | | | |
| Mean | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| SD | 0.9882 | 0.9963 | 0.9986 | 0.9882 | 0.9963 | 0.9986 | 0.9882 | 0.9963 | 0.9986 |
| CS | 1.5711 | 1.8525 | 1.9394 | 1.5710 | 1.8524 | 1.9394 | 1.5711 | 1.8524 | 1.9394 |
| CK | 5.7186 | 7.6584 | 8.3894 | 5.7185 | 7.6578 | 8.3894 | 5.7187 | 7.6580 | 8.3895 |
| $q = 0.50$ | | | | | | | | | |
| Mean | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| SD | 0.9882 | 0.9963 | 0.9986 | 0.9882 | 0.9963 | 0.9986 | 0.9882 | 0.9963 | 0.9986 |
| CS | 1.5711 | 1.8524 | 1.9394 | 1.5711 | 1.8524 | 1.9394 | 1.5711 | 1.8524 | 1.9394 |
| CK | 5.7189 | 7.6577 | 8.3894 | 5.7188 | 7.6577 | 8.3895 | 5.7188 | 7.6577 | 8.3895 |
| $q = 0.90$ | | | | | | | | | |
| Mean | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| SD | 0.9882 | 0.9963 | 0.9986 | 0.9882 | 0.9963 | 0.9986 | 0.9882 | 0.9963 | 0.9986 |
| CS | 1.5711 | 1.8524 | 1.9394 | 1.5711 | 1.8524 | 1.9394 | 1.5711 | 1.8524 | 1.9394 |
| CK | 5.7189 | 7.6577 | 8.3895 | 5.7188 | 7.6577 | 8.3895 | 5.7189 | 7.6577 | 8.3895 |

Table 9. Summary statistics of the RQ residuals ($\beta_0 = 0.5$; $\beta_1 = 1.0$).

| Statistic | $k = 0.50$ | | | $k = 1.00$ | | | $k = 2.00$ | | |
|------------|------------|-----------|-----------|------------|-----------|-----------|------------|-----------|-----------|
| | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ | $n = 50$ | $n = 200$ | $n = 600$ |
| $q = 0.10$ | | | | | | | | | |
| Mean | 0.0012 | 0.0004 | 0.0001 | 0.0012 | 0.0004 | 0.0001 | 0.0012 | 0.0004 | 0.0001 |
| SD | 1.0134 | 1.0033 | 1.0011 | 1.0134 | 1.0033 | 1.0011 | 1.0133 | 1.0033 | 1.0011 |
| CS | 0.0142 | 0.0026 | 0.0009 | 0.0142 | 0.0027 | 0.0009 | 0.0142 | 0.0027 | 0.0009 |
| CK | 2.7487 | 2.9258 | 2.9774 | 2.7486 | 2.9258 | 2.9774 | 2.7487 | 2.9258 | 2.9774 |
| $q = 0.50$ | | | | | | | | | |
| Mean | 0.0012 | 0.0003 | 0.0001 | 0.0012 | 0.0004 | 0.0001 | 0.0012 | 0.0003 | 0.0001 |
| SD | 1.0134 | 1.0033 | 1.0011 | 1.0134 | 1.0033 | 1.0011 | 1.0134 | 1.0033 | 1.0011 |
| CS | 0.0142 | 0.0027 | 0.0009 | 0.0142 | 0.0027 | 0.0009 | 0.0142 | 0.0027 | 0.0009 |
| CK | 2.7487 | 2.9258 | 2.9774 | 2.7487 | 2.9258 | 2.9774 | 2.7487 | 2.9258 | 2.9774 |
| $q = 0.90$ | | | | | | | | | |
| Mean | 0.0012 | 0.0004 | 0.0001 | 0.0012 | 0.0003 | 0.0001 | 0.0012 | 0.0003 | 0.0001 |
| SD | 1.0134 | 1.0033 | 1.0011 | 1.0134 | 1.0033 | 1.0011 | 1.0134 | 1.0033 | 1.0011 |
| CS | 0.0142 | 0.0027 | 0.0009 | 0.0142 | 0.0027 | 0.0009 | 0.0142 | 0.0027 | 0.0009 |
| CK | 2.7487 | 2.9258 | 2.9774 | 2.7487 | 2.9258 | 2.9774 | 2.7487 | 2.9258 | 2.9774 |

5. Local Influence

5.1. Perturbation Matrix and Potentially Influential Cases

Local influence techniques examine the effect of small perturbations in the model data and/or assumptions regarding the estimated parameters. Let $\ell(\theta)$ be the log-likelihood function for the parameter θ of the model defined by (6), which is named the non-perturbed model. Consider a vector of \mathbb{R}^n , ω namely, called the vector of perturbation, and we define $\ell(\theta; \omega)$ as the log-likelihood function of the perturbed model and $\hat{\theta}_\omega$ as the maximum likelihood estimate of θ obtained from $\ell(\theta; \omega)$. Further, let $\omega_0 \in \mathbb{R}^n$ be a non-perturbation vector such that $\ell(\theta) = \ell(\theta; \omega_0)$. The likelihood displacement function (LD) defined as

$$LD(\omega) = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_\omega)) \quad (17)$$

is used to detect the impact of ω . We study the local behavior of the surface plot $(\omega^\top, LD(\omega))^\top$ around ω_0 . The direction in which the LD locally changes most rapidly is evaluated; that is, the maximum curvature of the surface. For $LD(\omega)$ given in (17), the maximum curvature is established as

$$C_{\max} = \max_{\|d\|=1} 2|d^\top B d|, \quad (18)$$

where B defined in (18) is given by $B = -\Delta^\top \check{\ell}_{\hat{\theta}\hat{\theta}} \Delta$ and d is a unit-length direction vector; see [22]. The expression $\check{\ell}_{\hat{\theta}\hat{\theta}}$ is the Hessian matrix of $\ell(\theta)$ evaluated at $\hat{\theta}$ and Δ is a $(p+1) \times n$ perturbation matrix also evaluated at $\theta = \hat{\theta}$ and $\omega = \omega_0$. Hence, the elements of Δ are stated as

$$\Delta_{ij} = \left. \frac{\partial^2 \ell(\theta; \omega)}{\partial \theta_i \partial \omega_j} \right|_{\theta=\hat{\theta}, \omega=\omega_0}, \quad i \in \{0, 1, \dots, p\}, j \in \{1, \dots, n\}. \quad (19)$$

Then, d_{\max} is a unit-length eigenvector associated with the maximum absolute eigenvalue of B . The plot of d_{\max} versus the index i may be considered to detect whether case i is potentially influential on $\hat{\theta}$. The direction $d = e_i$, where e_i is an $n \times 1$ vector of zeros, with one at the i th position, is another relevant direction to analyze. For such a direction, the normal curvature is $C_i(\theta) = 2|b_{ii}|$, where b_{ii} is the i th element of the diagonal of the matrix B . If

$$C_i(\theta) > 2 \sum_{i=1}^n \frac{C_i(\theta)}{n} = 2\bar{C}, \quad (20)$$

then the case i is potentially influential; see [35]. Next, we describe the matrix Δ for different perturbation schemes with its elements defined in generic terms in (19).

5.2. Perturbation Schemes

5.2.1. Case-Weight Perturbation

Consider $\omega = (\omega_1, \dots, \omega_n)^\top$ as a weight vector. Then, the perturbed log-likelihood function is defined by $\ell(\theta; \omega) = \sum_{i=1}^n \omega_i \ell_i(Q_i, k)$, where $0 \leq \omega_i \leq 1$, for $i \in \{1, \dots, n\}$. Therefore, the n columns of Δ are given by

$$\delta_i = \left(\begin{array}{c} x_i a_i z_i \\ b_i \end{array} \right) \Big|_{\theta=\hat{\theta}, \omega=(1, \dots, 1)^\top}, \quad i \in \{1, \dots, n\}.$$

5.2.2. Perturbation on the Response

Now, consider an additive perturbation on the response i by making $y_i(\omega_i) = y_i + \omega_i s_Y$, where $\omega_i \in \mathbb{R}$ and s_Y is a scale factor that can be the sample SD of Y , for $i \in \{1, \dots, n\}$. Then, the perturbed log-likelihood function corresponds to $\ell(\theta; \omega) = \sum_{i=1}^n \ell_{\omega_i}(Q_i, k)$, with

$$\ell_{\omega_i}(Q_i, k) = \log(-\log(1-q)) + \log(k) + (k-1) \log(y_i(\omega_i)) - k \log(Q_i) + (y_i(\omega_i))^k Q_i^{-k} \log(1-q),$$

for $i \in \{1, \dots, n\}$. The column vectors of Δ may be expressed as

$$\delta_i = \left(\begin{array}{c} x_i a_i \phi_i \rho_i \\ \tau_i \rho_i \end{array} \right) \Big|_{\theta=\hat{\theta}, \omega=(0, \dots, 0)^\top}, \quad i \in \{1, \dots, n\},$$

where

$$\begin{aligned} \phi_i &= -k^2 Q_i^{-k-1} y_i^{k-1} \log(1-q), \\ \tau_i &= \frac{1}{y_i} + \frac{y_i^{k-1}}{Q_i^k} \left(k \log\left(\frac{y_i}{Q_i}\right) + 1 \right) \log(1-q), \\ \rho_i &= s_Y. \end{aligned}$$

5.2.3. Perturbation in the Continuous Covariate

Consider an additive perturbation on a particular continuous covariate, x_t namely, with $t \in \{1, \dots, p-1\}$, by making $x_{ti}(\omega_i) = x_{ti} + \omega_i s_{X_t}$, where s_{X_t} is, again, a scale factor, which can be taken as the sample SD of X_t , and $\omega_i \in \mathbb{R}$, for $i \in \{1, \dots, n\}$. Therefore, the perturbed log-likelihood function is given by $\ell(\theta; \omega) = \sum_{i=1}^n \ell_{\omega_i}(Q_i, k)$, where

$$\ell_{\omega_i}(Q_i, k) = \log(-\log(1-q)) + \log(k) + (k-1) \log(y_i) - k \log(Q_i(\omega_i)) + y_i^k Q_i(\omega_i)^{-k} \log(1-q),$$

with $Q_i(\omega_i) = h^{-1}(x_i^\top(\omega_i) \beta)$ and $x_i^\top(\omega_i) = (1, x_{i1}, \dots, x_{it}(\omega_i), \dots, x_{i(p-1)})^\top$, for $i \in \{1, \dots, n\}$. Hence, the perturbation matrix takes the form given by

$$\Delta = \left(\begin{array}{c} \Delta_\beta \\ \Delta_k \end{array} \right) \Big|_{\theta=\hat{\theta}, \omega=(0, \dots, 0)^\top},$$

where $\Delta_\beta = (\Delta_{\beta_{ij}})$ is a $p \times n$ matrix defined as

$$\Delta_{\beta_{ij}} = \begin{cases} s_{X_t} \beta_t a'_i x_{ij} z_i + s_{X_t} \beta_t x_{ij} a_i^2 c_i, & j \neq t, i \in \{1, \dots, n\}, \\ s_{X_t} a_i z_i + s_{X_t} \beta_t a'_i x_{it} z_i + s_{X_t} \beta_t x_{it} a_i^2 c_i, & j = t, i \in \{1, \dots, n\}, \end{cases}$$

with a'_i being the derivative of a_i defined in (8). Here, $\Delta_k = (\zeta_1, \dots, \zeta_n)$, with $\zeta_i = s_{X_t} \beta_t a_i m_i$.

5.2.4. Perturbation of the Parameter k

In this case, the perturbation scheme consists of changing k by making $k_i = k/\omega_i$, with $\omega_i > 0$. Then, the perturbed log-likelihood is $\ell(\theta; \omega) = \sum_{i=1}^n \ell_{\omega_i}(Q_i, k_i)$, where

$$\begin{aligned} \ell_{\omega_i}(Q_i, k_i) &= \log(-\log(1-q)) + \log(k_i) + (k_i - 1) \log(y_i) - k_i \log(Q_i) \\ &\quad + y_i^{k_i} Q_i^{-k_i} \log(1-q), \quad i \in \{1, \dots, n\}. \end{aligned}$$

The column vectors of Δ can be expressed as

$$\delta_i = \begin{pmatrix} x_i a_i \zeta_i \\ \eta_i \end{pmatrix} \Big|_{\theta=\hat{\theta}, \omega=(1, \dots, 1)^\top}, \quad i \in \{1, \dots, n\},$$

where $\zeta_i = -k m_i$ and $\eta_i = -k d_i$.

6. Illustrative Example

6.1. The Adjusted Weibull Quantile Regression

To illustrate the use of the Weibull quantile regression formulated in this paper, we assume $Y_i \sim \text{Wei}(Q_i, k)$ and that our goal is to model the median. Consider two link functions (logarithm and square root) for a systematic component of the regression model, which are respectively stated as

$$(L1) \quad \log(Q_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (21)$$

$$(L2) \quad \sqrt{Q_i} = \mathbf{x}_i^\top \boldsymbol{\beta},$$

for $i \in \{1, \dots, 41\}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ is the vector of regression coefficients, and $\mathbf{x}_i^\top = (1, x_i)$ is the vector of values of $\mathbf{X}_i^\top = (1, X_i)$.

We implement the function `quant.weibull.reg()` in the R software, which allows us to fit Weibull quantile regression models to a data set, computing information criteria and residuals. To select the best model amongst a set of options, the AIC, BIC, and CAIC can be used. These information criteria assume the existence of an unknown “true model”. The AIC chooses the model whose divergence in relation to the “true model” is the minimum within the competing models and may be computed by

$$\text{AIC} = -2\ell(\hat{\theta}) + 2m, \quad (22)$$

where $\ell(\hat{\theta})$ is the log-likelihood function evaluated at $\theta = \hat{\theta}$ and m is the number of parameters of the proposed model, in our case $m = p + 1$. When the number of parameters is large, the AIC can have a deficient behavior. For this reason, a correction to the AIC is proposed as

$$\text{CAIC} = \text{AIC} + \frac{2m(m+1)}{n-m-1}, \quad (23)$$

where n is the sample size. The BIC is another information criterion for model selection based on maximizing the probability of choosing the true model and corresponds to

$$\text{BIC} = -2\ell(\hat{\theta}) + m \log(n). \quad (24)$$

In all these criteria, the best model, among a set of candidates, has the smallest value.

Another measure to be employed to choose among competing models is R_M^2 , which works similarly to the usual R^2 measure in mean regression and is defined as

$$R_M^2 = 1 - \exp\left(\frac{2}{n}(\ell(\tilde{\theta}) - \ell(\hat{\theta}))\right), \quad (25)$$

where $\ell(\tilde{\theta})$ and $\ell(\hat{\theta})$ are the maximized log-likelihood for the regression model without any covariate and with all covariates, respectively.

Values of the AIC, BIC, and CAIC defined in (22), (23), and (24), R_M^2 stated in (25) and the corresponding log-likelihood functions are reported in Table 10. We conclude that the model with the logarithm link function (L1) should be used to describe the median.

Table 10. Values of AIC, BIC, CAIC, and log-likelihood function for Weibull median-regression models with the data of time to electrical breakdown of an insulating fluid.

| Model | AIC | CAIC | BIC | R_M^2 | Log-Likelihood |
|-------|--------|--------|--------|---------|----------------|
| L1 | 327.07 | 327.71 | 332.21 | 0.71 | −160.53 |
| L2 | 351.63 | 352.28 | 356.77 | 0.47 | −172.81 |

Now, we compare model L1 with the proposed model in [10]. This comparison is not obvious since the construction of both models is different. Then, we compare both models in terms of R_M^2 defined in (25) and the pseudo- R^2 proposed for the Koenker–Bassett model [11] given by

$$R_{KB}^2 = 1 - \frac{V^1(q)}{V^0(q)}, \quad (26)$$

where $V^1(q)$ is the sum of weighted distances for the full q th quantile regression model and $V^0(q)$ is the sum of weighted distance for the model that includes only an intercept; that is, with no covariates. For our data, using (25) and (26), we obtain $R_M^2 = 0.71$ and $R_{KB}^2 = 0.03$, allowing us to conclude that our model is a better option for describing this data set.

Another relevant comparison is to consider a GLM-type model based on the Weibull distribution and reparameterized by its mean. We fit this model to our data taking the logarithmic link function. The value of the mean squared prediction error for the Weibull mean regression is 425,939.8, and for the median regression it is 231,845.7, meaning that, in terms of prediction error, our adjusted quantile model outperforms a GLM-type model based on the Weibull distribution.

Table 11 reports the maximum likelihood estimates for the model parameters, their approximated standard errors (SEs), and p -values based on the Wald test (described in Section 2). Thus, the predictive model is given by $\log(\hat{Q}) = 20.97 - 0.56x$, for $x > 0$.

Table 11. Estimate, SE, and p -value of the indicated parameter for the data of time to electrical breakdown of an insulating fluid.

| Statistic | $\hat{\beta}_0$ | $\hat{\beta}_1$ | \hat{k} |
|------------|-----------------|-----------------|-----------|
| Estimate | 20.97 | −0.56 | 0.82 |
| SE | 1.86 | 0.06 | 0.10 |
| p -value | <0.01 | <0.01 | <0.01 |

We evaluate the distributional assumption of the model by using the RQ and GCS residuals; that is, r_i^{RQ} and r_i^{GCS} , respectively. The QQ plots with envelopes for these residuals are presented in Figure 4, where all points are inside the bands. Therefore, r_i^{RQ} (a) and r_i^{GCS} (b) follow approximately standard normal and standard exponential distributions, respectively. This result allows us to validate that the response variable follows a Weibull distribution.

To compare the Weibull quantile regression model with other direct competing models, we adjust the Birnbaum–Saunders quantile regression model [2,16] with logarithm link function, which considers a response variable with an asymmetric distribution. For the Birnbaum–Saunders model with our data, the CAIC, BIC, and R_M^2 values are 334.96, 339.45 and 0.67, respectively. Note that the CAIC and BIC values are greater than the corresponding values for model L1; see Table 10. Furthermore, the value of R_M^2 is less than for model L1. The CAIC and BIC values are 531.31 and 535.80 for the normal model given in (1). Additionally, the residual r_i^{RQ} has been computed for the Birnbaum–Saunders model, and the QQ plot with envelopes is shown in Figure 4c. We observe that, compared

with the QQ plot in Figure 4a, the behavior is less homogeneous around a straight line, and there are points outside of the bands. Therefore, by considering the CAIC, BIC, and QQ plots of residuals, we conclude that the Weibull quantile regression outperforms the Birnbaum–Saunders quantile regression as well.

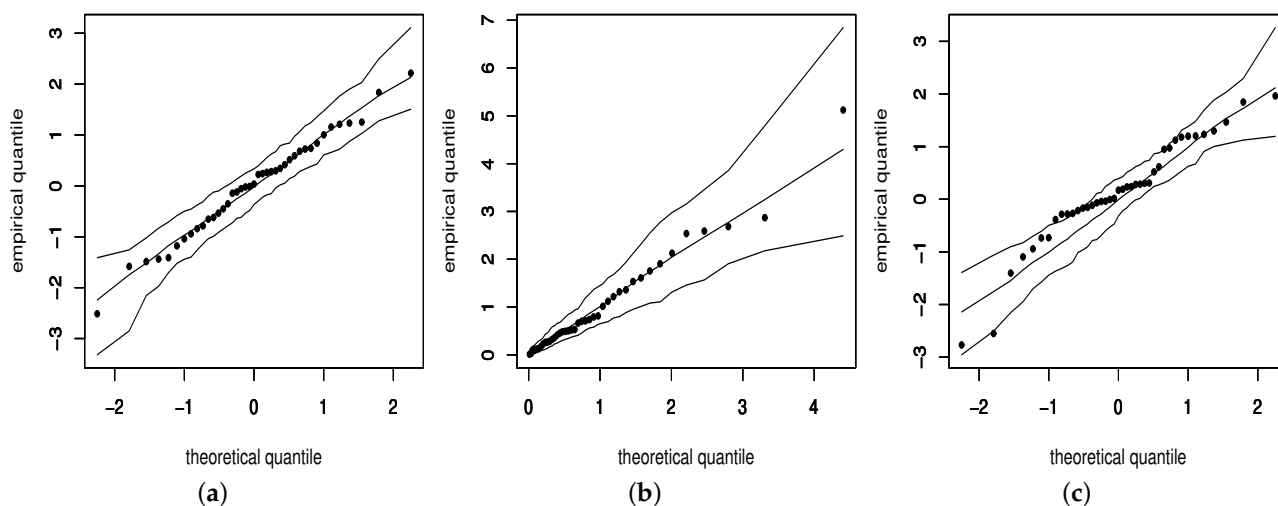


Figure 4. QQ plot with envelope of r_i^{RQ} (a) and r_i^{CGS} (b) for the Weibull median regression and of r_i^{RQ} for the Birnbaum–Saunders quantile regression model with logarithm link (c), using the data of the time to electrical breakdown of an insulating fluid.

6.2. Local Influence Analysis

Next, we analyze potentially influential cases by their local influence for the Weibull quantile regression with link L1, considering the four perturbation schemes as described in Section 5. In Figure 5, we show the index plots of $C_i(\theta)$ defined in (20) for each of them. Note that five cases are indicated as potentially influential, namely cases #1, #2, #3, #32, and #33. Observe that the local influence technique detects some atypical cases identified previously. From Figure 5d, note that small values for covariate X influence the estimates.

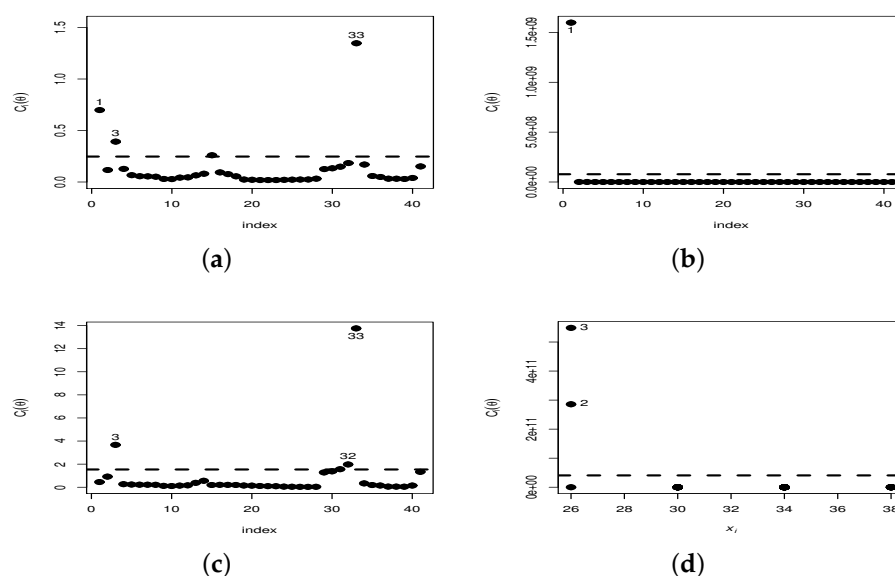


Figure 5. Index plots of $C_i(\theta)$ under case-weight perturbation (a), response perturbation (b), perturbation of the parameter k (c), and covariate perturbation X (d) for the data of time to electrical breakdown of an insulating fluid and the Weibull quantile regression.

We study the impact on the model inference considering the three cases most repeated in the index plots of Figure 5, which are cases #1, #3, and #33. The sets of cases $\{\#1\}$, $\{\#3\}$, $\{\#33\}$, $\{\#1, \#3\}$, $\{\#1, \#33\}$, $\{\#3, \#33\}$, and $\{\#1, \#3, \#33\}$ are removed and the model parameters are re-estimated. To determine the variation in the estimates of model parameters and in the associated SEs, we use the value of the relative changes (RCs) for each component of the parameter vector θ ; that is,

$$RC_{\theta_{j(i)}} = \left| \frac{\hat{\theta}_j - \hat{\theta}_{j(i)}}{\hat{\theta}_j} \right| \times 100\%, \quad RC_{SE(\hat{\theta}_j)(i)} = \left| \frac{\widehat{SE}(\hat{\theta}_j) - \widehat{SE}(\hat{\theta}_j)(i)}{\widehat{SE}(\hat{\theta}_j)} \right| \times 100\%,$$

where $\hat{\theta}_{j(i)}$ and $\widehat{SE}(\hat{\theta}_j)(i)$ denote the maximum likelihood estimates of θ_j and the estimated SE of the associated estimator, respectively, obtained after removing case i , for $j \in \{1, 2, 3\}$ and $i \in \{1, \dots, 41\}$, with $\theta_1 = \beta_0$, $\theta_2 = \beta_1$, and $\theta_3 = k$.

Table 12 reports the values of RCs for the data of time to electrical breakdown of an insulating fluid and the Weibull quantile regression. Note that the largest values of RCs are obtained when we remove cases #1 and #33, with the highest change occurring in the parameter k . The RCs of all parameters show a change close to 20%. However, we do not find any inferential change. Therefore, our study of local influence measures derived in this paper allows us to detect potentially influential cases, but these do not affect the model inference. Thus, the analysis of local influence presented with the data of the time to electrical breakdown of an insulating fluid permits us to conclude that the Weibull quantile regression model is nonsensitive to the atypical cases detected and exhibits an excellent performance to model this data set.

Table 12. RCs of maximum likelihood estimates and of the associated estimated SEs for the indicated cases, and respective p -values for the data of time to electrical breakdown of an insulating fluid and the Weibull quantile regression.

| | | Parameter | | |
|-----------------|--------------------|-----------|-----------|-------|
| Removed Case(s) | | β_0 | β_1 | k |
| None | $RC(\hat{\theta})$ | N/A | N/A | N/A |
| | $RC(\widehat{SE})$ | N/A | N/A | N/A |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#1} | $RC(\hat{\theta})$ | 3.41 | 3.41 | 5.81 |
| | $RC(\widehat{SE})$ | 2.52 | 2.78 | 5.38 |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#3} | $RC(\hat{\theta})$ | 4.87 | 5.23 | 0.77 |
| | $RC(\widehat{SE})$ | 18.86 | 18.31 | 0.14 |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#33} | $RC(\hat{\theta})$ | 1.46 | 1.98 | 8.16 |
| | $RC(\widehat{SE})$ | 13.23 | 13.26 | 12.43 |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#1, #3} | $RC(\hat{\theta})$ | 0.45 | 0.71 | 4.74 |
| | $RC(\widehat{SE})$ | 12.25 | 11.37 | 5.15 |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#1, #33} | $RC(\hat{\theta})$ | 4.46 | 4.92 | 15.72 |
| | $RC(\widehat{SE})$ | 15.21 | 15.50 | 20.27 |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#3, #33} | $RC(\hat{\theta})$ | 3.28 | 3.11 | 7.54 |
| | $RC(\widehat{SE})$ | 3.65 | 4.11 | 12.51 |
| | p -value | <0.01 | <0.01 | <0.01 |
| {#1, #3, #33} | $RC(\hat{\theta})$ | 0.56 | 0.76 | 14.77 |
| | $RC(\widehat{SE})$ | 5.40 | 6.19 | 20.12 |
| | p -value | <0.01 | <0.01 | <0.01 |

6.3. Coefficients across Quantiles

Quantile regression gives us a full description of how the covariates can affect the different values of the response variable. To show this, we consider the model given by $\log(Q) = \beta_0 + \beta_1 x$. If the covariate X increases from x_0 to $x_1 = x_0 + 1$, then the value of modeled quantile changes from $Q_0 = \exp(\beta_0 + \beta_1 x_0)$ to $Q_1 = \exp(\beta_0 + \beta_1(x_0 + 1))$, and then we have $(Q_1 - Q_0)/Q_0 = \exp(\beta_1) - 1$. Therefore, the coefficient β_1 is related to the percentage of change in the considered quantile when the covariate increases in one unit; see [36]. To illustrate this, we fit the Weibull regression model formulated in (21) considering the quantiles $q \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. In addition, we use a procedure to find the optimal value of q , q_{opt} namely, that is, the value of q that maximizes the log-likelihood function. We consider the profile log-likelihood method based on a grid of values of $q \in \{0.01, 0.02, \dots, 0.99\}$. Then, we estimate the Weibull regression parameters and compute the corresponding log-likelihood function. This procedure has been used in other contexts for Weibull models by [8] (pp. 426–433), where it is called a non-failing algorithm. The results are presented in Table 13. We observe that the covariate has the largest impact on higher levels of the response variable. For example, for values near to the 25th quantile of the response variable, if the voltage increases by 1 kV, the values of the response change by $(\exp(-0.62) - 1) \times 100\% = -47\%$. If we consider values close to the 90th quantile, it changes by $(\exp(-0.52) - 1) \times 100\% = -41\%$.

Table 13. Estimates of the parameters of the Weibull quantile regression model considering different quantiles, with insulating fluid data.

| Estimate | $q = 0.10$ | $q = 0.25$ | $q = 0.50$ | $q = 0.75$ | $q = 0.90$ | $q_{\text{opt}} = 0.32$ |
|-----------------|------------|------------|------------|------------|------------|-------------------------|
| $\hat{\beta}_0$ | 18.97 | 21.80 | 20.97 | 20.19 | 21.02 | 20.50 |
| $\hat{\beta}_1$ | −0.57 | −0.62 | −0.56 | −0.52 | −0.52 | −0.57 |
| \hat{k} | 0.84 | 0.81 | 0.82 | 0.84 | 0.84 | 0.85 |

7. Concluding Remarks

This paper has proposed novel quantile regression models for a response variable that follows an asymmetrical behavior based on a new parameterization of the Weibull distribution. We have estimated the new model parameters by using the maximum likelihood method and discussed hypothesis testing based on the Wald and likelihood ratio statistics. In addition, we have used the randomized quantile and generalized Cox–Snell residuals to evaluate the fit of the model. Monte Carlo simulation studies have found that (i) the maximum likelihood estimators are empirically consistent and asymptotically normal distributed (Tables 2–7), and (ii) the randomized quantile and generalized Cox–Snell residuals follow a standard normal distribution and exponential distribution with a parameter equal to one, respectively (Tables 8 and 9). Furthermore, we have derived local influence techniques to analyze the impact of a perturbation on the estimation of model parameters considering four schemes. We have applied the proposed model to a data set related to the time to electrical breakdown of an insulating fluid. The experimental results of this data analysis have shown the excellent performance of the proposed model to the data, making it a better choice than the usual normal regression model and other asymmetrical quantile regression models proposed in the literature. Some observations have been detected as potentially influential cases for our local diagnostic analysis (Figure 5) but without inferential change. Furthermore, we have studied the impact of the covariates on the quantiles of the response.

This work has evidenced that the new proposed model is helpful for independent data and a response variable with positive support. This new quantile regression model can also be suitable for small samples. However, we remark some limitations of our models and the proposed methodology. For example, diverse phenomena frequently provide other types of data to those analyzed in this study, such as censored, functional, spatial, and temporal data, as well as structures of measurement errors, and partial least squares, all of which

are suitable to be studied to increase the predictive capability in the modeling; see [37–41]. Then, it is necessary to formulate new models based on our approach to study these phenomena in such types of data and modeling structures. These structures are not an easy aspect to be explored, especially with spatially correlated data, because new multivariate distributions based on asymmetrical models need to be proposed and parameterized in terms of quantiles; see [38]. Furthermore, our proposal allows likelihood methods to be used, and thus this proposal can be applied to different distributions for modeling data, but adaptations of the corresponding methodology for each of these distributions must be performed.

An idea to enhance the empirical analysis of our proposal involves the following steps. First, consider the covariate with the highest simple correlation coefficient. Second, estimate the slope and intercept parameters in $h(Q)$. Third, taking a quantile as the median, create a data set on y and x stating a table with the observed values of y , the fitted values of y , and the residuals as their difference. Fourth, plot the observed and fitted values against the x values to allow the assessment of the model. In addition, least squares-fitted values can be displayed in the same graphical plot. A one-at-a-time cross-validation separates one observation for prediction from the remaining data, which adds a simple aspect about prediction that is also valuable. Other aspects related to k-fold cross validation are also appealing. Additionally, the relationship between a quantile and the covariates by means of a link function must be evaluated in each case, since it may not be correctly specified, implying extra analyses to achieve a better modeling. Moreover, measures such as the Cook distance and generalized leverage are essential diagnostic aspects of all statistical modeling, and they must be further studied for the newly proposed model. Weibull-type distributions with an extreme value index are widely used in many areas such as environmental sciences, hydrology, and meteorology; see [42]. Our proposed methodology can be adapted to this type of distributions. These and other aspects are part of our ongoing research.

Author Contributions: Data curation, L.S., H.S. and C.M.; investigation, L.S., V.L. and H.S.; formal analysis and methodology, L.S., V.L., H.S., C.M. and J.M.S.; writing—original draft, L.S., H.S. and C.M.; writing—review and editing, V.L. and J.M.S. All authors have read and agreed to the submitted version of the manuscript.

Funding: The research was partially funded by FONDECYT, project grant numbers 1200525 (V. Leiva and L. Sánchez) and 11190636 (C. Marchant) from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation; and by ANID-Millennium Science Initiative Program-NCN17_059 (C. Marchant).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The analyzed data and used codes are available under request.

Acknowledgments: The authors would also like to thank the Editor and three reviewers for their constructive comments which led to the improvement of the presentation of the manuscript.

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

References

1. Ventura, M.; Saulo, H.; Leiva, V.; Monsueto, S. Log-symmetric regression models: Information criteria, application to movie business and industry data with economic implications. *Appl. Stoch. Model. Bus. Ind.* **2019**, *35*, 963–977. [CrossRef]
2. Mazucheli, J.; Leiva, V.; Alves, B.; Menezes, A.F.B. A new quantile regression for modeling bounded data under a unit Birnbaum–Saunders distribution with applications in medicine and politics. *Symmetry* **2021**, *13*, 682. [CrossRef]
3. Johnson, N.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; Wiley: New York, NY, USA, 1994.
4. Castillo, E.; Hadi, A.S.; Balakrishnan, N.; Sarabia, J.M. *Extreme Value and Related Models with Applications in Engineering and Science*; Wiley: Hoboken, NJ, USA, 2005.
5. Saraiva, E.F.; Suzuki, A.K. Bayesian computational methods for estimation of two-parameters Weibull distribution in presence of right-censored data. *Chilean J. Stat.* **2017**, *8*, 25–43.
6. Weibull, W. A statistical distribution of wide applicability. *J. Appl. Mech.* **1951**, *18*, 293–297. [CrossRef]

7. Arnold, B.C.; Castillo, E.; Sarabia, J.M. Modeling the fatigue life of longitudinal elements. *Nav. Res. Logist. Q.* **1996**, *43*, 885–895. [\[CrossRef\]](#)
8. Rinne, H. *The Weibull Distribution*; Chapman and Hall: London, UK, 2009.
9. Laplace, P. *Theorie Analytique des Probabilités*; Editions Jacques Gabayr: Paris, France, 1818.
10. Koenker, R.; Bassett, G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [\[CrossRef\]](#)
11. Hao, L.; Naiman, D.Q. *Quantile Regression*. Sage Publications: Thousand Oaks, CA, USA, 2007.
12. Davino, C.; Furno, M.; Vistocco, D. *Quantile Regression: Theory and Applications*; Wiley: London, UK, 2013.
13. Koenker, R.; Chernozhukov, V.; He, X.; Peng, L. *Handbook of Quantile Regression*; CRC Press: Boca Raton, FL, USA, 2018.
14. Davison, A. *Statistical Models*; Cambridge University Press: Cambridge, UK, 2003.
15. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Chapman and Hall: London, UK, 1983.
16. Sánchez, L.; Leiva, V.; Galea, M.; Saulo, H. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. *Appl. Stoch. Model. Bus. Ind.* **2021**, *37*, 53–73. [\[CrossRef\]](#)
17. Saulo, H.; Dasilva, A.; Leiva, V.; Sánchez, L.; de la Fuente-Mella, H. Log-symmetric quantile regression models. *Stat. Neerl.* **2022**, in press. [\[CrossRef\]](#)
18. Cook, R.D.; Weisberg, S. *Residuals and Influence in Regression*; Chapman and Hall: London, UK, 1982.
19. Maddala, G.S. *Limited-Dependent and Qualitative Variables in Econometrics*; Cambridge University Press: Cambridge, UK, 1983.
20. Dunn, P.; Smyth, G. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.
21. Saulo, H.; Leão, J.; Leiva, V.; Aykroyd, R.G. Birnbaum-Saunders autoregressive conditional duration models applied to high-frequency financial data. *Stat. Pap.* **2019**, *60*, 1605–1629. [\[CrossRef\]](#)
22. Cook, R.D. Assessment of local influence. *J. R. Stat. Soc. B* **1986**, *48*, 133–169. [\[CrossRef\]](#)
23. Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V.; Barros, M. Reparameterized Birnbaum-Saunders regression models with varying precision. *Electron. J. Stat.* **2016**, *10*, 2825–2855. [\[CrossRef\]](#)
24. Garcia-Papani, F.; Leiva, V.; Uribe-Opazo, M.A.; Aykroyd, R.G. Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data. *Chemom. Intell. Lab. Syst.* **2018**, *177*, 114–128. [\[CrossRef\]](#)
25. Leiva, V.; Sanchez, L.; Galea, M.; Saulo, H. Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regression. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 1457–1471. [\[CrossRef\]](#)
26. Meeker, W.; Escobar, L. *Statistical Methods for Reliability Data*; Wiley: New York, NY, USA, 1998.
27. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
28. Therneau, T. A Package for Survival Analysis in R; R Package Version 3.2-10. 2021. Available online: <https://CRAN.R-project.org/package=survival> (accessed on 18 October 2021).
29. Maechler, M.; Rousseeuw, P.; Croux, C.; Todorov, V.; Ruckstuhl, A.; Salibian-Barrera, M.; Verbeke, T.; Koller, M.; Conceicao, E.L.; di Palma, M.A. Package ‘robustbase’. Basic Robust Statistics. 2021. Available online: <https://cran.r-project.org/web/packages/robustbase/robustbase.pdf> (accessed on 18 October 2021).
30. Noufaily, A.; Jones, M. Parametric quantile regression based on the generalized gamma distribution. *J. R. Stat. Soc. C* **2013**, *62*, 723–740. [\[CrossRef\]](#)
31. Mazucheli, J.; Menezes, A.; Fernandes, L.; Puziol, R.; Ghitany, M. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. *J. Appl. Stat.* **2019**, *47*, 954–974. [\[CrossRef\]](#)
32. Nocedal, J.; Wright, S. *Numerical Optimization*; Springer: New York, NY, USA, 2006.
33. Wald, A. *Sequential Analysis*; Wiley: New York, NY, USA, 1947.
34. Wilks, S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **1938**, *9*, 60–62. [\[CrossRef\]](#)
35. Lesaffre, E.; Verbeke, G. Local influence in linear mixed models. *Biometrics* **1998**, *54*, 570–582. [\[CrossRef\]](#)
36. Weisberg, S. *Applied Linear Regression*; Wiley: New York, NY, USA, 2014.
37. Huerta, M.; Leiva, V.; Liu, S.; Rodriguez, M.; Villegas, D. On a partial least squares regression model for asymmetric data with a chemical application in mining. *Chemom. Intell. Lab. Syst.* **2019**, *190*, 55–68. [\[CrossRef\]](#)
38. Sánchez, L.; Leiva, V.; Galea, M.; Saulo, H. Birnbaum-Saunders quantile regression models with application to spatial data. *Mathematics* **2020**, *8*, 1000. [\[CrossRef\]](#)
39. Calle-Saldarriaga, A.; Laniado, H.; Zuluaga, F.; Leiva, V. Homogeneity tests for functional data based on depth-depth plots with chemical applications. *Chemom. Intell. Lab. Syst.* **2021**, in press. [\[CrossRef\]](#)
40. Leiva, V.; Saulo, H.; Souza, R.; Aykroyd, R.G.; Vila, R. A new BISARMA time series model for forecasting mortality using weather and particulate matter data. *J. Forecast.* **2021**, *40*, 346–364. [\[CrossRef\]](#)
41. Figueroa-Zúñiga, J.I.; Bayes, C.L.; Leiva, V.; Liu, S. Robust Beta Regression Modeling with Errors-in-Variables: A Bayesian Approach and Numerical Applications. *Stat. Pap.* **2022**, in press. [\[CrossRef\]](#)
42. He, F.; Wang, H.J.; Tong, T. Extremal linear quantile regression with Weibull-type tails. *Stat. Sin.* **2020**, *30*, 1357–1377. [\[CrossRef\]](#)