

Article

Inverted Weibull Regression Models and Their Applications

Sarah R. Al-Dawsari * and Khalaf S. Sultan

Department of Statistics and Operations Research, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; ksultan@ksu.edu.sa

* Correspondence: s.aldawsari@psau.edu.sa

Abstract: In this paper, we propose the classical and Bayesian regression models for use in conjunction with the inverted Weibull (IW) distribution; there are the inverted Weibull Regression model (IW-Reg) and inverted Weibull Bayesian regression model (IW-BReg). In the proposed models, we suggest the logarithm and identity link functions, while in the Bayesian approach, we use a gamma prior and two loss functions, namely zero-one and modified general entropy (MGE) loss functions. To deal with the outliers in the proposed models, we apply Huber and Tukey's bisquare (biweight) functions. In addition, we use the iteratively reweighted least squares (IRLS) algorithm to estimate Bayesian regression coefficients. Further, we compare IW-Reg and IW-BReg using some performance criteria, such as Akaike's information criterion (AIC), deviance (D), and mean squared error (MSE). Finally, we apply the some real datasets collected from Saudi Arabia with the corresponding explanatory variables to the theoretical findings. The Bayesian approach shows better performance compare to the classical approach in terms of the considered performance criteria.

Keywords: Bayesian generalized linear models; Huber's function; identity link function; log link function; modified general entropy loss function; Tukey's bisquare function; zero-one loss function



Citation: Al-Dawsari, S.R.; Sultan, K.S. Inverted Weibull Regression Models and Their Applications. *Stats* **2021**, *4*, 269–290. <https://doi.org/10.3390/stats4020019>

Academic Editor: Carlos Alberto De Bragança Pereira

Received: 7 December 2020

Accepted: 29 March 2021

Published: 1 April 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

McCullagh and Nelder [1] published a book on the generalized linear models (GLMs) that led to their widespread use and appreciation. They extended the scoring method to maximum likelihood estimation (MLE) in exponential families. Nelder and Pregibon [2] described methods of jointly estimating the parameters of both the link and variance functions. The iteratively reweighted least squares (IRLS) algorithm is amenable to some statistics and measures that are common to all the GLMs. Nelder and Wedderburn [3] used the Newton-Raphson process for regression coefficients estimates. They reported that the Newton-Raphson process with expected second derivatives is equivalent to the Fisher's scoring technique. Additionally, De Jong and Heller [4] reported that the Newton-Raphson iteration equation leads to a sequence that often rapidly converges. These include the D statistic, along with some specific residuals and influence measures. Yuan and Bentler [5] reported that the convergence properties of the Fisher-scoring algorithm are affected by many factors. One of them is multicollinearity among the observed variables. If the sample or model implied covariance matrix is close to being singular, the Fisher-scoring algorithm may have difficulty reaching a set of converged solutions. Liao [6] introduced a systematic way of interpreting commonly used probability models: logit, probit, and the other GLMs.

The inverse Weibull (IW) model that was derived by Keller and Kamath [7] based on physical considerations on some mechanical components' failures subject to degradation phenomena, assuming that the strength of a component decreases with time with a power law. Calabria and Pulcini [8] proposed the IW distribution as a suitable model to describe mechanical degradation phenomena. They investigated a statistical property of the maximum likelihood estimator of the IW reliability. Jiang et al. [9] derived the Weibull (W) and IW mixture models with a common shape parameter for a system's components. They also used an example to illustrate that the proposed mixture model can be used to approximate

the reliability behaviors of the consecutive-k-out-of-n systems. Mahmoud et al. [10] considered the order statistics arising from IW distribution. They also derived an exact expression for the single moments of order statistics. Pasari and Dikshit [11] investigated the suitability of W distribution in the probabilistic assessment of earthquake hazards. The performance is also compared with two other popular models from the same W family, namely the two-parameter W model and the IW model. Jazi et al. [12] proposed a discrete IW distribution, a discrete version of the continuous IW variable of fitting discrete-time reliability and survival data sets. Kundu and Howlader [13] considered the Bayesian inference and prediction problems of the IW distribution based on Type-II censored data. Qingtian et al. [14] discussed the definition of environmental factors and restricting conditions of constant failure mechanism Based on generally accepted basic hypotheses. They estimated the IW distribution environment factor using the MLE and Bayes estimation methods. Musleh and Helu [15] considered two types of inference procedures: the classical (MLE, Approximate MLE and the least square method (LSE)) and the Bayesians (the squared error loss function (SQR), Linex loss function (LIN), General entropy loss function (GE), the Precautionary loss function (PRE)) to estimate the unknown parameters of the IW distribution when data under consideration are progressively type-II censoring. Akgul et al. [16] used IW distribution for modeling the seasonal wind speed using the modified maximum likelihood (MML) estimators of the parameters. The MML estimators' efficiencies are compared with the well-known maximum likelihood (ML) and the least-squares (LS) estimators via the Monte-Carlo simulation study. Nassar and Abo-Kasem [17] discussed the estimation problem of the unknown parameters of the IW distribution based on adaptive type-II progressively hybrid censored data. They used classical and Bayesian estimation methods to estimate the unknown parameters.

The Bayesian approach to modelling provides an alternative to the standard GLMs. The posterior mode estimation is an alternative to full posterior analysis or posterior mean estimation, which avoids numerical integrations or simulation methods. It has been proposed by many authors; see References [18,19]. Dey et al. [20] described how to conceptualize, perform, and critique the traditional GLMs from a Bayesian perspective and how to use modern computational methods to summarize inferences using simulation. Olsson [21] given an overview of the GLMs and has presented practical examples. The exponential family of distributions are discussed along with the maximum likelihood estimation and ways of assessing the fit of the model. Dobson and Barnett [22] presented a theoretical background of the GLMs. For the Bayesian estimation in this context, a useful asymmetric loss function, known as the LINEX loss function, was introduced by Varian (1975) and has been widely used by several authors. A suitable alternative to the modified LINEX loss is the general entropy (GE) loss function proposed by [23]. This loss function is a generalization of the entropy loss function used by several authors [23–25]. One highly used one is the zero-one loss function (for more details, see Reference [26]).

In order to reduce the influence of outliers on the estimate, some robust measures were proposed in the literature. The common robust estimation method can be divided into several categories: M, MM, median, L1, Msplit, R, S, least-trimmed squares, and sign-constraint robust least squares estimation. Among these, Huber's M estimation has become one of the main robust estimation methods by virtue of its simple calculation and convenience to implement [27]. The key aspect is the involvement of a loss function that is applied to data errors that was selected to less rapidly increase than the square loss function that is used in least-squares or maximum-likelihood procedures. There exist several well-known families of loss functions, such as Huber, Hampel, and Tukey's biweight (or bisquare) that can be used for the computation of M estimators [28]. The IW distribution is flexible distribution can be used as a competing to gamma and Weibull distributions to describe more widely real life data, failure characteristics, such as infant mortality, useful life and wear-out periods, applications in medicine and ecology, determining the cost-effectiveness, maintenance periods of reliability centered maintenance activities, and biological research (for more details, see References [7,9,13,29–35]).

This paper is structured as follows: In Section 2, we present an overview of the GLMs and propose the IW-Reg model under various link functions for estimating the model parameters. In Section 3, we estimate the IW-Bayesian regression (IW-BReg) model under a gamma prior, various link and two loss functions. In Section 4, we apply the theoretical results of both of IW-Reg and IW-BReg models to some real datasets collected from Saudi Arabia. Next, we investigate the performance of the proposed models in terms of some criteria, such as Akaike's information criterion (AIC), mean squared error (MSE), and deviance (D). In addition, we propose Huber and Tukey's bisquare (biweight) functions to improve IW-BReg models. In addition, we adopt the iteratively reweighted least squares (IRLS) algorithm to estimate the Bayesian regression coefficients. Finally, Section 5 draws a succinct conclusion to the findings.

2. Classical Approach

Nelder and Wedderburn [36] introduced the class of the GLMs, defined according to the assumption that y_1, y_2, \dots, y_n are observations of the response variable, with the density function of y_i as follows:

$$f(y_i; \theta_i) = e^{\theta_i y_i - \psi(\theta_i) + c(y_i)}, \quad i = 1, 2, \dots, n, \quad (1)$$

where $\psi(\cdot), c(\cdot)$ are known functions, with θ_i being the canonical parameter. A link function, $g(\cdot)$, relating to the regression coefficients, is given by

$$g(\mu_i) = \eta_i = x_i' \beta, \quad i = 1, 2, \dots, n, \quad (2)$$

where $g(\mu_i) = \theta_i$, $\beta = (\beta_1, \dots, \beta_p)'$ is a vector of p unknown regression parameters, $x_i' = (x_{i1}, x_{i2}, \dots, x_{ip})$ is a vector of explanatory variables, and η_i is a linear predictor of the vectors x_i' and β . Here, the $g(\cdot)$ is a link function, which is a monotonic differentiable invertible function. The model given by (1) and (2) is called the GLM. The GLM class includes, as special cases, linear regression and analysis of variance models, logit and probit models for quantal responses, log-linear models, and multinomial response models for counts; for more details, see Reference [36].

Consider that the probability density function of the IW distribution as follows [37]

$$f(y; \alpha, \gamma) = \alpha \gamma y^{-(\gamma+1)} e^{-\alpha y^{-\gamma}}; \quad y > 0, \quad \alpha, \gamma > 0; \quad (3)$$

here, $\alpha > 0$ and $\gamma > 0$ are the shape and scale parameters, respectively. The mean value of the response variable is given by

$$E(y) = \mu = \alpha^{\frac{1}{\gamma}} \Gamma\left(1 - \frac{1}{\gamma}\right). \quad (4)$$

The cumulative function of IW distribution is given by

$$F(y; \alpha) = e^{-\alpha y^{-\gamma}}; \quad y > 0.$$

Let y_i be a random sample from IW, and $\alpha_i = \left(\frac{\mu_i}{\Gamma(1 - \frac{1}{\gamma})}\right)^\gamma$, the log-likelihood function based on y_i , is given by

$$l_i = l(\mu_i | y_i) = \log \left[\frac{\mu_i}{\Gamma(1 - \frac{1}{\gamma})} \right]^\gamma + \log(\gamma) - (\gamma + 1) \log(y_i) - \left[\frac{\mu_i}{\Gamma(1 - \frac{1}{\gamma})} \right]^\gamma y_i^{-\gamma}. \quad (5)$$

The regression coefficients are estimated using the Fisher's scoring technique [3,36]. In order to develop the GLMs for our models, the IW-Reg are similar to GLMs, except that the

distribution of the response variable is not a member of the exponential family [38]. We also suggest some convenient link functions of $g(\cdot)$, in view of (2), as in the following lemmas.

Lemma 1 (The IW-Reg model with a log link function). *Let the response variable Y have an IW distribution, $i = 1, 2, \dots, n$, and let the link function of the form be*

$$g(\mu_i) = \log(\mu_i) = \eta_i = x_i' \beta, \quad i = 1, 2, \dots, n. \quad (6)$$

Thus, the estimated coefficients $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ using Fisher's scoring technique at the s^{th} iteration is given by

$$\hat{\beta}^{(s)} = \left(X' W \left(\hat{\beta}^{(s-1)} \right) X \right)^{-1} X' W \left(\hat{\beta}^{(s-1)} \right) Z, \quad s = 1, 2, 3, \dots, \quad (7)$$

where X is a covariates matrix, $\hat{\beta}_j^{(0)}$ is an initial vector, $W = \text{diag}(w_1, w_2, \dots, w_n)$,

$$w_i = [\gamma]^2, \quad (8)$$

and $Z' = (z_1, z_2, \dots, z_n)$, and

$$z_i = \sum_{j=1}^p x_{ij} \hat{\beta}_j^{(s-1)} + w_i^{-1} \gamma \left(1 - \frac{[\mu_i^{(s-1)}]^\gamma}{\left[\Gamma \left(1 - \frac{1}{\gamma} \right) y_i \right]^\gamma} \right). \quad (9)$$

The procedure in (7) can be repeated until $|\hat{\beta}^{(s)} - \hat{\beta}^{(s-1)}| \leq \epsilon$. The IW-Reg model in this case is given by

$$\hat{\mu}_i^{(s)} = e^{\left(\hat{\beta}_0^{(s)} + \hat{\beta}_1^{(s)} x_{i1} + \dots + \hat{\beta}_p^{(s)} x_{ip} \right)}. \quad (10)$$

Proof. Suppose that, in $y_i \sim f(y_i; \alpha, \gamma)$, as in (3), the parameter γ is assumed to be known. The log-likelihood function based on y_i , $i = 1, 2, \dots, n$ is given as in (5). The link function connecting the μ_i with the linear model $x_i' \beta$, in this case, is given as in (6). U_r for the log-likelihood is written from one observation as

$$U_r(\beta) = \frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^n \frac{\partial l_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r}, \quad r = 1, 2, \dots, p. \quad (11)$$

From (5) and (6), we have

$$U_r(\beta) = \sum_i^n \gamma \left(\frac{1}{\mu_i} - \frac{\mu_i^{\gamma-1}}{\left[\Gamma \left(1 - \frac{1}{\gamma} \right) y_i \right]^\gamma} \right) \mu_i x_{ir}, \quad (12)$$

which can be written in the matrix notation as

$$U(\hat{\beta}) = X^t Q(y, \mu(\hat{\beta})). \quad (13)$$

Taking the second derivatives of $l(\beta)$, we have

$$\frac{\partial U_r(\beta)}{\partial \beta_j} = \frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_r} = \sum_i^n \gamma \left(\frac{-1}{\mu_i^2} - \frac{(\gamma-1) \mu_i^{\gamma-2}}{\left[\Gamma \left(1 - \frac{1}{\gamma} \right) y_i \right]^\gamma} \right) \mu_i^2 x_{ij} x_{ir}, \quad j = 1, 2, \dots, p;$$

hence,

$$-E\left(\frac{\partial U_r(\beta)}{\partial \beta_j}\right) = \sum_{i=1}^n \gamma \left(\frac{1}{\mu_i^2} + \frac{(\gamma-1)\mu_i^{\gamma-2}E(y_i^{-\gamma})}{\left[\Gamma\left(1-\frac{1}{\gamma}\right)\right]^\gamma} \right) \mu_i^2 x_{ir} x_{ij}.$$

Since $E(y_i^{-\gamma}) = \left[\frac{\mu_i}{\Gamma\left(1-\frac{1}{\gamma}\right)} \right]^{-\gamma}$ and $\frac{\partial \mu_i}{\partial \beta_j} = \mu_i x_{ij}$, then $I_{rj} = -E\left(\frac{\partial U_r(\beta)}{\partial \beta_j}\right) = \sum_i^n \gamma^2 x_{ij} x_{ir}$, and

$$I(\hat{\beta}) = X'W(\hat{\beta})X, \quad (14)$$

where $W(\hat{\beta})$ is the diagonal matrix of weights, and w_i is as it is in (8). Then,

$$I(\hat{\beta}^{(s-1)})\hat{\beta}^{(s)} = I(\hat{\beta}^{(s-1)})\hat{\beta}^{(s-1)} + U(\hat{\beta}^{(s-1)}), \quad s = 1, 2, 3, \dots$$

From (13) and (14), we have

$$(X'W(\hat{\beta}^{(s-1)})X)\hat{\beta}^{(s)} = (X'W(\hat{\beta}^{(s-1)})X)\hat{\beta}^{(s-1)} + X'Q(y, \mu(\hat{\beta}^{(s-1)})). \quad (15)$$

Finally, the estimated coefficients $\hat{\beta}$ is given by

$$\hat{\beta}^{(s)} = (X'W(\hat{\beta}^{(s-1)})X)^{-1}X'W(\hat{\beta}^{(s-1)})[X\hat{\beta}^{(s-1)} + W^{-1}(\hat{\beta}^{(s-1)})Q(y, \mu(\hat{\beta}^{(s-1)}))],$$

and

$$\hat{\beta}^{(s)} = (X'W(\hat{\beta}^{(s-1)})X)^{-1}X'W(\hat{\beta}^{(s-1)})Z,$$

as given in (7).

To derive the MLS of β , the IRLS is used. Under certain regularity conditions on the likelihood function, the MLE $\hat{\beta}^{(s)}$ are asymptotically normal, unbiased, and efficient, with covariance matrix equal to the inverse of Fisher's information matrix [39]. Thus, $\hat{\beta}$ has asymptotically normal distribution,

$$\hat{\beta} \equiv N[\beta, (X'WX)^{-1}],$$

where $(X'WX)^{-1}$ is the inverse of Fisher's information matrix. \square

Lemma 2 (The IW-Reg model with identity link function). *Let the response variable Y have an IW distribution, $i = 1, 2, \dots, n$, and let the link function of the form be*

$$g(\mu_i) = \mu_i = \eta_i = x_i'\beta, \quad i = 1, 2, \dots, n. \quad (16)$$

Thus, the estimated coefficients $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ using Fisher's scoring technique at the s^{th} iteration is given by

$$\hat{\beta}^{(s)} = (X'W(\hat{\beta}^{(s-1)})X)^{-1}X'W(\hat{\beta}^{(s-1)})Z, \quad s = 1, 2, 3, \dots, \quad (17)$$

where X is a covariates matrix, $\hat{\beta}_j^{(0)}$ is an initial vector, $W = \text{diag}(w_1, w_2, \dots, w_n)$,

$$w_i = \left[\frac{\gamma}{\mu_i^{(s-1)}} \right]^2, \quad (18)$$

and $Z' = (z_1, z_2, \dots, z_n)$, and

$$z_i = \sum_{j=1}^p x_{ij} \hat{\beta}_j^{(s-1)} + w_i^{-1} \gamma \left(\frac{1}{\mu_i^{(s-1)}} - \frac{[\mu_i^{(s-1)}]^{-\gamma-1}}{\left[\Gamma\left(1 - \frac{1}{\gamma}\right) y_i\right]^\gamma} \right). \quad (19)$$

The procedure in (17) can be repeated until $|\hat{\beta}^{(s)} - \hat{\beta}^{(s-1)}| \leq \varepsilon$. The IW-Reg model in this case is given by

$$\hat{\mu}_i^{(s)} = \hat{\beta}_0^{(s)} + \hat{\beta}_1^{(s)} x_{i1} + \dots + \hat{\beta}_p^{(s)} x_{ip}. \quad (20)$$

Proof. Suppose that, in $y_i \sim f(y_i; \alpha, \gamma)$, as in (3), the parameter γ is assumed to be known. The log-likelihood function based on y_i , $i = 1, 2, \dots, n$ is given as in (5). The link function connecting the μ_i with the linear model $x_i' \beta$, in this case, is given in (16). From (5), (11), and (16), we have

$$U_r(\beta) = \sum_{i=1}^n \gamma \left(\frac{1}{\mu_i} - \frac{\mu_i^{\gamma-1}}{\left[\Gamma\left(1 - \frac{1}{\gamma}\right) y_i\right]^\gamma} \right) x_{ir},$$

which can be written in the matrix notation as in (13). Taking the second derivatives of $l(\beta)$, we have

$$\frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_r} = \sum_i \gamma \left(\frac{-1}{\mu_i^2} - \frac{(\gamma-1) \mu_i^{\gamma-2}}{\left[\Gamma\left(1 - \frac{1}{\gamma}\right) y_i\right]^\gamma} \right) x_{ir} x_{ij}$$

hence,

$$-E \left(\frac{\partial U_r(\beta)}{\partial \beta_j} \right) = -E \left(\frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_r} \right) = \sum_i \gamma \left(\frac{1}{\mu_i^2} + \frac{(\gamma-1) \mu_i^{\gamma-2} E(y_i^{-\gamma})}{\left[\Gamma\left(1 - \frac{1}{\gamma}\right)\right]^\gamma} \right) x_{ij} x_{ir},$$

since $E(y_i^{-\gamma}) = \left[\frac{\mu_i}{\Gamma\left(1 - \frac{1}{\gamma}\right)} \right]^{-\gamma}$ and $\frac{\partial \mu_i}{\partial \beta_j} = x_{ij}$, then $I_{rj} = -E \left(\frac{\partial U_r(\beta)}{\partial \beta_j} \right) = \sum_i^n x_{ij} \left[\frac{\gamma}{\mu_i^{(s-1)}} \right]^2 x_{ir}$ and $I(\hat{\beta})$ is given as in (14), where $W(\hat{\beta})$ is the diagonal matrix of weights, and w_i is as it is in (18). Using (15), we have $\hat{\beta}^{(s)}$ as in (17). \square

Lemma 3 (Convergence estimates in the Fisher's scoring process). *Let the response variable Y have an IW distribution, and let the link function of the form be $g(\mu_i) = \eta_i = x_i' \beta$ $i = 1, 2, \dots, n$, the estimated coefficients $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ using Fisher's scoring technique at the s^{th} iteration is given by*

$$\hat{\beta}^{(s)} = \left(X' W(\hat{\beta}^{(s-1)}) X \right)^{-1} X' W(\hat{\beta}^{(s-1)}) Z, \quad s = 1, 2, 3, \dots,$$

Then, $p(|\hat{\beta}^{(s)} - \hat{\beta}^{(s-1)}| \leq \varepsilon) = 1$, where X is a covariate matrix, W is the diagonal matrix of weights, and Z is a vector of the response variable.

Proof. Suppose that, in $y_i \sim f(y_i; \alpha, \gamma)$, as in (3), the parameter γ is assumed to be known. The log-likelihood function based on y_i , $i = 1, 2, \dots, n$ is given as in (5). Furthermore, suppose that the link function $g(\mu_i)$ connecting the μ_i with the linear model $x_i' \beta$ is given as

in (2). The Fisher's scoring process to obtain the MLEs estimates $\hat{\beta}$ is given by computing the iterations:

$$\hat{\beta}^{(s)} = \hat{\beta}^{(s-1)} + \left[I(\hat{\beta}^{(s-1)}) \right]^{-1} U(\hat{\beta}^{(s-1)}), \quad s = 1, 2, 3, \dots, \quad (21)$$

where $U(\hat{\beta}^{(s-1)})$ is the score vector for the log-likelihood (5), and $I(\hat{\beta}^{(s-1)})$ is the Fisher's information matrix. Taking the expectation of the Equation (21), we have

$$E[\hat{\beta}^{(s-1)}] = E[\hat{\beta}^{(s-1)}] = \xi, \quad (22)$$

since the estimates $\hat{\beta}^{(s-1)}$ are the MLEs, and $E\left[\left[I(\hat{\beta}^{(s-1)})\right]^{-1} U(\hat{\beta}^{(s-1)})\right] = 0$. Hence, we get $E[\hat{\beta}_j^{(s)}] = E[\hat{\beta}_j^{(s-1)}] = \xi_j$ for all $j = 1, 2, \dots, p$ where ξ is a vector of expectations. From (21), we have

$$\left| \hat{\beta}^{(s)} - \hat{\beta}^{(s-1)} \right| = \left| \left[I(\hat{\beta}^{(s-1)}) \right]^{-1} U(\hat{\beta}^{(s-1)}) \right|. \quad (23)$$

Using the Chebyshev inequality, for every $\omega > 0$, we find

$$p\left(\left| \hat{\beta}^{(s)} - \hat{\beta}^{(s-1)} \right| \leq \omega\right) \geq 1 - \frac{E\left(\left| \hat{\beta}^{(s)} - \hat{\beta}^{(s-1)} \right|\right)}{\omega}. \quad (24)$$

Now, by the Jensen inequality,

$$p\left(\left| \hat{\beta}^{(s)} - \hat{\beta}^{(s-1)} \right| \leq \omega\right) \geq 1 - \frac{\left| E\left(\hat{\beta}^{(s)} - \hat{\beta}^{(s-1)}\right) \right|}{\omega}. \quad (25)$$

Let $\omega = \varepsilon$ and, using (25), we then obtain

$$p\left(\left| \hat{\beta}^{(s)} - \hat{\beta}^{(s-1)} \right| \leq \varepsilon\right) = 1, \quad (26)$$

since $E(\hat{\beta}^{(s)} - \hat{\beta}^{(s-1)}) = 0$. On the other hand, by choosing $\omega = \frac{\varepsilon}{n}$ into the Equation (26), this becomes $p\left(\left| \hat{\beta}^{(s)} - \hat{\beta}^{(s-1)} \right| \leq 0\right) = 1$ as $n \rightarrow \infty$. \square

3. Bayesian Approach

Diaconis and Ylvisker [40] introduced a conjugate prior distribution for the exponential family, which, as in (1), can be shown as

$$\pi(\theta_i) = k_1 e^{m\mu_0\theta_i - m\psi(\theta_i)}, \quad i = 1, 2, \dots, n \quad (27)$$

where k_1 is a normalization constant, and m, μ_0 are natural parameters. The θ_i values are connected to the regression coefficients by the link function $\eta_i = x_i'\beta$ as

$$g^*(\eta_i) = \theta_i. \quad (28)$$

The posterior distribution of θ_i is given by

$$\pi(\theta_i | y_i) = k_2 e^{(y_i + m\mu_0)\theta_i - (1+m)\psi(\theta_i)}. \quad (29)$$

Das and Dey [41] suggested a Jacobian transformation and rewrote (29) with the term η_i , as

$$\pi(\eta_i|y_i) = K_2 e^{(y_i + m\mu_0)g^*(\eta_i) - (1+m)\psi(g^*(\eta_i))} \frac{\partial g^*(\eta_i)}{\partial \eta_i}, \quad (30)$$

where k_2 is a normalization constant, and $\frac{\partial g^*(\eta_i)}{\partial \eta_i} \neq 0$. They used a zero-one loss function to attain the posterior mode of (30) as $\hat{\eta}_i = h(y_i)$; hence, the estimated coefficients $\hat{\beta}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*, \dots, \hat{\beta}_p^*)'$ are given by

$$\hat{\beta}^* = (X'X)^{-1}X'\hat{\eta}, \quad (31)$$

where $\hat{\beta}^*$ is the least square estimates, and $\hat{\eta}' = (\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_n)$ [40,41]. Under regularity conditions, the estimator $\hat{\beta}^*$ has a asymptotically normal distribution $\hat{\beta} \equiv N[\beta, F^{-1}]$, where F^{-1} is the inverse of Bayesian Fisher's information (BIF). Note that the BFI is given by

$$F = \left[\frac{\partial}{\partial \eta} \log \pi(\eta|y) \right]^2,$$

where $\pi(\eta|y)$ is the posterior pdf of η [39,42,43].

In order to develop the Bayesian approach, we propose a modified loss function of the general entropy (MGE) loss function to be appropriate for the Bayesian estimates. The MGE loss function is introduced in the following lemma.

Lemma 4 (The MGE loss function). *Consider that the posterior distribution of the η_i is $\pi(\eta_i|y_i)$, $\hat{\eta}_i$ is an estimate of η_i , and y_i , $i = 1, 2, \dots, n$ are independent observations. A suitable alternative loss function to the GE loss is the MGE loss function, given as*

$$l(g^*(\hat{\eta}_i); g^*(\eta_i)) \propto \left(\frac{g^*(\hat{\eta}_i)}{g^*(\eta_i)} \right)^\kappa - \kappa \log \left(\frac{g^*(\hat{\eta}_i)}{g^*(\eta_i)} \right) - 1, \quad \kappa \neq 0. \quad (32)$$

Thus, the posterior Bayes estimates of η_i is given by solving the equation

$$E([g^*(\eta_i)]^{-\kappa}|y_i) = [g^*(\hat{\eta}_i)]^{-\kappa} \frac{\partial g^*(\hat{\eta}_i)}{\partial \hat{\eta}_i}. \quad (33)$$

Proof. Consider the MGE loss function is given as in (32). The posterior expectation of the loss function with respect to the posterior $\pi(\eta_i|y_i)$ is given as

$$\begin{aligned} E[l(g^*(\hat{\eta}_i); g^*(\eta_i)|y_i)] &= \int_{-\infty}^{\infty} \left[\left(\frac{g^*(\hat{\eta}_i)}{g^*(\eta_i)} \right)^\kappa - \kappa \log \left(\frac{g^*(\hat{\eta}_i)}{g^*(\eta_i)} \right) - 1 \right] \pi(\eta_i|y_i) d\eta_i, \\ &= [g^*(\hat{\eta}_i)]^\kappa \int_{-\infty}^{\infty} \left[\left(\frac{1}{g^*(\eta_i)} \right)^\kappa \right] \pi(\eta_i|y_i) d\eta_i - \kappa \int_{-\infty}^{\infty} [\log g^*(\hat{\eta}_i) - \log g^*(\eta_i)] \pi(\eta_i|y_i) d\eta_i - 1. \end{aligned}$$

The value of $\hat{\eta}_i$ that minimizes the posterior expectation of the MGE loss function is obtained by solving the equation

$$\frac{\partial E[l(g^*(\hat{\eta}_i); g^*(\eta_i)|y_i)]}{\partial \hat{\eta}_i} = \kappa [g^*(\hat{\eta}_i)]^{\kappa-1} E([g^*(\eta_i)]^{-\kappa}|y_i) - \kappa \frac{1}{g^*(\hat{\eta}_i)} \frac{\partial g^*(\hat{\eta}_i)}{\partial \hat{\eta}_i} = 0;$$

hence,

$$[g^*(\hat{\eta}_i)]^{\kappa-1} E([g^*(\eta_i)]^{-\kappa}|y_i) = \frac{1}{g^*(\hat{\eta}_i)} \frac{\partial g^*(\hat{\eta}_i)}{\partial \hat{\eta}_i},$$

and

$$E([g^*(\eta_i)]^{-\kappa} | y_i) = [g^*(\hat{\eta}_i)]^{-\kappa} \frac{\partial g^*(\hat{\eta}_i)}{\partial \hat{\eta}_i},$$

as given in (33). \square

In order to develop a Bayesian approach, we suggest inverted Weibull Bayesian generalized linear models (IW-BReg) that are similar to the approach in Section 3, except that the distribution of the response variable is not a member of the exponential family. We use the general form of the posterior in (30), and since $g(\cdot)$ is a monotonic differentiable function, then we attain the posterior Bayes estimates. Moreover, we use the a log and identity link functions with different loss functions. The IW-BReg estimates correspond to the link functions using different loss functions, as in the following lemmas.

Lemma 5 (The IW-BReg model based on zero-one loss function). *Let the response variable Y have an IW distribution and let the link function of the form be as in (2). Consider that α has a gamma prior $G(\nu, \lambda)$ with the following density function:*

$$\pi(\alpha) = \frac{\lambda^\nu}{\Gamma(\nu)} e^{-\lambda\alpha} \alpha^{\nu-1}, \quad \alpha > 0, \lambda, \nu > 0. \quad (34)$$

Thus, the posterior mode of η_i by using a zero-one loss function can be derived by solving the following equation:

$$\frac{\nu}{g^*(\eta_i)} - (y_i^\gamma + \lambda) + \frac{1}{\left[\frac{\partial g^*(\eta_i)}{\partial \eta_i}\right]^2} \frac{\partial^2 g^*(\eta_i)}{\partial \eta_i^2} = 0, \quad i = 1, 2, \dots, n, \quad (35)$$

where $g^*(\eta_i) = \mu_i$, η_i is defined as in (6) and (16). The estimated coefficients $\hat{\beta}^*$ are given as in (31). The IW-BReg model in this case is given by

$$\hat{\mu}_i^* = g^{-1}(x_i' \beta), \quad i = 1, 2, \dots, n. \quad (36)$$

Proof. Suppose that $y_i \sim f(y_i; \alpha, \gamma)$ is as it is in (3), the parameter γ is assumed to be known, and the density function of y_i is given by

$$f(y_i; \alpha, \gamma) = \alpha \gamma y_i^{-(\gamma+1)} e^{-\alpha y_i^{-\gamma}}. \quad (37)$$

Consider a gamma prior for α_i , which can be written as in (34). The posterior distribution of α_i is given by

$$\pi(\alpha_i | y_i) = \frac{\alpha_i^\nu \gamma \lambda^\nu}{\Gamma(\nu)} y_i^{-(\gamma+1)} e^{-\alpha_i (y_i^{-\gamma} + \lambda)}.$$

Using Jacobian transformation from α_i to η_i , we have

$$\pi(\eta_i | y_i) \propto [g^*(\eta_i)]^\nu e^{-g^*(\eta_i)(y_i^{-\gamma} + \lambda)} \frac{\partial g^*(\eta_i)}{\partial \eta_i}. \quad (38)$$

Taking the derivative of the log posterior, we have

$$\frac{\partial \log(\pi(\eta_i | y_i))}{\partial \eta_i} \propto \frac{\nu}{g^*(\eta_i)} \frac{\partial g^*(\eta_i)}{\partial \eta_i} - (y_i^{-\gamma} + \lambda) \frac{\partial g^*(\eta_i)}{\partial \eta_i} + \frac{1}{\frac{\partial g^*(\eta_i)}{\partial \eta_i}} \frac{\partial^2 g^*(\eta_i)}{\partial \eta_i^2} = 0;$$

hence, we get the equation as in (35), and the posterior mode of η_i is given by solving it. \square

Lemma 6 (The IW-BReg model based on MGE loss function). *Let the response variable Y have an IW distribution, and let the link function of the form be as given in (2). Consider that α_i has a gamma prior with a density function as given in (34). Thus, the posterior Bayes estimates of η_i , by using an MGE loss function, can be derived by solving the equation*

$$\frac{-\gamma\lambda^\nu y_i^{-(\gamma+1)}}{\Gamma(\nu)(y_i^{-\gamma} + \lambda)} \lim_{t \rightarrow \infty} e^{-g^*(t)(y_i^{-\gamma} + \lambda)} \Big|_0^t - [g^*(\eta_i)]^{-\nu} \frac{\partial g^*(\eta_i)}{\partial \eta_i} = 0. \quad (39)$$

where $g^*(\eta_i) = \mu_i$, η_i is defined as in (6) and (16). The estimated coefficients $\hat{\beta}^*$ are given as in (31). The IW-BReg model in this case is given by

$$\hat{\rho}_i^* = g^{-1}(x_i' \beta), \quad i = 1, 2, \dots, n. \quad (40)$$

Proof. Suppose that $y_i \sim f(y_i; \alpha, \gamma)$ is as it is in (3). The parameter γ is assumed to be known, and the density function of y_i is given as in (37). Consider the gamma prior for α_i , which can be written as given in (34). Using the posterior distribution of η_i that is given in (38), we have

$$E([g^*(\eta_i)]^{-\nu} | y_i) = \int_0^\infty [g^*(\eta_i)]^{-\nu} \pi(\eta_i | y_i) d\eta_i,$$

$$E([g^*(\eta_i)]^{-\nu} | y_i) = \frac{-\gamma\lambda^\nu y_i^{-(\gamma+1)}}{\Gamma(\nu)(y_i^{-\gamma} + \lambda)} \int_0^\infty - (y_i^{-\gamma} + \lambda) e^{-g^*(\eta_i)(y_i^{-\gamma} + \lambda)} \frac{\partial g^*(\eta_i)}{\partial \eta_i} d\eta_i;$$

hence,

$$E([g^*(\eta_i)]^{-\nu} | y_i) = \frac{-\gamma\lambda^\nu y_i^{-(\gamma+1)}}{\Gamma(\nu)(y_i^{-\gamma} + \lambda)} \lim_{t \rightarrow \infty} e^{-g^*(t)(y_i^{-\gamma} + \lambda)} \Big|_0^t.$$

Using Lemma (4), we have

$$\frac{-\gamma\lambda^\nu y_i^{-(\gamma+1)}}{\Gamma(\nu)(y_i^{-\gamma} + \lambda)} \lim_{t \rightarrow \infty} e^{-g^*(t)(y_i^{-\gamma} + \lambda)} \Big|_0^t = [g^*(\eta_i)]^{-\nu} \frac{\partial g^*(\eta_i)}{\partial \eta_i}.$$

Thus, the posterior Bayes estimates of η_i by using the MGE loss function can be derived by solving the Equation (39). \square

4. Data Analysis

In this section, we show the usefulness and performance of the IW-Reg and IW-BReg models by applying the theoretical findings in Sections 2 and 3 to some real datasets. For simplicity, we use the following notations for the proposed models used throughout the applications.

Model	Description
Model I	IW-Reg model based on identity link function
Model II	IW-Reg model based on log link function
Model III	IW-BReg model based on identity link and zero-one loss function
Model IV	IW-BReg model based on log link and zero-one loss
Model V	IW-BReg model based on identity link and MGE loss
Model VI	IW-BReg model based on log link and MGE loss

4.1. Application 1: (The Minimum Temperatures)

Dataset in this application was collected from the meteorology station at King Khalid International Airport, Saudi Arabia during (2014–2018). This data contains 54 observations (monthly data), in which the response variable Y be the minimum of dry bulb temperatures in Celsius. The explanatory variables are; X_1 ; mean of relative humidity, X_2 ; mean of vapor pressure (mm), X_3 ; mean of sky cover oktes, X_4 ; maximum of station-level pressure (mm).

In order to aid in distributional assessment of the response variable Y , the empirical cumulative distribution function (ECDF) plot was proposed. Kolmogorov–Smirnov goodness of fit test (K–S) was calculated based on IW, Gaussian, and gamma distributions. The IW-Reg and IW-BReg models based on log and identity link, and loss functions were fitted using the proved Lemmas in Sections 2 and 3. Bayes coefficients were obtained using a gamma prior $G(\nu, \lambda)$ with some known values of the hyperparameters ν and λ . In addition, Huber's function was suggested to avoid such distortions due to an outlier in X_2 ; see Appendix A. In this case, under regularity conditions, estimator $\hat{\beta}^*$ has asymptotically normal distribution $\hat{\beta}^* \equiv N[\beta^*, (X'WX)^{-1}]$ [39,42]. The performance of all these models were compared. Modeling performance is measured in terms of some criteria, such as AIC, D, D/df, and MSE [4]. We also used Thiel's inequality coefficient $TIC = \frac{\sqrt{\sum (y_i - \hat{y}_i)^2}}{\sqrt{\sum y_i^2} + \sqrt{\sum \hat{y}_i^2}}$ to compare the prediction accuracy of the selected models [44,45]. The backward-selection method was used in the IW-BReg model to select the best fit in view of the covariates.

To check the adequacy for the selected models, we consider Pearson residuals [36,40]. R software was used to carry out calculations. In order to compare with known distributions, the `glm()` function in “stats” was used to fit the GLMs [46]. Functions `qqPlot()`, `ecdf()`, `boxplot`, and `ks.test()` in R package “stats” were used for the assessment distributions [47]. To solves n roots of n nonlinear equations in Section 3, the function `multiroot()` in R package “rootSolve” was used [48]. The fitting results and the relative errors (RE) of the selected model, and other numerical results are shown in Tables 1–3.

Based on the results obtained from K–S test, the p -value = 0.315 for the test indicates that the IW distribution fits the response variable in the given data quite well. Figure 1 provides the ECDF plot, and it is clear that the IW distribution fits these data well.

To compare between the Bayesian fitting results, we observe that the results based on MGE loss function are better than zero-one loss function. Table 1 shows that the IW-BReg models based on MGE loss function (Model V and VI) are good in terms of MSE, AIC, and D statistics. Table 1 also shows that the D/df of IW-Reg and IW-BReg models (I, II, III, IV, V, and VI) are less than 1, indicating that the fitting degree is very good. If the model is correct, the Pearson residuals $r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$ and Pearson statistics $Q = \sum_{i=1}^{54} r_i^2$ have an approximately normal distribution with mean 0 and chi-square distribution $\chi_{n-k'}^2$ respectively. For the IW-BReg model based on identity link and MGE loss function (Model V), the Pearson statistics is $Q = 4.0396$, the p -value for Anderson-Darling is 0.0001, and the Cox Stuart test is 1, so the Pearson residuals are not normal but randomly scattered around zero at the level of significant $\alpha = 0.05$. For the IW-BReg model based on log link and MGE loss function (Model VI), the Pearson statistics is $Q = 0.4460$, the p -value for Anderson-Darling is 0.06379, and the Cox Stuart test is 1, so the Pearson residuals are normal and randomly scattered around zero.

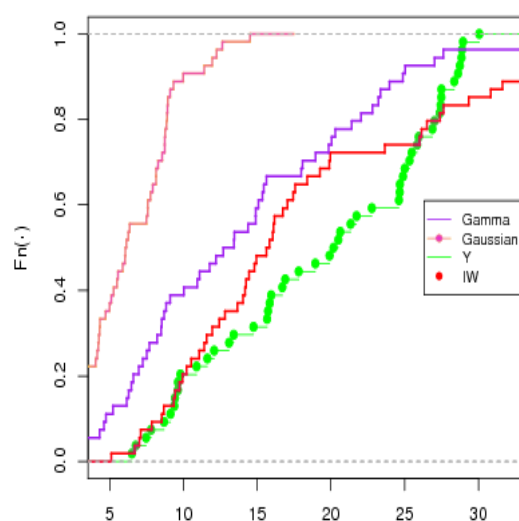


Figure 1. The empirical cumulative distribution function (ECDF) plot of the minimum temperatures (Y) based on IW distribution and some other distributions.

Table 1. Efficiency Gamma, Gaussian, inverted Weibull Regression (IW-Reg), IW-Bayesian regression (IW-BReg) models.

Model	AIC	D	D/df	MSE	Person Statistics	K-S Test p -Value
Model I	344.997	43.950	0.897	191.828	2.122	
Model II	579.782	0.009	0.000	7.026		
Model III	477.534	15.311	0.312	9.137		
Model IV	484.655	24.183	0.494	20.282		
Model V	294.264	0.459	0.009	1.774	4.040	
Model VI	364.333	0.350	0.007	2.349	0.446	0.3149
Gaussian (log)	186.800	80.491	0.936	1.643		0.0000
Gamma (log)	215.040	0.437	0.005	2.412		0.0321

Table 2. Akaike's information criterion (AIC), deviance (D), and mean squared error (MSE) of the Model VI (backward selection method).

Function	Variables	$\hat{\beta}$	Standard Error (SE)	z	p-Value	AIC	D	MSE
Step 1								
No weight	Intercept	26.6539	6.5930	4.0427	0.0001	364.3330	0.3500	2.3490
	X ₁	−0.0268	0.0036	−7.4222	0.0000			
	X ₂	0.1471	0.0180	8.1613	0.0000			
	X ₃	−0.0178	0.0282	−0.6296	0.5290			
	X ₄	−0.0255	0.0070	−3.6408	0.0003			
Step 2								
No weight	Intercept	26.5295	6.6037	4.0174	0.0001	364.3539	0.3709	2.3463 *
	X ₁	−0.0271	0.0036	−7.6350	0.0000			
	X ₂	0.1422	0.0160	8.8633	0.0000			
	X ₄	−0.0254	0.0070	−3.6132	0.0003			
Huber	Intercept	26.7509	6.7973	3.9355	0.0001	363.2006	0.8045	2.3451 **
	X ₁	−0.0269	0.0037	−7.3672	0.0000			
	X ₂	0.1420	0.0165	8.6036	0.0000			
	X ₄	−0.0256	0.0072	−3.5423	0.0004			

* The Model VI that was shown in Equation (41); ** The Model VI based on Huber's function that was shown in Equation (42).

Based on this analysis, we conclude that the Model VI is more appropriate for fitting these data, leading to the following equation

$$\hat{Y} = e^{26.6539 - 0.0268x_{i1} + 0.1471x_{i2} - 0.0178x_{i3} - 0.0255x_{i4}}, i = 1, 2, \dots, 54.$$

For the backward selection method results, Table 2, we can conclude that the predictive model is given as follows:

$$\hat{Y} = e^{26.5295 - 0.0271x_{i1} + 0.1422x_{i2} - 0.0254x_{i4}}, i = 1, 2, \dots, 54. \quad (41)$$

We also can see that, this model has AIC = 364.3539 and a low MSE = 2.3463, and there was also a significant relationship among variables when using level of significance $\alpha = 0.05$. For the residuals, the Pearson statistics is $Q = 0.4520$, p -value for Anderson-Darling is 0.0443 and for the Cox Stuart test is 1.

Because of the presence of an outlier, we can conclude that the Model VI based on Huber's function is the best for our data, and it is given as follows:

$$\hat{Y} = e^{26.7509 - 0.0269x_{i1} + 0.1420x_{i2} - 0.0256x_{i4}}, i = 1, 2, \dots, 54. \quad (42)$$

From Table 2, we can see, this model has AIC = 363.2006 and a low MSE = 2.3451, and there was also a significant relationship among variables when using the level of significance $\alpha = 0.05$. For the residuals, the Pearson statistics is $Q = 0.4532$, the p -value for Anderson-Darling is 0.052, and the Cox Stuart test is 1. Hence, the Pearson residuals are normal randomly scattered around zero; see Figure 2. The fitting results for this model during the year 2014 are shown in Table 3. We can also see that the fitting accuracy is good because the TIC value is closer to 0 than 1.

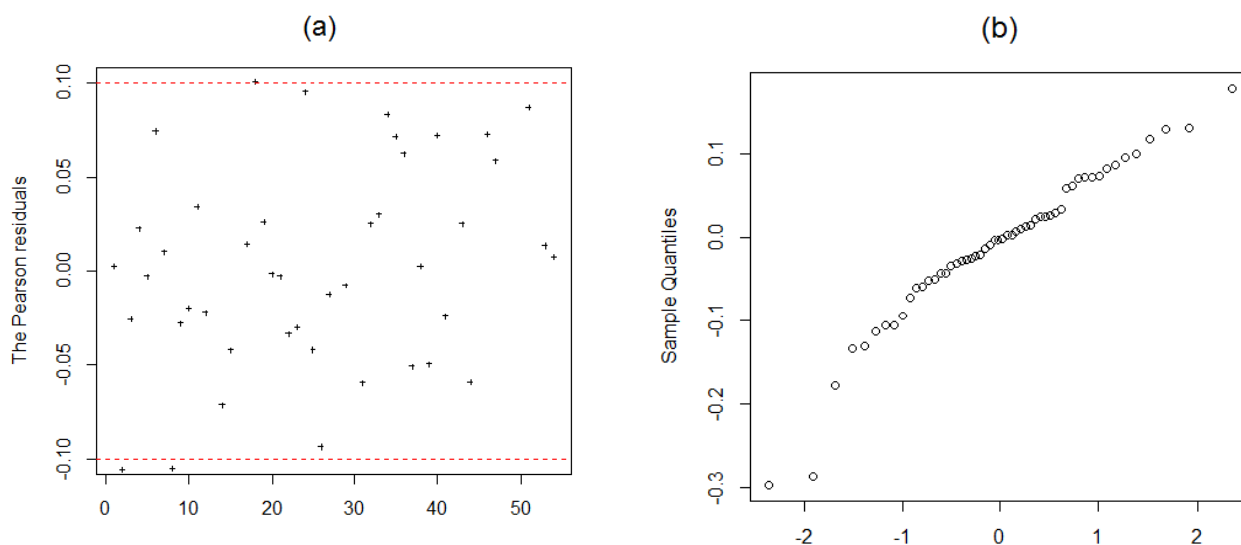


Figure 2. (a) Pearson residuals plot and (b) Normal Q-Q plot of the residuals using Model VI.

Table 3. Fitting results \hat{y}_i of the Model VI based on Huber's function (during fitting interval, 2014).

Months	y_i	\hat{y}_i	$ y_i - \hat{y}_i $	RE	Months	y_i	\hat{y}_i	$ y_i - \hat{y}_i $	RE
1	7.813	7.790	0.023	0.295	7	27.484	27.202	0.282	1.027
2	9.626	10.749	1.123	11.668	8	27.435	30.610	3.174	11.571
3	15.671	16.080	0.409	2.607	9	24.677	25.372	0.694	2.813
4	20.177	19.730	0.448	2.218	10	21.335	21.765	0.430	2.014
5	24.665	24.721	0.056	0.229	11	13.410	12.967	0.443	3.301
6	25.842	24.066	1.776	6.872	12	9.426	9.639	0.213	2.262
TIC		0.011							

4.2. Application 2: (Wind Speed Data)

The dataset in this application was taken again from the meteorology station at King Khalid International Airport, Saudi Arabia, in 2016. This data contains 91 observations, during 7 June and 5 September, (summer season), in which the response variable Y be the mean wind speed (km/h). The explanatory variables are; X_1 ; maximum's wind direction, X_2 ; maximum of station-level pressure (mm), X_3 ; mean of sea-level pressure (mm), X_4 ; mean of dry bulb temperatures of air (Celsius), X_5 ; mean of wet bulb temperatures (Celsius), X_6 ; mean of relative humidity, X_7 ; mean of vapor pressure (mm), X_8 ; mean of sky cover oktes, X_9 ; maximum of station-level pressure (mm), X_{10} ; maximum of sea-level pressure (mm), X_{11} ; maximum of dry bulb temperatures (Celsius), X_{12} ; maximum of the wet bulb temperatures (Celsius), X_{13} ; maximum of relative humidity, X_{14} ; minimum of station-level pressure (mm), X_{15} ; minimum of sea-level pressure (mm), X_{16} ; minimum of dry bulb temperatures, X_{17} ; minimum of the wet bulb temperatures, X_{18} ; minimum of relative humidity, X_{19} ; time of maximum daily wind (HH:MM).

Proceeding similarly, as in Application 1 to aid in the distributional assessment. In this dataset, we identify the outliers, different plots as the quantile-quantile (Q-Q) plot, ECDF, and box plot were proposed. Again, Lemmas in Sections 2 and 3 were applied to these data to fit the IW-Reg based on log and identity link functions were used. Besides being an alternative analysis, the IW-BReg models were obtained using a log, identity link, and a gamma prior with known hyperparameters ν and λ parameters. We also compare the performance of all these models. In addition, biweight function was suggested to avoid such distortions due to outliers; see Appendix A. In this case, under regularity conditions, estimator $\hat{\beta}^*$ has asymptotically normal distribution $\hat{\beta}^* \equiv N[\beta^*, (X'WX)^{-1}]$ [39,42]. The modeling performance was measured in terms of some criteria, such as AIC, D, D/df, and MSE [4]. We also used Theil's Inequality coefficient (TIC) to measure the prediction accuracy of the selected models [44,45]. To compare the residual for all models, we consider Pearson residuals to check the adequacy of the regression model fitted to the data [36,40].

Furthermore, to detect the influential cases, we use the Cook's distance measure using the formula $C_{(i)} = \frac{(\hat{\beta} - \hat{\beta}_{(i)})(X'X)(\hat{\beta} - \hat{\beta}_{(i)})}{k\hat{\sigma}^2}$ and $\hat{\sigma}^2 = \frac{\hat{\eta}'(I - X(X'X)^{-1}X')\hat{\eta}}{n-k}$ in the case of Bayesian analysis [40,49]. The backward selection method was used in the IW-Reg model to remove the input variable; see Table 4. R software was used to carry out the calculations. In order to compare with known distributions, the function glm in "stats" is used to fit the GLMs. The functions qqPlot, ecdf, boxplot, and ks.test in the R package "stats" are used for the assessment distributions [47]. To solves n roots of n nonlinear equations in Section 3, the function multiroot() in R package "rootSolve" was used [48]. The fitting, predictive results of these models and the other numerical results are shown on the Tables 4–8.

Based on the results obtained from K–S test, the p -value = 0.139 for the test indicates that the IW distribution fits the response variable in the given data quite well. Figure 3 provides the Q-Q plot and ECDF, and it is clear that the IW distribution fits these data well. Figure 4 provides box plot corresponding to the mean wind speed variable Y , and this chart mapped one outlier (leverage point) that exceeds the values of Q_3 .

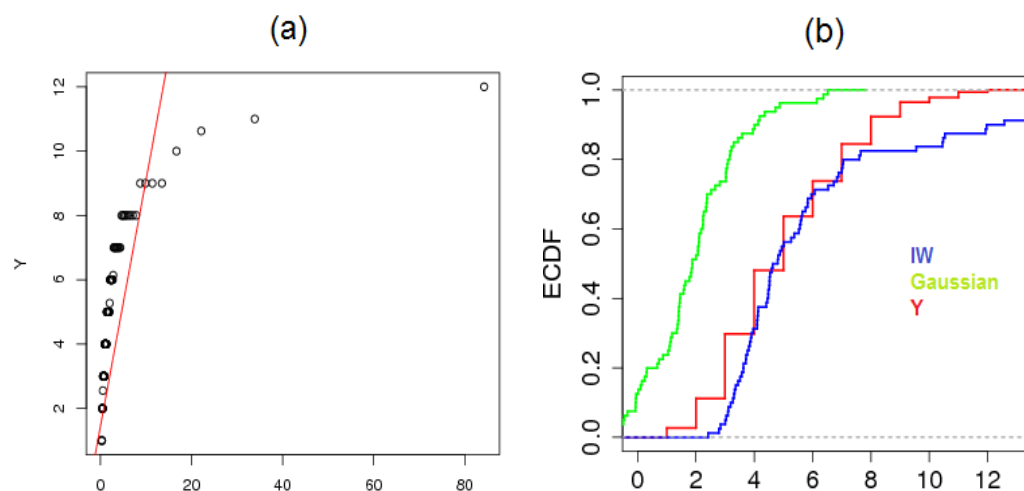


Figure 3. (a) Q-Q plots of the wind speed (Y) based on IW distribution and (b) The ECDF plot of Y based on IW and some other distributions.

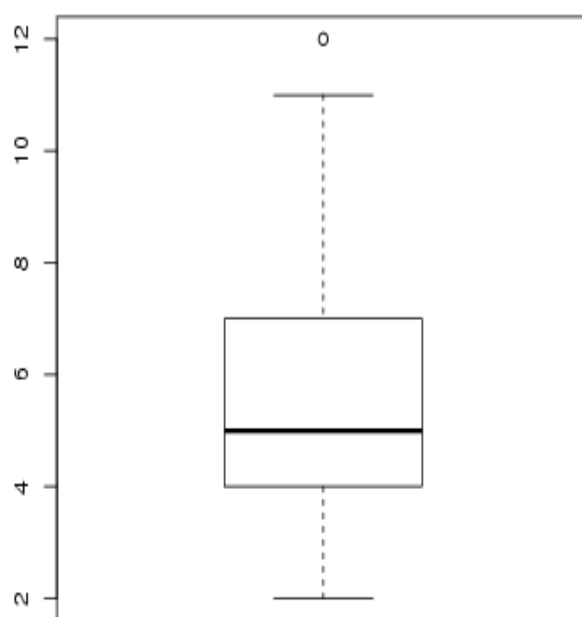


Figure 4. Box plot of the wind speed variable (Y).

From Table 4, we can observe that the variables X_2 , X_5 , X_8 , and X_{16} are significant for the model, so there is a significant relationship among variables. In these models, $\hat{\beta}^{(s)}$ is stabilized when the Fisher's scoring procedure is converged at $s = 6$ and $s = 15$, respectively, because of $|\hat{\beta}^{(s)} - \hat{\beta}^{(s-1)}| < \varepsilon$. To compare the Bayesian fitting results we observe that the results based on MGE loss function (Model V and VI) better than zero-one loss function (Model III and IV); see Table 5. Table 5 also shows that the D/df of the models I, II, III, IV, V, and VI are less than 1, indicating that the fitting degree is very good.

Table 4. The Model I (backward selection method).

Step	Variables	$\hat{\beta}$	SE	z-Statistics	p-Value	AIC	D	MSE
Step 1	Intercept	385.557	129.470	2.978	0.0029	352.790	54.4782	3.549
	X ₁	−0.013	0.014	−0.945	0.3446			
	X ₂	−0.883	0.851	−1.038	0.2993			
	X ₃	2.199	1.159	1.897	0.0579			
	X ₄	1.798	0.981	1.833	0.0668			
	X ₅	−4.039	2.254	−1.792	0.0732			
	X ₆	0.240	0.521	0.460	0.6456			
	X ₇	0.624	0.876	0.713	0.4760			
	X ₈	−0.537	0.188	−2.859	0.0042			
	X ₉	−0.741	0.882	−0.840	0.4010			
	X ₁₀	0.012	0.679	0.018	0.9855			
	X ₁₁	−0.555	0.225	−2.468	0.0136			
	X ₁₂	0.543	0.337	1.612	0.1070			
	X ₁₃	0.038	0.077	0.493	0.6218			
	X ₁₄	0.030	2.656	0.118	0.9065			
	X ₁₅	−1.110	0.690	−1.609	0.1075			
	X ₁₆	0.720	0.216	3.330	0.0009			
	X ₁₇	−0.124	0.386	−0.322	0.7475			
	X ₁₈	−0.055	0.271	−0.205	0.8376			
	X ₁₉	−0.048	0.062	−0.764	0.4447			
Step 2	Intercept	385.734	128.470	3.003	0.0027	352.790	54.4785	3.735
	X ₁	−0.014	0.014	−0.951	0.3414			
	X ₂	−0.884	0.847	−1.044	0.2964			
	X ₃	2.208	1.105	1.999	0.0456			
	X ₄	1.793	0.962	1.865	0.0622			
	X ₅	−4.034	2.248	−1.795	0.0727			
	X ₆	0.239	0.521	0.458	0.6466			
	X ₇	0.622	0.874	0.712	0.4765			
	X ₈	−0.537	0.188	−2.858	0.0043			
	X ₉	−0.731	0.590	−1.238	0.2155			
	X ₁₁	−0.555	0.225	−2.467	0.0136			
	X ₁₂	0.543	0.337	1.614	0.1065			
	X ₁₃	0.038	0.077	0.495	0.6205			
	X ₁₄	0.030	0.251	0.120	0.9048			
	X ₁₅	−1.116	0.647	−1.726	0.0844			
	X ₁₆	0.721	0.214	3.369	0.0008			
	X ₁₇	−0.126	0.368	−0.342	0.7324			
	X ₁₈	−0.055	0.270	−0.204	0.8385			
	X ₁₉	−0.048	0.062	−0.770	0.4415			
Step 16	Intercept	355.1198	75.107	4.728	0.0000	369.9813	71.6697	3.192 *
	X ₂	−0.3737	0.078	−4.798	0.0003			
	X ₅	−1.3347	0.227	−5.880	0.0000			
	X ₈	−0.6819	0.145	−4.711	0.0000			
	X ₁₆	0.7912	0.116	6.837	0.0000			

* The Model I.

Table 5. Efficiency Gaussian, inverted Weibull (IW-Reg), IW-Bayesian regression (IW-BReg) models.

Model	AIC	D	D/df	MSE	Person Statistics
Model I	369.981	71.670	0.833	3.192	57.937
Model II	818.882	0.232	0.003	3.008	
Model III	372.847	26.195	0.305	3.275	19.807
Model IV	447.162	3.152	0.037	3.627	0.011
Model V	369.476	48.737	0.567	3.181	33.636
Model VI	446.981	3.312	0.039	3.093	0.103
Gaussian (Identity)	369.357	270.422	3.144	3.144	
Gaussian (log)	364.041	255.077	2.966	2.966	

Based on this analysis, we also conclude that the Model VI is more appropriate for fitting these data, leading to the following equation

$$\hat{Y} = e^{42.8281 - 0.0443x_{i2} - 0.2023x_{i5} - 0.0976x_{i8} + 0.1308x_{i16}}, \quad i = 1, 2, \dots, 91. \quad (43)$$

For the residuals, the Pearson statistics is $Q = 0.1032$, p -value for Anderson-Darling is 0.0496, and for the Cox Stuart test is 1; see Table 5. This residuals have a large positive residual at the observation 91. However, for the model, this case is non-influential according to $C_{91} = 0.1065 < F_{(0.05,1,n-k)}$ where $F_{(0.05,1,n-k)}$ corresponding to upper α -percentile from the F distribution [50].

Because of the presence of an outlier, we can conclude that the Model VI based on biweight function is the best for our data, and it is given as follows:

$$\hat{Y} = e^{47.259 - 0.049x_{i2} - 0.196x_{i5} - 0.112x_{i8} + 0.136x_{i16}}, \quad i = 1, 2, \dots, 91. \quad (44)$$

From Table 6, we can see that this model has $AIC = 446.515$ and a low $MSE = 3.046$, and there was also a significant relationship among variables when using the level of significance $\alpha = 0.05$. For the residuals, the Pearson statistics is $Q = 0.1049$, the p -value for Anderson-Darling is 0.0612, and the Cox Stuart test is 1. Hence, the Pearson residuals are normal randomly scattered around zero at the level of significant $\alpha = 0.05$; see Table 7 and Figure 5. This Figure shows no large positive residual. The fitting and predicted results for this model during 2016 and 2017 are shown in Table 8. We can also see that the prediction accuracy is good because the TIC value is closer to 0 than 1.

Table 6. AIC, BIC, D, and MSE of the Model VI based biweight function.

Function	Variables	Coefficient Estimate	SE	z-Statistics	p-Value	AIC	D	MSE
No weight	Intercept	42.8281	13.437	3.187	0.0014	446.9806	3.3117	3.093 *
	X_2	−0.0443	0.014	−3.161	0.0016			
	X_5	−0.2023	0.042	−4.867	0.0000			
	X_8	−0.0976	0.026	−3.702	0.0002			
	X_{16}	0.1308	0.018	7.267	0.0000			
Biweight	Intercept	47.259	13.956	3.386	0.0007	446.515	3.777	3.046 **
	X_2	−0.049	0.015	−3.385	0.0007			
	X_5	−0.196	0.043	−4.571	0.0000			
	X_8	−0.112	0.027	−4.070	0.0000			
	X_{16}	0.136	0.019	7.326	0.0000			

* The Model VI that was shown in Equation (43). ** The Model VI based on biweight function that was shown in Equation (44).

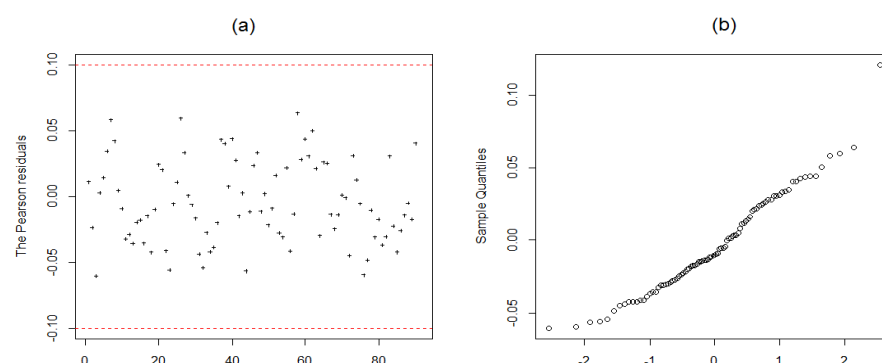


Figure 5. (a) Pearson residuals plot and (b) Normal Q-Q plot of the Pearson residuals for the Model VI based on biweight function as mentioned in Table 6.

Table 7. Anderson-Darling and Cox Stuart test for Pearson residuals of the Model VI.

Function	Anderson-Darling Statistics	<i>p</i> -Value	Cox Stuart Statistics	<i>p</i> -Value
No weight function	0.7483	0.0496	25, <i>n</i> = 45	0.5515
Biweight function	0.7117	0.0612	26, <i>n</i> = 45	0.3713

Table 8. Fitting and Predicted results \hat{y}_i for the Model VI based on biweight function.

Date	y_i	\hat{y}_i	$ y_i - \hat{y}_i $	RE%	Date	y_i	\hat{y}_i	$ y_i - \hat{y}_i $	RE%
Fitting results, 2016					Predicted results, 2017				
7/21	5	5.685	0.685	13.708	8/21	8	10.553	2.553	31.915
7/22	7	5.867	1.133	16.190	8/22	8	7.211	0.789	9.862
7/23	11	8.784	2.217	20.150	8/23	6	6.974	0.974	16.238
7/24	6	6.859	0.859	14.322	8/24	4	3.657	0.343	8.571
7/25	5	4.880	0.120	2.393	8/25	5	4.510	0.491	9.811
7/26	4	5.067	1.067	26.686	8/26	7	7.082	0.082	1.165
7/27	5	5.672	0.672	13.433					
TIC	0.087					0.088			

5. Conclusions

In this paper, the regression models IW-Reg and IW-BReg for modeling Saudi datasets are considered. Zero-one and MGE loss functions were used to attain the Bayesian estimates based on a log and identity functions. In the classical approach, parameter estimation is done by the Fisher's scoring technique, and closed-form expressions are provided for the score function, and for Fisher's information matrix and its inverse. In the Bayesian approach, parameter estimation is performed using a gamma prior distribution, Jacobian transformation, and least-squares estimates. The IW-Reg and IW-BReg models were compared to find which model predicted better. To deal with outlier problems, IW-BReg based on Huber's and biweight functions, and the adopted algorithm based on IRLS to find the estimates, were proposed. For distributional assessment, Q-Q, ECDF, box plots, and the K-S test were applied. Some criteria, namely AIC, D, D/df, and MSE, were also computed for all regression models.

According to the results of the Application (1), the IW-BReg model based on Huber's and MGE loss with a log link function, performed the best in terms of the AIC, D, D/df, and MSE statistics, so it is recommended for these data. In contrast, the IW-Reg model showed poor results compared with those of the other models. Results indicated that the IW-BReg model based on Huber's and MGE loss is highly capable of improving

regression models' performance to a greater extent in predicting the minimum of dry bulb temperatures (Celsius) in Saudi Arabia. It is found the following regressors are significant for the model: Explanatory variables are: X_1 , the mean of relative humidity, X_2 , the mean of vapor pressure (mm); and X_4 , the maximum of station-level pressure (mm). Application (2), the IW-BReg model based on biweight function and MGE loss with a log link function, performed the best in terms of the AIC, D, D/df, and MSE statistics, so it is recommended for these data. In contrast, IW-Reg and IW-BReg based on zero-one loss function showed poor results than those of the other models. Finally, the results in this application indicated that the IW-BReg model based on biweight function and MGE loss with a log link function is highly capable of improving regression models' performance to a greater extent in predicting the mean wind speed (km/h) in Saudi Arabia. It is found the following regressors are significant for the model: Explanatory variables are: X_2 , the mean of station-level pressure (mm); X_5 , the mean of wet-bulb temperatures (Celsius); X_8 , the mean of sky cover oktes; and X_{16} , the minimum of dry bulb temperatures. From these discussions, we conclude that IW-BReg model based on log link and MGE loss has good performance for the response variables in the considered applications.

Author Contributions: Conceptualization, (S.R.A.-D., K.S.S.); methodology, S.R.A.-D.; software, S.R.A.-D.; validation and formal analysis, S.R.A.-D.; resources, (S.R.A.-D., K.S.S.); supervision, K.S.S.; writing—original draft preparation, S.R.A.-D.; writing—review and editing, (S.R.A.-D., K.S.S.). All authors have read and agreed to the published version of the manuscript.

Funding: This article was funded by the Deanship of Scientific Research at King Saud University (RG-1435-056).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this article was taken from the National Center for Meteorology in Saudi Arabia on the link <https://ncm.gov.sa/Ar/About/Branches/Pages/default.aspx>, accessed on 7 December 2020.

Acknowledgments: The authors would like to thank the editor and referees for their helpful comments, which improved the presentation of the paper. In addition, the authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding this Research Group (RG-1435-056).

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

Appendix A. Robust IW-BReg Models

M-estimation is considered to be the most common method of robust regression. It was proposed by [51,52] in the presence of outliers, and it is more efficient than ordinary least squares (OLS) [53,54]. The Huber's function takes the following form:

$$\rho(r) = \begin{cases} \frac{r^2}{2}, & |r| \leq k, \\ k(|r| - \frac{k}{2}), & |r| > k, \end{cases} \quad (\text{A1})$$

where k is the tuning constant, r is the residual corresponding to the observation in OLS, and $\rho(\cdot)$ is the objective function that satisfies certain properties. Often, $\rho(\cdot)$ can be formed by using a linear combination of the residuals. Defining function $\frac{\partial}{\partial r}\rho(r)$ and the corresponding weight function in this case is as follows:

$$\frac{\psi(r)}{r} = w(r) = \begin{cases} 1, & |r| \leq k, \\ \frac{k}{|r|}, & |r| > k. \end{cases} \quad (\text{A2})$$

Another M-estimation function is the Tukey bisquare (biweight) function. This is based on Tukey's function, taking the form of that in Reference [28]

$$\rho(r) = \begin{cases} \frac{k^2}{6}(1 - [1 - (\frac{r}{k})^2]^3), & |r| \leq k, \\ \frac{k^2}{6}, & |r| > k, \end{cases} \quad (\text{A3})$$

where k is the tuning constant, and r is the residual corresponding to the observation in OLS. Defining function $\frac{\partial}{\partial r}\rho(r) = \psi(r)$ and the corresponding weight function in this case is given as follows:

$$\frac{\psi(r)}{r} = w(r) = \begin{cases} [1 - (\frac{r}{k})^2]^2, & |r| \leq k, \\ 0, & |r| > k. \end{cases} \quad (\text{A4})$$

To make the IW-BReg models are robust, we suggest Huber's and biweight functions for these models. There are also many other versions of the M-estimation function that could be used here.

Let the response variable Y have an IW distribution, and let the link function of the form be as given in (2). Consider that α has a gamma prior with density function is given as in (34). Using the Jacobian transformation from α_i to η_i and using the link function, we have the posterior distribution of η_i is given as in (38). Thus, the estimated coefficients $\hat{\beta}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*, \dots, \hat{\beta}_p^*)'$ are given as

$$\hat{\beta}^{*(q)} = \left(X'W^* \left(\hat{\beta}^{*(q-1)} \right) X \right)^{-1} X'W^* \left(\hat{\beta}^{*(q-1)} \right) \hat{\eta}, \quad q = 1, 2, 3, \dots, \quad (\text{A5})$$

where $\hat{\eta}_i = h(y)$ and $\hat{\eta}' = (\hat{\eta}_1, \hat{\eta}_2, \dots, \hat{\eta}_n)$ are the posterior Bayes estimates of η_i using the zero-one or MGE loss functions, and $W^* = \text{diag}(w_1^*, w_2^*, \dots, w_n^*)$, w_i are the selected weights depending on M-estimation functions. In this case, coefficients are estimated using an adopted IRLS algorithm [27,55–57] as follows:

- i. Setting the iteration counter at $q = 0$, finding an initial estimates of regression coefficients $\hat{\beta}_j^{*(q)}$, $j = 0, 1, 2, \dots, p - 1$ using IW-Reg estimates.
- ii. The initial residuals $r_{(i)}^{*(q)} = Y_i - g^{-1}(x_i' \hat{\beta}_j^{*(q)})$ are based on the link function that is given as in (2), and calculate an initial scale estimate $s^{*(q)} = 1.4826(\text{median}|r_i^{*(q)}|)$.
- iii. An initial standardized residuals $u_i^{*(q)} = \frac{r_i^{*(q)}}{s^{*(q)}}$ are calculated and used to calculate initial estimates for the weight function. Preliminary weights are $w_i^{*(q)} = w(u_i^{*(q)})$.
- iv. Calculate Bayes estimates $\hat{\eta}_i = h(y)$; $i = 1, \dots, n$ using a gamma prior $G(v, \lambda)$ with known parameters and zero-one or MGE loss functions.
- vii. Using weights from Steps i–iii and Step iv to find estimators in (A5).
- viii. Set $q = q + 1$; then, go to Step ii. Steps ii to vii are repeated until the estimate of $\hat{\beta}^{*(q)}$ is stabilized from the previous iteration, which means: $|\hat{\beta}^{*(q+1)} - \hat{\beta}^{*(q)}| \leq \varepsilon$.

References

1. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Number 37 in Monographs on Statistics and Applied Probability; Chapman and Hall: London, UK, 1983.
2. Nelder, J.A.; Pregibon, D. An extended quasi-likelihood function. *Biometrika* **1987**, *74*, 221–232. [\[CrossRef\]](#)
3. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [\[CrossRef\]](#)
4. De Jong, P.; Heller, G.Z. *Generalized Linear Models for Insurance Data*; Cambridge University Press: New York, NY, USA, 2008.
5. Yuan, K.H.; Bentler, P.M. Improving the convergence rate and speed of Fisher-scoring algorithm: Ridge and anti-ridge methods in structural equation modeling. *Ann. Inst. Stat. Math.* **2017**, *69*, 571–597. [\[CrossRef\]](#)
6. Liao, T.F. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*; SAGE Publications: Thousand Oaks, CA, USA, 1994.

7. Keller, A.Z.; Kamath, K. Alternate reliability models for mechanical systems. In Proceedings of the 3rd International Conference on Reliability and Maintainability, Toulouse, France, 16–21 October 1982; pp. 411–415.
8. Calabria, R.; Pulcini, G. Confidence limits for reliability and tolerance limits in the inverse Weibull distribution. *Reliab. Eng. Syst. Saf.* **1989**, *24*, 77–85. [\[CrossRef\]](#)
9. Jiang, R.; Zuo, M.J.; Li, H.X. Weibull and inverse Weibull mixture models allowing negative weights. *Reliab. Eng. Amnd Syst. Saf.* **1999**, *66*, 227–234. [\[CrossRef\]](#)
10. Mahmoud, M.A.W.; Sultan, K.S.; Amer, S.M. Order statistics from inverse Weibull distribution and associated inference. *Comput. Stat. Data Anal.* **2003**, *42*, 149–163. [\[CrossRef\]](#)
11. Pasari, S.; Dikshit, O. Impact of three-parameter Weibull models in probabilistic assessment of earthquake hazards. *Pure Appl. Geophys.* **2014**, *171*, 1251–1281. [\[CrossRef\]](#)
12. Jazi, M.A.; Lai, C.D.; Alamatsaz, M.H. A discrete inverse Weibull distribution and estimation of its parameters. *Stat. Methodol.* **2010**, *7*, 121–132. [\[CrossRef\]](#)
13. Kundu, D.; Howlader, H. Bayesian inference and predication of the inverse Weibull distribution for Type-II censoring data. *Comput. Stat. Data Anal.* **2010**, *54*, 1547–1558. [\[CrossRef\]](#)
14. Han, Q.; Li, L.; Gao, X. Statistical inference of the environment factor for inverse weibull distribution. In Proceedings of the 2010 The 2nd Conference on Environmental Science and Information Application Technology, Wuhan, China, 17–18 July 2010; Volume 3, pp. 613–616.
15. Musleh, R.M.; Helu, A. Estimation of the inverse Weibull distribution based on progressively censored data: Comparative study. *Reliab. Eng. Syst. Saf.* **2014**, *131*, 216–227. [\[CrossRef\]](#)
16. Akgul, F.; Senoglu, B.; Arslan, T. An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. *Energy Convers. Manag.* **2016**, *114*, 234–240. [\[CrossRef\]](#)
17. Nassar, M.; Abo-Kasem, O.E. Estimation of the inverse Weibull parameters under adaptive type-II progressive hybrid censoring scheme. *J. Comput. Appl. Math.* **2017**, *315*, 228–239. [\[CrossRef\]](#)
18. Cepeda, E.; Gamerman, D. Bayesian methodology for modeling parameters in the two parameter exponential family. *Rev. Estad.* **2005**, *57*, 93–105.
19. Fahrmeir, L.; Tutz, G. *Multivariate Statistical Modelling Based on Generalized Linear Models*; Springer: Berlin/Heidelberg, Germany, 2013.
20. Dey, D.K.; Ghosh, S.K.; Mallick, B.K. *Generalized Linear Models: A Bayesian Perspective*; CRC Press: New York, NY, USA, 2000.
21. Olsson, U. *Generalized Linear Models, an Applied Approach*; Student Litteratur Lund.: Lund, Sweden, 2002.
22. Dobson, A.J.; Barnett, A. *An Introduction to Generalized Linear Models*; CRC Press: Boca Raton, FL, USA, 2008.
23. Calabria, R.; Pulcini, G. An engineering approach to Bayes estimation for the Weibull distribution. *Microelectron. Reliab.* **1994**, *34*, 789–802. [\[CrossRef\]](#)
24. Dey, D.K.; Ghosh, M.; Srinivasan, C. Simultaneous estimation of parameters under entropy loss. *J. Stat. Plan. Inference* **1986**, *15*, 347–363. [\[CrossRef\]](#)
25. Calabria, R.; Pulcini, G. Point estimation under asymmetric loss functions for left-truncated exponential samples. *Commun. Stat.-Theory Methods* **1996**, *25*, 585–600. [\[CrossRef\]](#)
26. Sano, N.; Suzuki, H.; Koda, M. A robust ensemble learning using zero-one loss function. *J. Oper. Res. Soc. Jpn.* **2008**, *51*, 95–110. [\[CrossRef\]](#)
27. Li, Y.; Hou, L.; Yang, Y.; Tong, J. Huber's M-Estimation-Based Cubature Kalman Filter for an INS/DVL Integrated System. In *Mathematical Problems in Engineering*; Hindawi: London, UK, 2020.
28. Sinova, B.; Van Aelst, S. Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function. *Int. J. Approx. Reason.* **2018**, *93*, 219–237. [\[CrossRef\]](#)
29. Drapella, A. Complementary Weibull distribution: unknown or just forgotten. *Qual. Reliab. Eng. Int.* **1993**, *9*, 383–385 [\[CrossRef\]](#)
30. Mudholkar, G.S.; Kollia, G.D. Generalized Weibull family: A structural analysis. *Commun. Stat. Theory Methods* **1994**, *23*, 1149–1171. [\[CrossRef\]](#)
31. Murthy, D.P.; Xie, M.; Jiang, R. *Weibull Models*; John Wiley and Sons: Hoboken, NJ, USA, 2004; Volume 505.
32. Khan, M.S.; Pasha, G.R.; Pasha, A.H. Theoretical analysis of inverse Weibull distribution. *WSEAS Trans. Math.* **2008**, *7*, 30–38.
33. De Gusmao, F.R.; Ortega, E.M.; Cordeiro, G.M. The generalized inverse Weibull distribution. *Stat. Pap.* **2011**, *52*, 591–619. [\[CrossRef\]](#)
34. Singh, S.; Singh, U.; Sharma, V. Bayesian prediction of observations from inverse Weibull distribution based on Type-II hybrid censored sample. *Int. J. Adv. Stat. Probab.* **2013**, *1*, 32–43. [\[CrossRef\]](#)
35. Elbatal, I.; El Gebaly, Y.M.; Amin, E.A. The Beta Generalized Inverse Weibull Geometric Distribution. *Pak. J. Stat. Oper. Res.* **2017**, *75*–90. [\[CrossRef\]](#)
36. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; CRC Press: Boca Raton, FL, USA, 1989; Volume 37.
37. Muhammed, H.Z. Bivariate inverse Weibull distribution. *J. Stat. Comput. Simul.* **2016**, *86*, 2335–2345. [\[CrossRef\]](#)
38. Ferrari, S.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [\[CrossRef\]](#)
39. Houston, W.M.; Woodruff, D.J. *Empirical Bayes Estimates of Parameters from the Logistic Regression Model*; ACT Research Report Series 97-6; ACT, Inc.: Iowa, IA, USA, 1997; 34p.
40. Das S.; Dey, D.K. *On Bayesian Analysis of Generalized Linear Models: A New Perspective, Technical Report*; University of Connecticut, Department of Statistics: Storrs, CT, USA, 2007; 33p.

41. Das S.; Dey, D.K. On Bayesian analysis of generalized linear models using the Jacobian technique. *Am. Stat.* **2006**, *60*, 264–268. [[CrossRef](#)]
42. Tellinghuisen, J. Least squares with non-normal data: Estimating experimental variance functions. *Analyst* **2008**, *133*, 161–166. [[CrossRef](#)]
43. Clarkson, E. Bayesian Fisher Information and Detection of a Small Change in a Parameter. In Proceedings of the 2020 54th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 18–20 March 2020; pp. 1–5.
44. Leuthold, R.M. On the use of Theil's inequality coefficients. *Am. J. Agric. Econ.* **1975**, *57*, 344–346. [[CrossRef](#)]
45. Niu, T.; Zhang, L.; Zhang, B.; Yang, B.; Wei, S. An Improved Prediction Model Combining Inverse Exponential Smoothing and Markov Chain. In *Mathematical Problems in Engineering*; Hindawi: London, UK, 2020; 11p.
46. Faraway, J.J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*; CRC Press: Boca Raton, FL, USA, 2016.
47. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*; Sage Publications: Thousand Oaks, CA, USA, 2018.
48. Soetaert, K. *rootSolve: Nonlinear Root Finding, Equilibrium and Steady-State Analysis of Ordinary Differential Equations*; R Package 1.6; 2009. Available online: <https://cran.r-project.org/web/packages/rootSolve/index.html> (accessed on 7 December 2020).
49. Agresti, A. *Foundations of Linear and Generalized Linear Models*; John Wiley and Sons: Hoboken, NJ, USA, 2015.
50. Diaz-Garcia, J.A.; Gonzalez-Farias, G. A note on the Cook's distance. *J. Stat. Plan. Inference* **2004**, *120*, 119–136. [[CrossRef](#)]
51. Huber, P.J. *Robust Statistics*; John Wiley and Sons: Hoboken, NJ, USA, 1981.
52. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [[CrossRef](#)]
53. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*; John Wiley and Sons: Hoboken, NJ, USA, 1987.
54. Chang, L.; Hu, B.; Chang, G.; Li, A. Robust derivative-free Kalman filter based on Huber's M-estimation methodology. *J. Process Control* **2013**, *23*, 1555–1561. [[CrossRef](#)]
55. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics: Theory and Methods*; John Wiley and Sons: Hoboken, NJ, USA, 2006.
56. Wen, F.; Liu, W. Iteratively reweighted optimum linear regression in the presence of generalized Gaussian noise. In Proceedings of the 2016 IEEE International Conference on Digital Signal Processing (DSP), Beijing, China, 16–18 October 2016; pp. 657–661.
57. Kikuchi, H.; Yasunaga, H.; Matsui, H.; Fan, C.I. Efficient privacy-preserving logistic regression with iteratively Re-weighted least squares. In Proceedings of the 2016 11th Asia Joint Conference on Information Security (AsiaJCIS), Fukuoka, Japan, 4–5 August 2016; pp. 48–54.