Journal of Econometrics and Statistics
Vol. 3, Issue 1, 2023, pp. 55-68
https://doi.org/10.47509/JES.2023.v03i01.04

The Unit Inverse Weibull Distribution and its Associated Regression Model

Lucas David Ribeiro-Reis
Department of Statistics
Federal University of Pernambuco
Recife-PE, Brazil
E-mail: econ.lucasdavid@gmail.com

Abstract

In this paper, from the inverse Weibull distribution, a new bi-parametric distribution is proposed. This new distribution has support in $(0, 1)$, being an alternative to other distributions for double limited data analysis. Based on this distribution, a regression model is proposed. This model has a regression structure at the median and is called the unit inverse Weibull regression model. The maximum likelihood method is used to obtain the estimates of the unknown parameters. Analytical expressions for the score vector and for the Fisher observed information matrix are performed. The good performance of the maximum likelihood estimators, for the proposed model, is demonstrated through Monte Carlo simulations. The usefulness of the model is shown through an application to real data. In this application, the proposed model is superior to beta, Kumaraswamy and unit Weibull regression models.
Keywords: Unit inverse Weibull; Quantile function; Maximum likelihood; Regression model; Monte Carlo simulation
Mathematical Subject Classification 2010: 60E05, 62F12, 65C05, 65C10, 62Jxx

1 Introduction

In data analysis, several phenomena are defined in the unit interval, such as rates, proportions, and the human development index. The distributions most often used to model these types of data are the beta and Kumaraswamy distributions.

In recent years there has been a growing number of proposals of new distributions for the unit interval. These new distributions are alternatives to the beta and Kumaraswamy distributions. Some of these new distributions are: log-Lindley (Gómez-Déniz et al., 2014), unit-Weibull (Mazucheli et al., 2018), log-xgamma (Altun and Hamedani, 2018), unit Gompertz (Mazucheli et al., 2019), log-Bilal (Altun et al., 2021), unit log-logistic (Ribeiro-Reis, 2021).

In this paper, based on a transformation of the inverse Weibull distribution, a new distribution for the interval $(0, 1)$ is proposed, called the unit inverse Weibull. This new distribution, is asymmetric and has simple forms for the cumulative distribution function (cdf) and for probability density function (pdf), in contrast to the beta distribution, which depends on special functions such as the gamma function and the incomplete beta function.

Article History
Received : 08 March 2023; Revised : 28 March 2023; Accepted : 11 April 2023; Published : 27 May 2023

To cite this paper
Lucas David Riberiro-Reis (2023). The Unit Inverse Weibull Distribution and its Associated Regression Model. *Journal of Econometrics and Statistics.* 3(1), 55-68. https://doi.org/10.47509/JES.2023.v03i01.04

Under reparameterization, a new regression model for the unit interval is introduced. Here, the response variable is assumed to follow the unit inverse Weibull distribution and the regression structure is done on the median. Application to real data shows that this regression model is better than the popular beta (Ferrari and Cribari-Neto, 2004) and Kumaraswamy (Mitnik and Baek, 2013) regression models.

The rest of the paper is organized as follows. In Section 2, the unit inverse Weibull distribution is introduced. In Section 3, a reparameterization on the median is taken. In Section 4, a regression model is proposed. The estimates of the unknown parameters by the maximum likelihood method is discussed. In Sections 5 and 6, respectively, Monte Carlo simulations and application to real data show the good performance of the regression model. Finally, the Section 7 concludes the paper.

2 Unit inverse Weibull distribution

The Weibull distribution is one of the most famous continuous distributions. It is a distribution that is indexed by two parameters and plays a very important role in positive data analysis, mainly in survival analysis. The cdf of the Weibull distribution is defined by

$$F_{\text{W}}(v;\lambda,\phi) = 1 - \exp\left[-\lambda v^{\phi}\right], \quad v > 0,$$

where where $\lambda > 0$ is scale parameter and $\phi > 0$ is shape parameter.

Let V is a random variable with Weibull distribution, then the random variable $T = 1/V$, will have inverse Weibull (IW) distribution, and its cdf and pdf are given by

$$F_{\text{IW}}(t;\lambda,\phi) = \exp\left[-\lambda t^{-\phi}\right], \quad t > 0$$

and

$$f_{\text{IW}}(t;\lambda,\phi) = \lambda\phi t^{-\phi-1}\exp\left[-\lambda t^{-\phi}\right], \quad t > 0,$$

respectively, where $\lambda > 0$ is scale parameter and $\phi > 0$ is shape parameter.

Taking $Y = \exp(-T)$, the cdf and pdf of Y are

$$F(y;\lambda,\phi) = 1 - \exp\left[-\lambda(-\log y)^{-\phi}\right], \quad 0 < y < 1 \tag{1}$$

and

$$f(y;\lambda,\phi) = \frac{\lambda\phi}{y}(-\log y)^{-\phi-1}\exp\left[-\lambda(-\log y)^{-\phi}\right], \quad 0 < y < 1, \tag{2}$$

respectively. The random variable Y with pdf (2) has unit inverse Weibull (UIW) distribution and is denoted as $Y \sim \text{UIW}(\lambda,\phi)$. Figure 1 presents some forms for the density function (2). Note that the density function of $Y \sim \text{UIW}(\lambda,\phi)$ can have decreasing, U-shaped, U-unimodal, right-skewed and left-skewed shapes.

The corresponding failure rate function (frf) of Y is

$$r(y;\lambda,\phi) = \frac{\lambda\phi}{y}(-\log y)^{-\phi-1}, \quad 0 < y < 1.$$

Note that the parameter λ is of acceleration, it does not influence in the shape of the frf. The first derivative of the log-frf is

$$\xi(y) = \frac{d}{dy}\log r(y;\lambda,\phi) = -\frac{1}{y} - (\phi+1)\frac{1}{y\log y}.$$

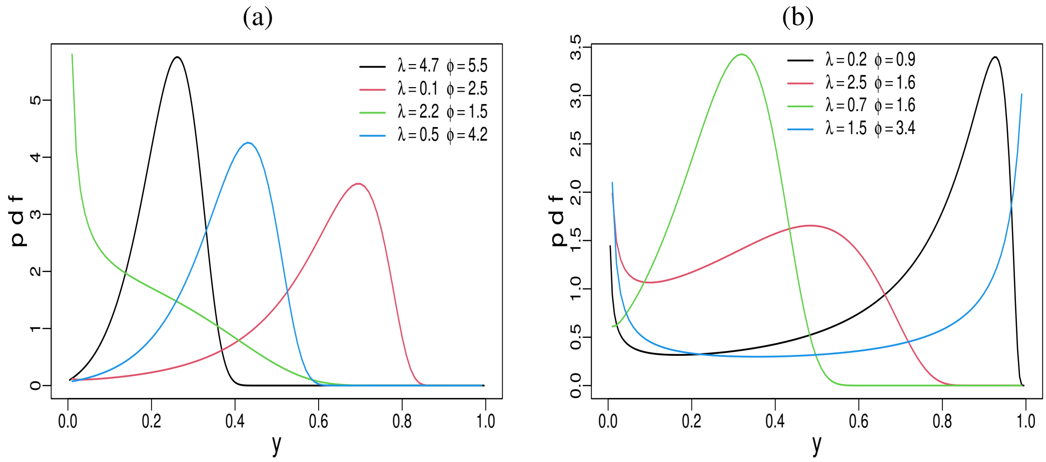

Figure 1: Plots of the pdf of Y for selected parameters.

From $\xi(y) = 0$, the root is $y_0 = \exp(-\phi - 1)$. Note that $y_0 \in (0, 1)$ of fact. The parameter ϕ controls the shape of the frf of Y, and it is seen that $\lim_{\phi\to 0} y_0 = \exp(-1)$ and $\lim_{\phi\to\infty} y_0 = 0$. Then for large values of ϕ, y_0 tends to 0 and the frf is increasing. In turn, for very small values of ϕ, y_0 tends to $\exp(-1)$, and then the frf of Y assumes the shape of bathtub.

Taking second derivative of the log-frf,

$$\xi'(y) = \frac{d^2}{dy^2}\log r(y;\lambda,\phi) = \frac{1}{y^2} + (\phi+1)\frac{1+\log y}{y^2(\log y)^2}.$$

Evaluating in y_0,

$$\xi'(y_0) = \frac{1}{y_0^2(1+2\phi+\phi^2)} > 0.$$

So, y_0 corresponds to a minimum point. Thus, the frf can have the shape of a bathtub, which is a very high merit for a bi-parametric distribution. Figure 2 shows some shapes of the frf of Y. It is noted that the frf of Y can assume the increasing and bathtub forms.

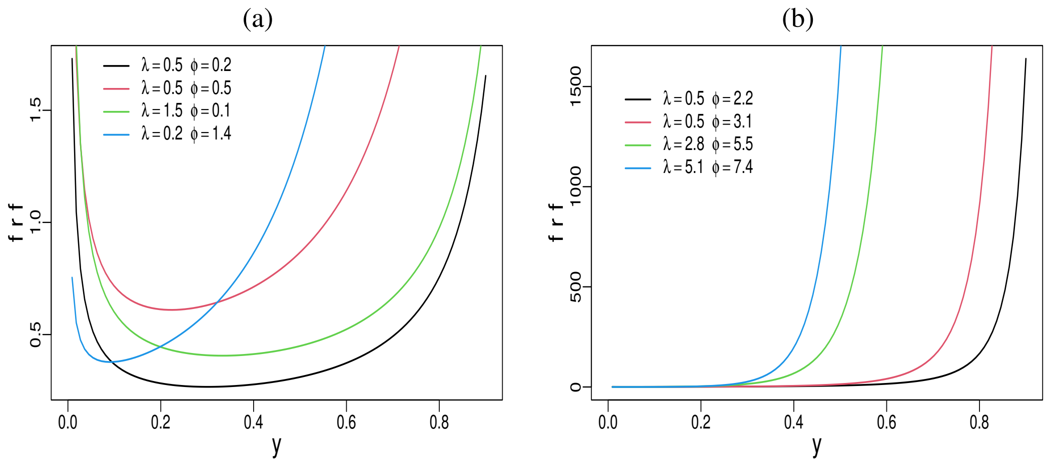

Figure 2: Plots of the frf of Y for selected parameters.

By inverting $F(y;\lambda,\phi)=u$, the quantile function is given by

$$Q(u;\lambda,\phi)=\exp\left\{-[-\lambda^{-1}\log(1-u)]^{-1/\phi}\right\},\quad 0<u<1.$$

For $u=1/2$, the median is obtained. So, the median of Y is

$$\text{median}(Y)=\exp\left\{-[-\lambda^{-1}\log 0.5]^{-1/\phi}\right\}.$$

3 Parameterization in the median

Taking median$(Y)=\tau$ and solving for λ, it follows

$$\lambda=-\frac{\log 0.5}{(-\log\tau)^{-\phi}}.$$

Under parameterization on median, the cdf (1) and pdf (2) becomes

$$F(y;\tau,\phi)=1-2^{-\left(\frac{\log y}{\log\tau}\right)^{-\phi}} \tag{3}$$

and

$$f(y;\tau,\phi)=\frac{\phi\log 0.5}{y\log\tau}\left(\frac{\log y}{\log\tau}\right)^{-\phi-1}2^{-\left(\frac{\log y}{\log\tau}\right)^{-\phi}}, \tag{4}$$

respectively, where $0<\tau<1$ is the median parameter and $\phi>0$ is shape parameter. The random variable with pdf (4) is denoted as $Y\sim \text{UIW}(\tau,\phi)$. Figure 3 presents some shapes of the density function of $Y\sim \text{UIW}(\tau,\phi)$.

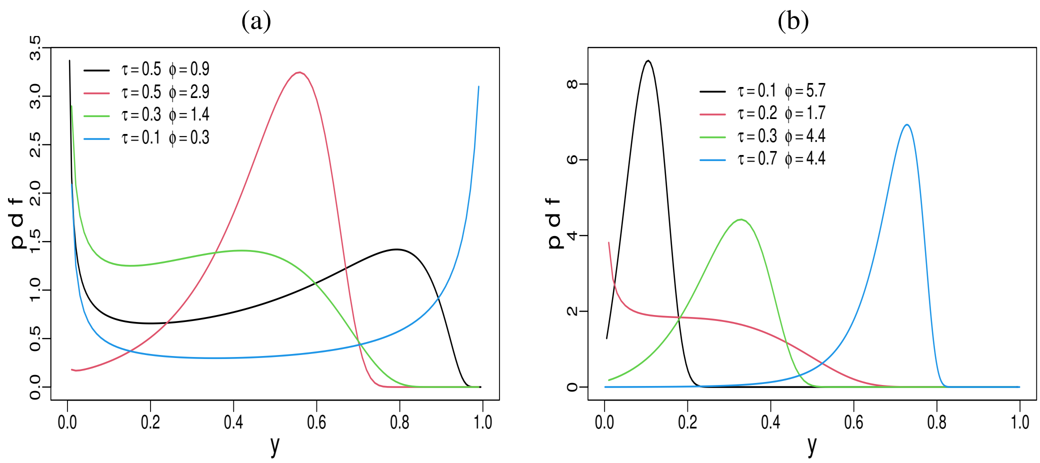

Figure 3: Plots of the pdf of Y under parameterization in the median, for selected parameters.

The quantile function corresponding to cdf (3) is

$$Q(u;\tau,\phi)=\exp\left\{\log(\tau)\left[\frac{\log(1-u)}{\log 0.5}\right]^{-1/\phi}\right\},\quad 0<u<1.$$

Note that $Q(0.5; \tau, \phi) = \tau$, as expected. Since that τ denotes the median of $Y \sim \text{UIW}(\tau, \phi)$. Using the quantile function, the random variable

$$Y = \exp\left\{\log(\tau)\left[\frac{\log(1-U)}{\log 0.5}\right]^{-1/\phi}\right\} \tag{5}$$

has density function (4), where U is a uniform random variable over the interval $(0, 1)$.

4 Regression model

Let the independent random variables $Y_i \sim \text{UIW}(\tau_i, \phi)$, with observed values y_i, $i = 1, \cdots, n$. The proposed regression model for the median of y_i is given by

$$\eta_i = g(\tau_i) = \mathrm{x}_i^\top \beta = \sum_{m=1}^{k} x_{im}\beta_m, \tag{6}$$

where $\beta = (\beta_1, \ldots, \beta_k)^\top$ is k-vector of unknown parameters, $\mathrm{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ is vector of k covariates $(k < n)$, which are assumed fixed and known and η_i is the linear predictor. For model with intercept, it is assumed that $x_{i1} = 1, \forall\, i$. The $g(\cdot)$ is a link function strictly monotonic and twice differentiable. Since $\tau \in (0, 1)$, the $g(\cdot)$ link function must map the interval $(0, 1)$ to the reals, i.e., $g : (0, 1) \to \mathbb{R}$. Examples of some link functions can be the quantile functions of the following distributions: standard logistic, standard Gumbel type I, standard Gumbel type II and standard Cauchy. Thus, these link functions are:

- logit: $g(\tau) = \log[\tau/(1-\tau)]$;
- Gumbel type I: $g(\tau) = -\log(-\log \tau)$;
- Gumbel type II: $g(\tau) = \log[-\log(1-\tau)]$;
- Cauchy: $g(\tau) = \tan(\pi(\tau - 0.5))$.

Several estimation methods can be adopted to estimate the model parameters, such as ordinary least squares, percentiles, maximum product of spacing, Bayesian and maximum likelihood. By simplicity and for having an asymptotic distribution for the estimators, the maximum likelihood method is then adopted in this paper.

The log-likelihood function for a sample of n independent observations with pdf (4), under the structure of the regression model (6), is given by

$$\mathcal{L}(\beta, \phi) = \sum_{i=1}^{n} \mathcal{L}_i(\tau_i, \phi),$$

where

$$\begin{aligned}\mathcal{L}_i(\tau_i, \phi) = &\log \phi + \log\left(\frac{\log 0.5}{\log \tau_i}\right) - \log y_i - (\phi + 1)\log\left(\frac{\log y_i}{\log \tau_i}\right) \\ &- \log(2)\left(\frac{\log \tau_i}{\log y_i}\right)^{\phi}.\end{aligned}$$

The first derivatives of $\mathcal{L}_i(\tau_i, \phi)$ with respect to τ_i and ϕ are

$$\begin{aligned}
\frac{\partial \mathcal{L}_i(\tau_i,\phi)}{\partial \tau_i} &= -\frac{1}{\tau_i \log \tau_i} + \frac{\phi+1}{\tau_i \log \tau_i} - \frac{\phi \log(2)}{\tau_i \log y_i}\left(\frac{\log \tau_i}{\log y_i}\right)^{\phi-1} \\
&= \frac{\phi}{\tau_i \log \tau_i} - \frac{\phi \log(2)}{\tau_i \log \tau_i}\left(\frac{\log \tau_i}{\log y_i}\right)^{\phi} \\
&= b_i(1-\dot{y}_i), \qquad (7) \\
\frac{\partial \mathcal{L}_i(\tau_i,\phi)}{\partial \phi} &= \frac{1}{\phi} + \log\left(\frac{\log \tau_i}{\log y_i}\right) - \log(2)\left(\frac{\log \tau_i}{\log y_i}\right)^{\phi} \log\left(\frac{\log \tau_i}{\log y_i}\right) \\
&= \frac{1}{\phi} + \ddot{y}_i - \dot{y}_i \ddot{y}_i, \qquad (8)
\end{aligned}$$

respectively, where $b_i = \frac{\phi}{\tau_i \log \tau_i}$, $\dot{y}_i = \log(2)\left(\frac{\log \tau_i}{\log y_i}\right)^{\phi}$ and $\ddot{y}_i = \log\left(\frac{\log \tau_i}{\log y_i}\right)$.

The differential total of $\mathcal{L}(\beta, \phi)$ is given by

$$\begin{aligned}
U_{\beta_j}(\beta,\phi) &= \frac{\partial \mathcal{L}(\beta,\phi)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \mathcal{L}_i(\tau_i,\phi)}{\partial \tau_i} \frac{d\tau_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j}, \\
U_{\phi}(\beta,\phi) &= \frac{\partial \mathcal{L}(\beta,\phi)}{\partial \phi} = \sum_{i=1}^{n} \frac{\partial \mathcal{L}_i(\tau_i,\phi)}{\partial \phi}.
\end{aligned}$$

Note that, $d\tau_i/d\eta_i = 1/g'(\tau_i)$ and $\partial \eta_i/\partial \beta_j = x_{ij}$, then the score vector of β_j and ϕ are given by

$$\begin{aligned}
U_{\beta_j}(\beta,\phi) &= \sum_{i=1}^{n} \frac{w_i}{g'(\tau_i)} x_{ij}, \\
U_{\phi}(\beta,\phi) &= \sum_{i=1}^{n} c_i,
\end{aligned}$$

respectively, where $g'(\tau_i) = dg(\tau_i)/d\tau_i$, $w_i = b_i(1-\dot{y}_i)$ and $c_i = \frac{1}{\phi} + \ddot{y}_i - \dot{y}_i\ddot{y}_i$.

In matrix form, the score vector of β and ϕ can be written as $U_\beta(\beta,\phi) = X^\top Q\, w$ and $U_\phi(\beta,\phi) = c^\top \mathbf{1}_n$, respectively, where X is a $n \times k$ matrix whose ith row is $\mathrm{x}_i^\top$, $Q = \mathrm{diag}\{1/g'(\tau_1), \cdots, 1/g'(\tau_n)\}$ (diagonal matrix), $w = (w_1, \cdots, w_n)^\top$, $c = (c_1, \cdots, c_n)^\top$ and $\mathbf{1}_n$ is a n-dimensional vector of 1's.

The maximum likelihood estimators (MLEs) of β and ϕ, says $\hat{\beta}$ and $\hat{\phi}$, are the solutions of

$$\begin{cases} U_\beta(\beta,\phi) = \mathbf{0}, \\ U_\phi(\beta,\phi) = 0, \end{cases}$$

which do not have closed-form expressions. Thus, these estimates are obtained through numerical optimization. These days with advanced computing, it can be done without any difficulty. These iterative processes require initial values for the parameters. Initial guesses for β can be the ordinary least squares estimator of the regression of $g(y)$ on X. So, initial guess for β is $\beta^{(0)} = (X^\top X)^{-1} X^\top g(y)$. For ϕ, the initial guess is $\phi^{(0)} = 1$. Some software that can be used to obtain these MLEs are: Ox Console (Doornik, 2018) with `MaxBFGS` function and R Project (R Core Team, 2020) with `optim` or `nlminb` functions.

4.1 Observed information matrix

Note that, the second derivatives of $\mathcal{L}(\beta,\phi)$ are given by

$$
\begin{aligned}
\frac{\partial^2\mathcal{L}(\beta,\phi)}{\partial\beta_j\partial\beta_l} =& \sum_{i=1}^{n}\frac{\partial}{\partial\tau_i}\left(\frac{\partial\mathcal{L}_i(\tau_i,\phi)}{\partial\tau_i}\frac{1}{g'(\tau_i)}x_{ij}\right)\frac{d\tau_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_l}\\
=& \sum_{i=1}^{n}\left(\frac{\partial^2\mathcal{L}_i(\tau_i,\phi)}{\partial\tau_i^2}\frac{1}{g'(\tau_i)^2}x_{ij}x_{il} - \frac{\partial\mathcal{L}_i(\tau_i,\phi)}{\partial\tau_i}\frac{g''(\tau_i)}{g'(\tau_i)^3}x_{ij}x_{il}\right),\\
\frac{\partial^2\mathcal{L}(\beta,\phi)}{\partial\beta_j\partial\phi} =& \sum_{i=1}^{n}\frac{\partial^2\mathcal{L}_i(\tau_i,\phi)}{\partial\tau_i\partial\phi}\frac{1}{g'(\tau_i)}x_{ij},\\
\frac{\partial^2\mathcal{L}(\beta,\phi)}{\partial\phi^2} =& \sum_{i=1}^{n}\frac{\partial^2\mathcal{L}_i(\tau_i,\phi)}{\partial\phi^2},
\end{aligned}
$$

where $g''(\tau_i) = d^2g(\tau_i)/d\tau_i^2$. If $g(\cdot)$ logit link function, then $g'(\tau_i) = 1/(\tau_i - \tau_i^2)$ and $g''(\tau_i) = -(1-2\tau_i)/(\tau_i-\tau_i^2)^2$. Already for $g(\cdot)$ Cauchy link function, $g'(\tau_i) = \pi/\cos(\pi(\tau_i - 0.5))^2$ and $g''(\tau_i) = 2\pi^2\tan(\pi(\tau_i-0.5))/\cos(\pi(\tau_i-0.5))^2$.

From Equations (7) and (8),

$$
\begin{aligned}
\frac{\partial\mathcal{L}_i^2(\tau_i,\phi)}{\partial\tau_i^2} =& -\frac{\phi(1+\log\tau_i)}{(\tau_i\log\tau_i)^2}(1-\dot{y}_i) - b_i^2\dot{y}_i = p_i,\\
\frac{\partial\mathcal{L}_i^2(\tau_i,\phi)}{\partial\tau_i\partial\phi} =& \frac{1-\dot{y}_i}{\tau_i\log\tau_i} - b_i\dot{y}_i\ddot{y}_i = r_i,\\
\frac{\partial\mathcal{L}_i^2(\tau_i,\phi)}{\partial\phi^2} =& -\frac{1}{\phi^2} - \dot{y}_i\ddot{y}_i^2 = s_i.
\end{aligned}
$$

Finally, the second derivatives are

$$
\begin{aligned}
\frac{\partial^2\mathcal{L}(\beta,\phi)}{\partial\beta_j\partial\beta_l} =& \sum_{i=1}^{n}\left(\frac{p_i}{g'(\tau_i)^2}x_{ij}x_{il} - w_i\frac{g''(\tau_i)}{g'(\tau_i)^3}x_{ij}x_{il}\right),\\
\frac{\partial^2\mathcal{L}(\beta,\phi)}{\partial\beta_j\partial\phi} =& \sum_{i=1}^{n}\frac{r_i}{g'(\tau_i)}x_{ij},\\
\frac{\partial^2\mathcal{L}(\beta,\phi)}{\partial\phi^2} =& \sum_{i=1}^{n}s_i.
\end{aligned}
$$

Let $P = \text{diag}\{p_1,\cdots,p_n\}$, $R = \text{diag}\{w_1g''(\tau_1),\cdots,w_ng''(\tau_n)\}$, $r = (r_1,\cdots,r_n)^\top$ and $s = (s_1,\cdots,s_n)^\top$.

In matrix form, this expressions can be written as

$$
\begin{aligned}
U_{\beta\beta}(\beta,\phi) =& X^\top P\,Q^2X - X^\top R\,Q^3X,\\
U_{\beta\phi}(\beta,\phi) =& U_{\phi\beta}(\beta,\phi)^\top = X^\top Q\,r,\\
U_{\phi\phi}(\beta,\phi) =& s^\top\mathbf{1}_n.
\end{aligned}
$$

The Fisher observed information matrix is given by

$$
\begin{aligned}
\mathcal{J}(\beta,\phi) =& -\begin{bmatrix} U_{\beta\beta}(\beta,\phi) & U_{\beta\phi}(\beta,\phi)\\ U_{\phi\beta}(\beta,\phi) & U_{\phi\phi}(\beta,\phi)\end{bmatrix}\\
=& -\begin{bmatrix} X^\top P\,Q^2X - X^\top R\,Q^3X & X^\top Q\,r\\ r^\top Q\,X & s^\top\mathbf{1}_n\end{bmatrix}.
\end{aligned}
$$

Under the usual regularity conditions of the MLEs, for n large,

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \quad \overset{a}{\sim} \quad \mathcal{N}_{k+1}\left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, \mathcal{J}(\beta,\phi)^{-1}\right),$$

where $\overset{a}{\sim}$ denotes asymptotic distribution. So, confidence intervals and hypothesis testing can be performed using the normal distribution. Based on asymptotic distribution, the $100(1-\gamma)\%$ confidence intervals for β_j and ϕ are given by

$$\hat{\beta}_j \pm z_{(1-\gamma/2)}\sqrt{L_{jj}} \quad \text{and} \quad \hat{\phi} \pm z_{(1-\gamma/2)}\sqrt{L_{(j+1)(j+1)}}, \quad j = 1, \cdots, k, \tag{9}$$

respectively, where $z_{(1-\gamma/2)}$ is the $(1-\gamma/2)$ quantile of the standard normal distribution and L_{jj} denotes the jth diagonal element of the matrix $\mathcal{J}(\beta,\phi)^{-1}$.

4.2 Diagnostic analysis

The analysis of residuals is crucial to check the adequacy of the estimated model. Here, we will discuss the Dunn-Smith residuals (Dunn and Smyth, 1996) residuals, which are given by

$$\hat{r}_i = Q_N(F(y_i; \hat{\tau}_i, \hat{\phi})),$$

where $F(y_i; \hat{\tau}_i, \hat{\phi})$ is the cdf (3) evaluated in $\hat{\tau}_i$ and $\hat{\phi}$, and $Q_N(\cdot)$ is the quantile function of the standard normal distribution. If the model is valid, the Dunn-Smith residuals approximately follow the standard normal distribution. Thus, the Dunn-Smith residuals have a behavior around zero with about 95% of the values in the interval $(-2, 2)$.

Simulated envelope plot (Atkinson, 1985) of the Dunn-Smith residuals can be used to check the quality of the estimated model. The simulated envelope plot can be produced as follows:

1. Estimate the model and generate a sample with n observations, considering the estimated model as the true model;
2. Estimate the model of the generated sample and calculate the ordered Dunn-Smith residuals;
3. repeat the steps 1 and 2 k times;
4. From the k groups of the Dunn-Smith ordered residuals, compute the mean, median, minimum and maximum.

The envelope is determined by the minimum and maximum values. If a large number of points of the Dunn-Smith residuals are outside the envelope, then this is evidence against the adequacy of the estimated model.

5 Simulation

In this Section to show the performance of the MLEs of the UIW regression model, two Monte Carlo simulations are performed. The behavior of the MLEs is evaluated through the average estimates (AEs), the mean square errors (MSEs) and the coverage rates (CRs) of the confidence intervals, calculated from Equation (9). The simulations are carried out with sample size of $n = \{30, 60, 100, 200, 300\}$ and with $R = 10,000$ replications.

In each experiment the values of the true parameters are the same, only the link function used is changed. In the first experiment, the logit link function is chosen (Model I). In the second experiment, the Cauchy link function is used (Model II). The simulations were made using the matrix programming language Ox Console (Doornik, 2018) through of the `MaxBFGS` function, with analytical derivatives.

The simulated models are given by

$$\begin{array}{ll} \text{Modelo I:} & \log\left(\frac{\tau_i}{1-\tau_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} \\ \text{Modelo II:} & \tan(\pi(\tau_i - 0.5)) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} \end{array}, \quad i = 1, \cdots, n,$$

where as covariates are generated from the standard normal distribution and are kept fixed in the simulations. The true parameters adopted are: $\beta_1 = 0.7$, $\beta_2 = 1.5$, $\beta_3 = -0.8$, $\beta_4 = -2.4$ and $\phi = 3.3$.

The following mechanism is used to simulate the response variable:

1. Generate $x_{im} \sim \mathcal{N}(0, 1)$, $m = 2, 3, 4$;
2. Write $\eta_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$ and obtain $\tau_i = g^{-1}(\eta_i) = \mathrm{e}^{\eta_i}/(1 + \mathrm{e}^{\eta_i})$ (for Modelo I) and $\tau_i = g^{-1}(\eta_i) = \frac{1}{\pi}\arctan(\eta_i) + 0.5$ (for Modelo II);
3. From Equation (5), generate $y_i \sim \mathrm{UIW}(\tau_i, \phi)$.

The simulation results for models I and II are given in Tables 1 and 2, respectively. As can be seen, in both models, when the sample size increases, the MLEs tend towards the true parameters and the MSEs decrease. Note also that when n grows, the CRs approach of the true nominal levels. All these results show the consistency of the MLEs for the UIW regression model.

6 Application

In this section, the usefulness of the regression model in practice is shown from an application to real data. The competitive regression models are the beta (Ferrari and Cribari-Neto, 2004), Kumaraswamy (Mitnik and Baek, 2013) and the unit Weibull Mazucheli et al. (2020) regressions models. The pdfs (for $0 < y < 1$) of the beta, Kumaraswamy and unit Weibull regressions models are given by

$$f_{\mathrm{B}}(y;\tau,\phi) = \frac{\Gamma(\phi)}{\Gamma(\tau\phi)\Gamma((1-\tau)\phi)} y^{\tau\phi-1}(1-y)^{(1-\tau)\phi-1},$$

$$f_{\mathrm{K}}(y;\tau,\phi) = \frac{\phi\log 0.5}{\log(1-\tau^{\phi})} y^{\phi-1}(1-y^{\phi})^{\frac{\log 0.5}{\log(1-\tau^{\phi})}-1}$$

and

$$f_{\mathrm{UW}}(y;\tau,\phi) = \frac{\phi\log 0.5}{y\log\tau}\left(\frac{\log y}{\log\tau}\right)^{\phi-1} 2^{-\left(\frac{\log y}{\log\tau}\right)^{\phi}},$$

where $0 < \tau < 1$ denotes the mean (for beta model) and median (for Kumaraswamy and unit Weibull models), $\phi > 0$ is a shape parameter and $\Gamma(p) = \int_0^\infty u^{p-1}\mathrm{e}^{-u}du$, $p > 0$, is the gamma function. For beta and Kumaraswamy models, ϕ can be interpreted as precision parameter.

The information criteria adopted for choosing the best model are: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Hannan–Quinn Information Criterion (HQIC).

$$\mathrm{AIC} = -2\mathcal{L}(\hat{\boldsymbol{\eta}}) + 2p,$$

$$\mathrm{BIC} = -2\mathcal{L}(\hat{\boldsymbol{\eta}}) + p\log n$$

Table 1: Simulation results for Modelo I.

n	Par	AE	MSE	CR		
				90%	95%	99%
30	β_1	0.69672	0.00966	86.06	91.50	97.48
	β_2	1.49424	0.00606	85.44	91.26	97.29
	β_3	−0.79703	0.00859	85.57	91.38	96.90
	β_4	−2.38082	0.01208	84.60	90.71	96.50
	ϕ	3.70372	0.51265	84.13	91.86	98.46
60	β_1	0.69871	0.00503	88.16	93.58	98.21
	β_2	1.49596	0.00310	87.63	93.14	98.07
	β_3	−0.79712	0.00305	88.13	93.47	98.32
	β_4	−2.39031	0.00602	86.97	92.75	97.92
	ϕ	3.48830	0.17335	86.86	93.46	98.66
100	β_1	0.69922	0.00307	88.51	94.03	98.48
	β_2	1.49735	0.00186	88.59	93.79	98.57
	β_3	−0.79794	0.00196	89.03	94.56	98.62
	β_4	−2.39748	0.00289	88.86	94.08	98.59
	ϕ	3.41008	0.08598	88.39	94.02	98.93
200	β_1	0.69998	0.00157	89.15	94.33	98.69
	β_2	1.49848	0.00101	89.71	94.75	98.86
	β_3	−0.79888	0.00096	89.26	94.30	98.69
	β_4	−2.39847	0.00130	89.47	94.76	98.74
	ϕ	3.35650	0.03836	89.02	94.40	99.01
300	β_1	0.69965	0.00101	89.79	94.82	98.89
	β_2	1.49953	0.00065	89.67	94.92	98.91
	β_3	−0.79973	0.00059	89.48	94.26	98.79
	β_4	−2.39878	0.00078	89.17	94.30	98.81
	ϕ	3.33467	0.02411	89.18	94.86	98.95

and

$$\text{HQIC} = -2\mathcal{L}(\hat{\boldsymbol{\eta}}) + 2p\log(\log n),$$

where $\mathcal{L}(\hat{\boldsymbol{\eta}})$ is the log-likelihood evaluated in the MLE $\hat{\boldsymbol{\eta}}$, p is the number of parameters in the model and n is the number of observations. The best model is the one with the lowest values of these information criteria.

The data from this application contains 32 observations concerning gasoline yield (Prater, 1956). The response variable (y) is proportion of crude oil converted to gasoline after distillation and fractionation. The exogenous variables are batch and temperature (ºF) at which all the gasoline is vaporized. The batch variable is a categorical variable denoting 10 different conditions involved in the experiment.

Since the model here is estimated considering the intercept, then one dummy variable is omitted. The estimated model, under logit link function, is given by

$$\log\left(\frac{\tau_i}{1-\tau_i}\right) = \beta_1 + \sum_{m=2}^{11}\beta_m x_{im}, \quad i = 1, \ldots, 32,$$

where $(x_{i2}, \ldots, x_{i10})$ denotes the nine dummy variables and x_{i11} is the variable temperature.

All calculations were done using the BFGS method in the Ox Console programming language (Doornik, 2018) with analytical derivatives. Table 3 presents the summary of the MLEs. As can be

Table 2: Simulation results for Modelo II.

n	Par	AE	MSE	CR		
				90%	95%	99%
30	β_1	0.68181	0.01860	86.04	92.10	97.72
	β_2	1.51145	0.02403	85.66	91.44	97.18
	β_3	−0.81530	0.01997	85.93	91.69	97.19
	β_4	−2.41789	0.07631	85.49	91.59	97.13
	ϕ	3.69455	0.48738	84.02	91.31	98.33
60	β_1	0.69114	0.00890	87.92	93.53	98.19
	β_2	1.50769	0.01277	87.59	93.01	98.12
	β_3	−0.80658	0.00989	88.18	93.76	98.29
	β_4	−2.41009	0.03395	87.86	93.04	98.18
	ϕ	3.48909	0.16359	86.58	92.99	98.64
100	β_1	0.69606	0.00514	88.28	94.09	98.54
	β_2	1.50616	0.00788	88.78	94.09	98.51
	β_3	−0.80263	0.00593	88.76	93.95	98.44
	β_4	−2.40804	0.01809	88.52	94.17	98.47
	ϕ	3.41336	0.08068	88.08	93.59	98.91
200	β_1	0.69765	0.00255	89.45	94.41	98.66
	β_2	1.50220	0.00412	89.29	94.73	98.81
	β_3	−0.80219	0.00296	89.38	94.38	98.91
	β_4	−2.40586	0.00797	89.82	94.57	98.67
	ϕ	3.35809	0.03544	88.89	94.24	98.99
300	β_1	0.69842	0.00169	89.65	94.49	98.59
	β_2	1.50184	0.00265	89.71	94.86	98.94
	β_3	−0.80149	0.00177	88.90	94.41	98.86
	β_4	−2.40243	0.00522	89.44	94.55	98.75
	ϕ	3.34024	0.02274	88.79	94.59	98.85

seen, for all four models, all estimated coefficients are highly significant. The information criteria, shown in Table 4, indicate that the UIW model is the best model, since this model has the lowest AIC, BIC and HQIC values.

Figure 4 shows the plots of the Dunn-Smith residuals and of the simulated envelope. This figure reveals that the residuals present good behavior, indicating that the model is well estimated.

7 Conclusions

A new bi-parametric distribution is introduced in this paper. The new distribution has support in the interval (0,1) and is obtained from a transformation of the inverse Weibull distribution. The failure rate function of this new distribution can be increasing and bathtub-shaped.

Subsequently, based on this distribution, a regression model is proposed. This model has a regression structure at the median. The unknown parameters are obtained by the maximum likelihood method. Analytical expressions for the score vector and for the Fisher observed information matrix are demonstrated.

Monte Carlo simulation studies demonstrate the good consistency of the maximum likelihood estimators. Finally, an application to real Petroleum data is performed, showing that the proposed regression model is better than three others known regression models.

Table 3: Summary of the MLEs.

Par	Estimate	Std. error	p-value	Estimate	Std. error	p-value
	UIW			unit Weibull		
β_1	−5.99149	0.17483	0.00000	−6.42857	0.21942	0.00000
β_2	1.67859	0.09041	0.00000	1.66828	0.09478	0.00000
β_3	1.17744	0.08981	0.00000	1.45079	0.11858	0.00000
β_4	1.46966	0.08865	0.00000	1.65779	0.12960	0.00000
β_5	0.94507	0.07936	0.00000	1.11451	0.09897	0.00000
β_6	1.03651	0.07763	0.00000	1.21315	0.10626	0.00000
β_7	0.92085	0.08158	0.00000	1.10537	0.10354	0.00000
β_8	0.46318	0.07898	0.00000	0.53008	0.12999	0.00005
β_9	0.42795	0.08204	0.00000	0.46854	0.12087	0.00011
β_{10}	0.29092	0.09352	0.00187	0.41987	0.11747	0.00035
β_{11}	0.01077	0.00042	0.00000	0.01157	0.00051	0.00000
ϕ	21.79434	3.30620	0.00000	16.76020	2.35583	0.00000
	Kumaraswamy			beta		
β_1	−6.27295	0.15239	0.00000	−6.15957	0.18232	0.00000
β_2	1.82388	0.09804	0.00000	1.72773	0.10123	0.00000
β_3	1.24510	0.09688	0.00000	1.32260	0.11790	0.00000
β_4	1.51274	0.09913	0.00000	1.57231	0.11610	0.00000
β_5	0.98056	0.08810	0.00000	1.05971	0.10236	0.00000
β_6	0.98650	0.09343	0.00000	1.13375	0.10352	0.00000
β_7	0.93776	0.09311	0.00000	1.04016	0.10604	0.00000
β_8	0.40937	0.08569	0.00000	0.54369	0.10913	0.00000
β_9	0.39060	0.08951	0.00001	0.49590	0.10893	0.00001
β_{10}	0.31594	0.10030	0.00163	0.38579	0.11859	0.00114
β_{11}	0.01153	0.00036	0.00000	0.01097	0.00041	0.00000
ϕ	11.02289	1.38850	0.00000	440.27838	110.02562	0.00006

Table 4: Information criteria.

Model	AIC	BIC	HQIC
UIW	**−150.0440**	**−132.4551**	**−144.2138**
unit Weibull	−138.8193	−121.2305	−132.9891
Kumaraswamy	−143.2682	−125.6793	−137.4380
beta	−145.5951	−128.0063	−139.7649

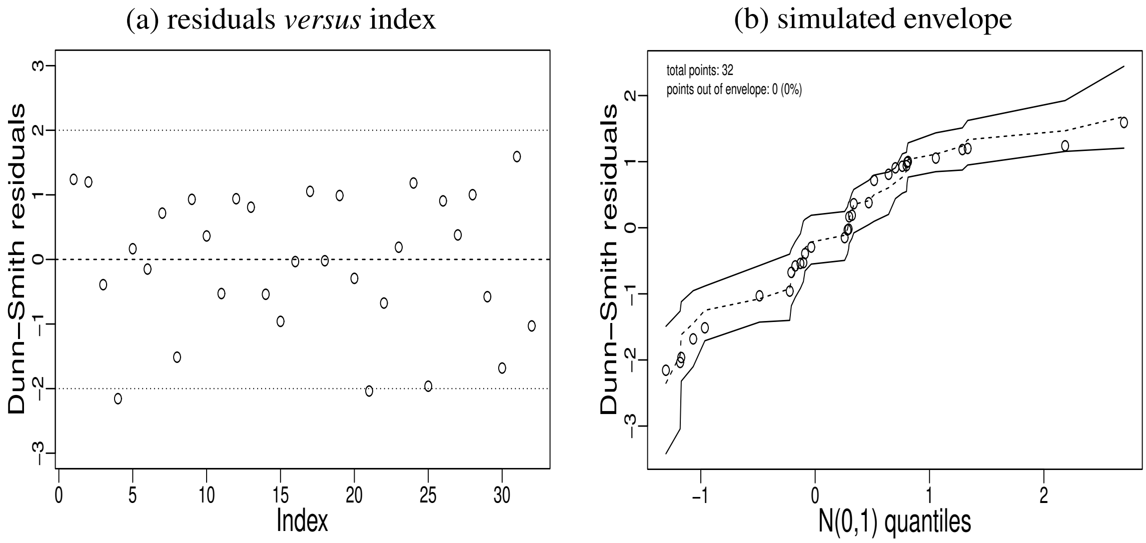

Figure 4: Dunn-Smith residuals.

References

Altun, E., El-Morshedy, M., and Eliwa, M. S. (2021). A new regression model for bounded response variable: An alternative to the beta and unit-Lindley regression models. *PLOS ONE*, 16(1):1–15.

Altun, E. and Hamedani, G. G. (2018). The log-xgamma distribution with inference and application. *Journal de la Société Française de Statistique*, 159(3):40–55.

Atkinson, A. C. (1985). *Plots, Transformations and Regression; an Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford University Press, New York.

Doornik, J. A. (2018). *Ox: an object-oriented matrix programming language*. Timberlake Consultants and Oxford, London, United Kingdown.

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244.

Ferrari, S. L. P. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815.

Gómez-Déniz, E., Sordo, M. A., and Calderín-Ojeda, E. (2014). The log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insurance: Mathematics and Economics*, 54:49–57.

Mazucheli, J., Menezes, A. F. B., and Dey, S. (2019). Unit-Gompertz distribution with applications. *Statistica*, 79(1):25–43.

Mazucheli, J., Menezes, A. F. B., Fernandes, L. B., de Oliveira, R. P., and Ghitany, M. E. (2020). The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. *Journal of Applied Statistics*, 47(6):954–974.

Mazucheli, J., Menezes, A. F. B., and Ghitany, M. E. (2018). The unit-Weibull distribution and associated inference. *Journal of Applied Probability and Statistics*, 13(2):1–22.

Mitnik, P. A. and Baek, S. (2013). The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54(1):177–192.

Prater, N. H. (1956). Estimate gasoline yields from crudes. *Petroleum Refiner*, 35(5):236–238.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ribeiro-Reis, L. D. (2021). Unit log-logistic distribution and unit log-logistic regression model. *Journal of the Indian Society for Probability and Statistics*, 22(2):375–388.