Unit-Power Half-Normal Distribution Including Quantile Regression with Applications to Medical Data

Karol I. Santoro [1], Yolanda M. Gómez [2,*], Darlin Soto [2] and Inmaculada Barranco-Chamorro [3]

[1] Departamento de Estadística y Ciencia de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; karol.santoro@uantof.cl
[2] Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile; dmsoto@ubiobio.cl
[3] Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla, 41012 Sevilla, Spain; chamorro@us.es
* Correspondence: ygomez@ubiobio.cl

Abstract: In this paper, we present the unit-power half-normal distribution, derived from the power half-normal distribution, for data analysis in the open unit interval. The statistical properties of the unit-power half-normal model are described in detail. Simulation studies are carried out to evaluate the performance of the parameter estimators. Additionally, we implement the quantile regression for this model, which is applied to two real healthcare data sets. Our findings suggest that the unit power half-normal distribution provides a robust and flexible alternative for existing models for proportion data.

Keywords: power half-normal distribution; proportion data; maximum likelihood estimation; regression model

MSC: 60E05; 62E15; 62F10

1. Introduction

To study and interpret real events, new statistical models must continuously be built. In recent times, models limited to the interval $(0, 1)$ have generated a lot of interest. This type of models is mainly used to study proportion data, such as survival rate data and product failure, among others. Different areas, such as health, actuarial and financial sciences, require this type of distributions. So, recently, several distributions with positive support have been transformed to distributions with unit support, such as the cases of Jones [1] in the Kumaraswamy distribution, Gómez-Déniz et al. [2] in the log-Lindley distribution, Mazucheli et al. [3], in the Lindley distribution, Abd El-Monsef et al. [4], who proposed a new two-parameter omega unitary distribution, Altun et al. [5], who studied a distribution called enhanced second-order Lindley distribution modeling data on the interval (0,1), and, recently, Ahmad et al. [6], who studied the exponential pareto distribution.

One of the distributions that is mainly used for this type of data is the Beta distribution. Interest in this type of model has been increasing, and many researchers have transformed known distributions with positive support into distributions with unitary support, an example of which can be found in articles such as Grassia [7], based on the Gamma distribution, Ghitany et al. [8], based on the Inverse Gamma distribution, Mazucheli et al. [9], based on the Birnbaum-Saunders distribution, Modi et al. [10], based on Burr distribution III, Korkmaz and Chesneau [11], based on Burr distribution XII, Haq et al. [12], based on the modified Burr III distribution, and, more recently, Bakouch et al. [13], based on half-normal distribution.

The aim of this paper is to introduce a new distribution with a bounded domain on (0,1). This distribution is originated by modifying the representation of the power half-normal (PHN) distribution proposed by Gómez et al. [14]. One of the motivations of this

work is to generate alternatives to well-known distributions used in the statistical analysis of certain type of data. This work is based on the PHN distribution whose probability density function f (pdf), cumulative distribution function F (cdf), and quantile function Q are:

$$f_Y(y;\sigma,\alpha) = \frac{2\alpha}{\sigma}\phi\left(\frac{y}{\sigma}\right)\left(2\Phi\left(\frac{y}{\sigma}-1\right)\right)^{\alpha-1}, \quad y > 0, \tag{1}$$

$$F_Y(y;\sigma,\alpha) = \left(2\Phi\left(\frac{y}{\sigma}\right)-1\right)^{\alpha}, \quad y > 0, \tag{2}$$

$$Q(p;\sigma,\alpha) = \sigma\Phi^{-1}\left(\frac{1+p^{1/\alpha}}{2}\right), \quad 0 < p < 1, \tag{3}$$

where the parameters are $\sigma > 0$ and $\alpha > 0$. $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of the standard normal distribution, respectively.

At present, this model continues to be studied. For example, Barrios et al. [15] performed an extension of the distribution using the slash process, resulting in a model with higher kurtosis, i.e., with heavier tails. In addition, Pallini [16] introduced a discretized model based on the PHN distribution. The interest in further investigating models with support on the interval (0,1) is of great importance, since in many areas, such as finance, actuarial sciences, engineering, and health, databases that belong to this type of interval are handled. To consider a new model with unit support based on the PHN model is a promising alternative for more accurate inferential studies.

The rest of the paper is organized as follows. In Section 2, we introduce our proposal, the unit-power half-normal (UPHN) distribution. Several important properties of this new model are presented. In Section 3, inference is performed, and maximum likelihood (ML) estimators are obtained. In Section 4, the reparametrized model in terms of a quantile is presented. In Section 5, a simulation study is carried out to analyse the performance of ML estimators in finite samples for the proposed model without and with covariates. In Section 6, two real data applications are presented, and we are again dealing with cases without and with covariates. Finally, in Section 7, some concluding comments are presented.

2. Unit Power Half-Normal Distribution

In this section, we will discuss the stochastic representation of the UPHN model, including its pdf, cdf, and some properties of the model.

2.1. Stochastic Representation

Definition 1. *A random variable X follows a UPHN distribution with parameters $\sigma > 0$ and $\alpha > 0$ if its stochastic representation is given by:*

$$X = \frac{1}{1+Y} \tag{4}$$

where $Y \sim PHN(\sigma,\alpha)$. The notation $X \sim UPHN(\sigma,\alpha)$ will be used.

2.2. Pdf, Cdf, Survival, and Hazard Functions

Proposition 1. *Let $X \sim UPHN(\sigma,\alpha)$. Then, the pdf of X is given by:*

$$f_X(x;\sigma,\alpha) = \frac{2\alpha}{\sigma x^2}\phi\left(\frac{1-x}{\sigma x}\right)\left(2\Phi\left(\frac{1-x}{\sigma x}\right)-1\right)^{\alpha-1}, \quad 0 < x < 1, \quad \sigma,\alpha > 0. \tag{5}$$

Proof. By considering the stochastic representation given in (4), $x = g(y) = 1/(1+y)$, we have that

$$f_X(x;\sigma,\alpha) = f_Y(g^{-1}(x);\sigma,\alpha)\left|\frac{dg^{-1}(x)}{dx}\right| = \frac{2\alpha}{\sigma}\phi\left(\frac{g^{-1}(x)}{\sigma}\right)\left(2\Phi\left(\frac{g^{-1}(x)}{\sigma}\right)-1\right)^{\alpha-1}\frac{1}{x^2}.$$

Taking into account that $g^{-1}(x) = (1-x)/x$, (5) follows. □

Proposition 2. *Let $X \sim UPHN(\sigma, \alpha)$. Then, its cdf is given by:*

$$F(x; \sigma, \alpha) = \begin{cases} 0, & x \leq 0, \\ 1 - \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha}, & 0 < x < 1, \\ 1, & x \geq 1. \end{cases} \quad (6)$$

Proof.

- If $x \leq 0$, then $f(x) = 0$. Therefore, $F(x, \sigma, \alpha) = 0$.
- It follows from the fact that for $0 < x < 1$,

$$F(x; \sigma, \alpha) = \int_0^x \frac{2\alpha}{\sigma t^2} \phi\left(\frac{1-t}{\sigma t}\right) \left(2\Phi\left(\frac{1-t}{\sigma t}\right) - 1\right)^{\alpha-1} dt .$$

Making $u = 2\Phi\left(\frac{1-t}{\sigma t}\right) - 1$, we obtain the result.

- If $x \geq 1$, then:

$$\begin{aligned} F(x; \sigma, \alpha) &= \int_{-\infty}^x f(t) dt \\ &= \int_{-\infty}^0 0 \, dt + \int_0^1 \frac{2\alpha}{\sigma t^2} \phi\left(\frac{1-t}{\sigma t}\right) \left(2\Phi\left(\frac{1-t}{\sigma t}\right) - 1\right)^{\alpha-1} dt + \int_1^x 0 \, dt = 1. \end{aligned}$$

□

Figure 1 shows the pdf and cdf of the UPHN(σ, α) model of $\sigma = 1$ with different values for α. Figure 2 shows the pdf and cdf of the UPHN(σ, α) model, for different values of (σ, α).
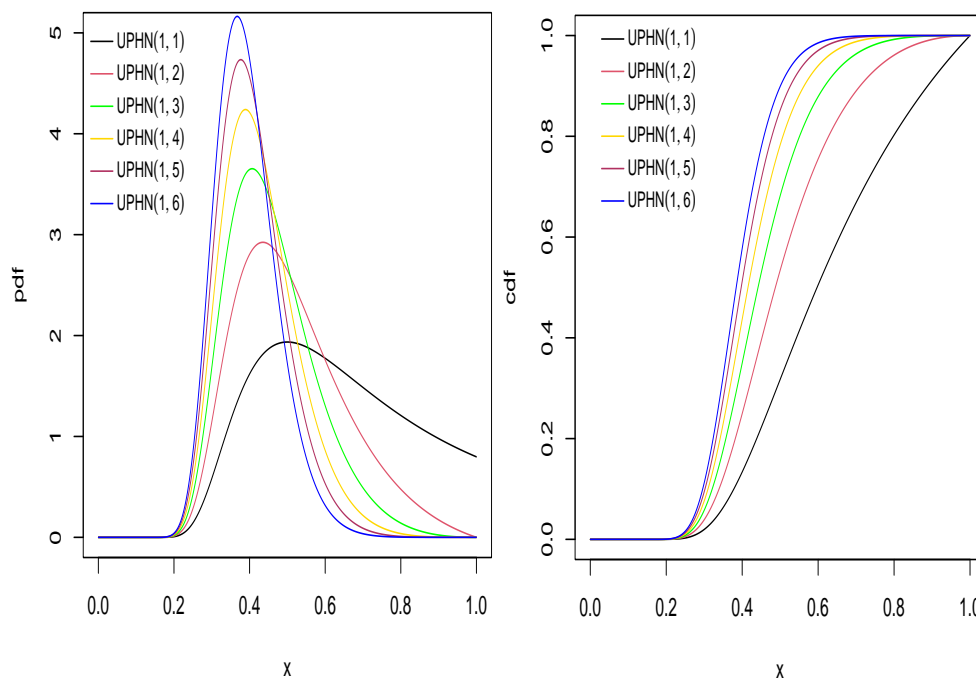
Figure 1. Pdf and cdf for UPHN$(\sigma = 1, \alpha)$ model with different values of α.
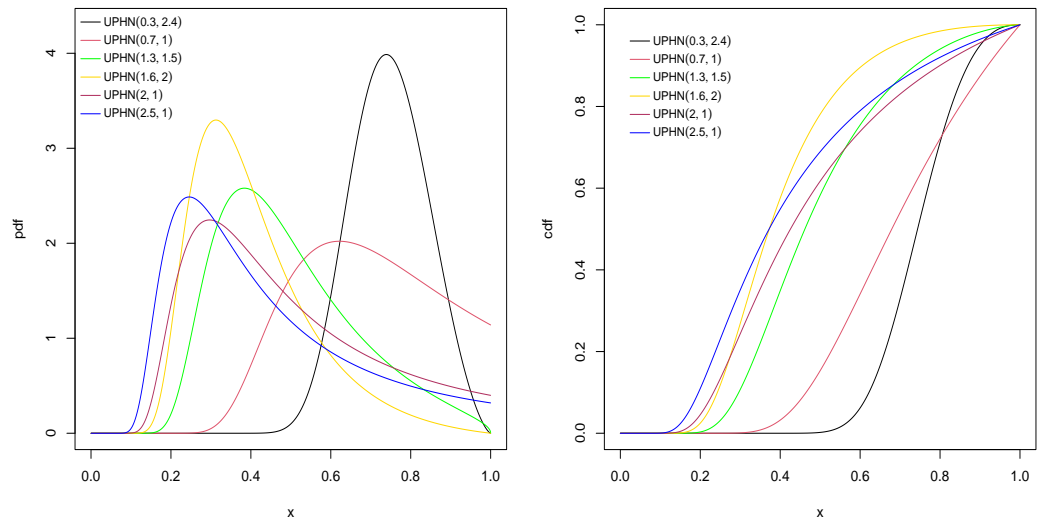
Figure 2. Pdf and cdf for UPHN(σ, α) model for some values of parameters σ and α.

Proposition 3. *Let* $X \sim UPHN(\sigma, \alpha)$. *Then, the survival and hazard function of X are given by:*

$$S(x; \sigma, \alpha) = \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha} \quad and \quad h(x; \sigma, \alpha) = \frac{2\alpha}{\sigma x^2}\phi\left(\frac{1-x}{\sigma x}\right)\left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{-1},$$

for $0 < x < 1$.

Proof. Both functions are immediately obtained from their definitions, since $S(x) = 1 - F(x)$ and $h(x) = \frac{f(x)}{1-F(x)}$. $\square$

Figure 3 shows the survival and the hazard function for the UPHN(σ, α) model, considering several values of σ and α. We highlight that the hazard function is increasing.
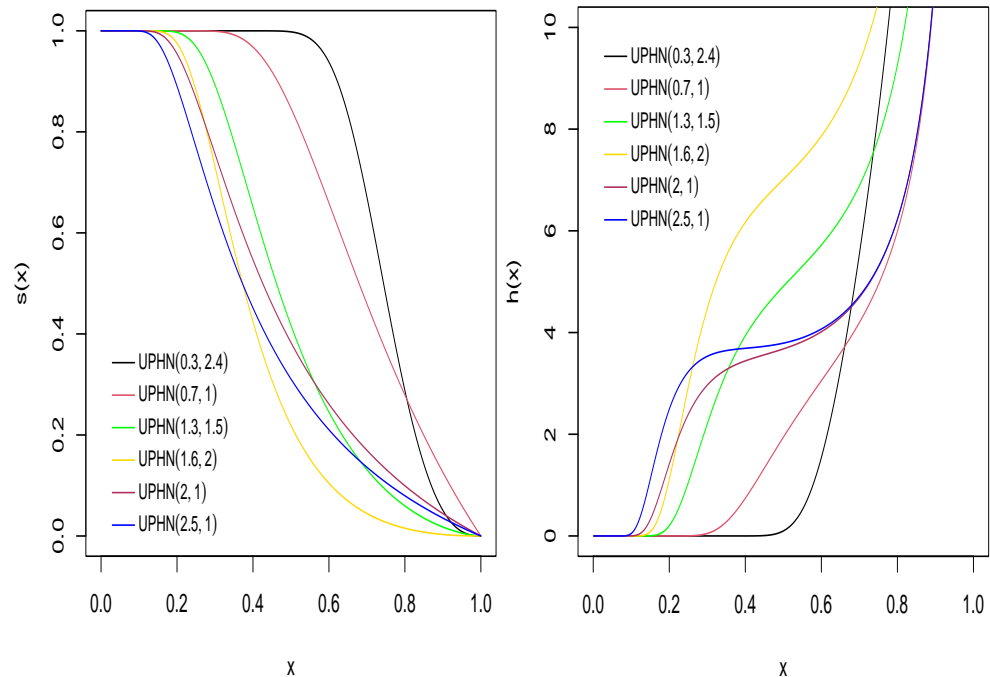
Figure 3. Survival and hazard functions for UPHN(σ, α) model for some values of parameters σ and α.

2.3. Moments

Proposition 4. *Let* $X \sim UPHN(\sigma, \alpha)$. *Then, the n-th moment of X, where n is a positive integer, is given by:*

$$\mu_X = \alpha b_n(\sigma, \alpha), \tag{7}$$

where $b_n(\sigma, \alpha) = \int_0^1 \left(\sigma \Phi^{-1} \left(\frac{u+1}{2} \right) + 1 \right)^{-n} u^{\alpha-1} du$, *are computed numerically.*

Proof. The moments are immediately obtained from their definition.

$$E[X^n] = \frac{2\alpha}{\sigma} \int_0^1 x^{n-2} \phi \left(\frac{1-x}{\sigma x} \right) \left(2\Phi \left(\frac{1-x}{\sigma x} \right) - 1 \right)^{\alpha-1} dx. \tag{8}$$

Making the change of variable $u = 2\Phi \left(\frac{1-x}{\sigma x} \right) - 1$, we have that:

$$E[X^n] = \alpha \int_0^1 \left(\sigma \Phi^{-1} \left(\frac{u+1}{2} \right) + 1 \right)^{-n} u^{\alpha+1} du,$$

and defining $b_n(\sigma, \alpha)$ as the integral, the result is obtained. $\square$

Corollary 1. *Let* $X \sim UPHN(\sigma, \alpha)$. *Then, the skewness* $\left(\sqrt{\beta_1} \right)$ *and kurtosis* (β_2) *coefficients are*

$$\sqrt{\beta_1} = \frac{b_3 - 3b_1 b_2 + 2b_1^3}{(b_2 - b_1^2)^{3/2}} \quad and \quad \beta_2 = \frac{b_4 - 4b_1 b_3 + 6b_1^2 b_2 - 3b_1^4}{(b_2 - b_1^2)^2},$$

where $b_n = b_n(\sigma, \alpha)$.

Proof. It is immediately obtained from the definition of these coefficients. $\square$

Tables 1 and 2 provide us $\left(\sqrt{\beta_1} \right)$ and (β_2) for several values of parameters σ and α. We can observe that for low values of σ and α, the skewness coefficient is negative and for high values, positive skewness is obtained. As for the kurtosis coefficient, the observed effect is more dispersed.

Table 1. Skewness coefficient of UPHN(σ, α) model for different values of σ and α.

	α				
σ	0.3	0.7	1.5	3	5
0.1	−1.7390	−0.9557	−0.5196	−0.3301	−0.2892
0.6	−1.1043	−0.3191	0.1764	0.3817	0.3805
1.4	−0.7097	0.1218	0.7115	0.9300	0.8517
2.5	−0.4381	0.4597	1.1669	1.3915	1.2040
5	−0.1228	0.9035	1.8445	2.0545	1.6279
10	0.1805	1.4033	2.7419	2.8675	2.0197

Table 2. Kurtosis coefficient of UPHN(σ, α) model for different values of σ and α.

	α				
σ	0.3	0.7	1.5	3	5
0.1	5.8356	3.4239	2.8539	2.8706	2.9599
0.6	3.1896	2.0788	2.3258	2.8454	3.0395
1.4	2.1877	1.8542	2.9189	4.0065	4.1003
2.5	1.7492	2.0303	4.0774	5.8378	5.5157
5	1.4777	2.6956	6.9298	10.0505	8.0999
10	1.4435	4.0179	12.8648	18.3334	11.6586

2.4. Incomplete Moments

The k-th incomplete moments of the UPHN distribution are given by:

$$
\begin{aligned}
T_k(t) &= E\left(X^k | x < t\right) \\
&= \frac{2\alpha}{\sigma} \int_0^t x^{k-2} \phi\left(\frac{1-x}{\sigma x}\right) \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha-1} dx, \quad k = 0, 1, 2, \ldots
\end{aligned}
$$

This integral cannot be solved analytically. In particular, for $k = 0, 1, 2$, we have:

$$
T_0(t) = F(t; \sigma, \alpha), \quad T_1(t) = \alpha \gamma_1(a(t)), \quad T_2(t) = \alpha \gamma_2(a(t)),
$$

where $\gamma_j(a(t)) = \int_{a(t)}^1 \left(\alpha \Phi^{-1}\left(\frac{u+1}{2}\right) + 1\right)^{-j} u^{\alpha-1} du$ and $a(y) = 2\Phi\left(\frac{1-y}{\sigma y}\right) - 1$. $\gamma_j(a(t))$, are computed numerically.

2.5. The Lorenz Curve and the Gini Index

The standard definition of the Lorenz curve [17] is provided in terms of the first incomplete moment and the expected value of X. Specifically, for the UPHN model, the following closed form expression is obtained

$$
\begin{aligned}
L(p) &= \frac{1}{\mu_X} \int_0^t x f(x) dx = \frac{1}{\mu_X} T_1(t); \\
&= \frac{\alpha}{\mu_X} \gamma_1(a(t)).
\end{aligned}
$$

The Gini index, also known as the Gini coefficient (see [18,19]), is a statistical dispersion metric associated with the Lorenz curve, intended to represent income inequality, wealth inequality, or consumption inequality within a nation or social group. The Gini index is defined as:

$$
G(\sigma, \alpha) = 1 - \frac{1}{\mu_X} \int_0^\infty [1 - T_0(t)]^2 dt.
$$

Proposition 5. *Let $X \sim UPHN(\sigma, \alpha)$. Then, the Gini index is given by:*

$$
G(\sigma, \alpha) = 1 - \frac{4^\alpha}{\mu_X} \int_0^\infty \left[\Phi\left(\frac{1-t}{\sigma t}\right)\right]^{2\alpha} dt.
$$

Proof. By definition, the proof is direct. □

Proposition 6. *Let $X \sim UPHN(\sigma, \alpha)$; the Renyi entropy of order δ for X is given by:*

$$
R_\delta(x) = \frac{1}{1-\delta}\left(\delta \log(\alpha) - \log((\alpha-1)\delta + 1) + \log\left(1 - (1 - 2\Phi(1/\sigma))^{(\alpha-1)\delta+1}\right)\right),
$$

where $\delta > 0$ and $\delta \neq 1$.

Proof. The Renyi entropy is defined as $R_\delta(x) = \frac{1}{1-\delta} \log\left(\int_0^\infty f(x)^\delta dx\right)$. Therefore, we have:

$$
R_\delta(y) = \frac{1}{1-\delta} \log\left(\int_0^\infty \left(\frac{2\alpha}{\sigma x^2} \phi\left(\frac{1-x}{\sigma x}\right) \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha-1}\right)^\delta dx\right),
$$

Making the change of variable $u = 2\Phi\left(\dfrac{1-x}{\sigma x}\right) - 1$, the result is obtained. $\square$

2.6. Shannon Entropy

The entropy of a random variable X is a measure of its uncertainty. The Shannon entropy measure is defined by:

$$\mathbf{J}_S = -E(\log(f(x))).$$

If follows, after extensive algebraic manipulations, that the Shannon entropy for the UPHN model is:

$$\mathbf{J}_S = \log\left(\frac{\sigma\sqrt{2\pi}}{2\alpha}\right) + 2m_1 + \frac{1}{2\sigma^2}\left(\alpha b_2(\sigma,\alpha) - 2\alpha b_1(\sigma,\alpha) + 1\right) + (\alpha+1)m_2.$$

where $m_1 = E(\log(x))$, $m_2 = E\left(\log\left(2\Phi\left(\dfrac{1-x}{\sigma x}\right) - 1\right)\right)$, and with $b_1(\sigma,\alpha)$ and $b_2(\sigma,\alpha)$ as given above.

2.7. Quantiles

Proposition 7. *Let* $X \sim UPHN(\sigma,\alpha)$. *Then, the quantile function is given by:*

$$Q(p;\sigma,\alpha) = \left[\sigma\Phi^{-1}\left(\frac{(1-p)^{1/\alpha}}{2} + \frac{1}{2}\right) + 1\right]^{-1}, \quad 0 < p < 1. \tag{9}$$

Proof. It follows from a direct computation, by applying the definition of quantile function. $\square$

Corollary 2. *The quartiles for the UPHN distribution are:*

1. *(First quartile)* $Q(0.25;\sigma,\alpha) = \left[\sigma\Phi^{-1}\left(\dfrac{0.75^{1/\alpha}}{2} + \dfrac{1}{2}\right) + 1\right]^{-1}.$

2. *(Median)* $Q(0.5;\sigma,\alpha) = \left[\sigma\Phi^{-1}\left(\dfrac{0.5^{1/\alpha}}{2} + \dfrac{1}{2}\right) + 1\right]^{-1}.$

3. *(Third quartile)* $Q(0.75;\sigma,\alpha) = \left[\sigma\Phi^{-1}\left(\dfrac{0.25^{1/\alpha}}{2} + \dfrac{1}{2}\right) + 1\right]^{-1}.$

The quantile function can also be used to study the kurtosis and skewness in the UPHN model. A classical measurement for skewness was introduced by MacGillivray [20], which is given by:

$$\rho_{X_i}(p) = \frac{Q_{X_i}(1-p;\sigma,\alpha) + Q_{X_i}(p;\sigma,\alpha) - 2Q_{X_i}(0.5;\sigma,\alpha)}{Q_{X_i}(1-p;\sigma,\alpha) - Q_{X_i}(p;\sigma,\alpha)}, \quad p \in (0,1). \tag{10}$$

In particular, the MacGuillevray skewness measurement can efficiently describe the effect of the parameters (σ,α) on asymmetry. In Figure 4, the behavior of the asymmetry coefficient is plotted with respect to the values of the parameters.
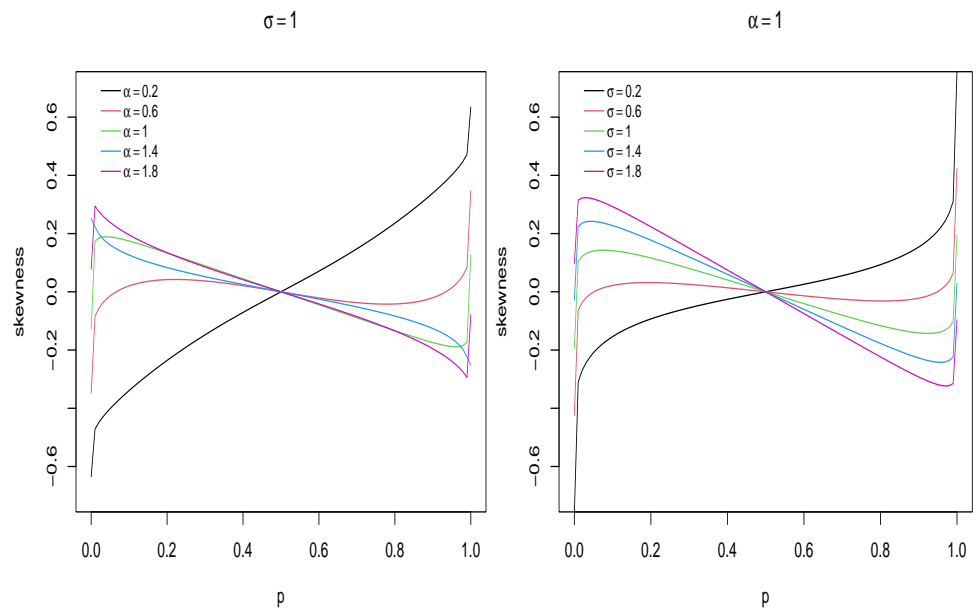
Figure 4. Plots of the MacGillivray skewness coefficient in UPHN model.

The kurtosis of the UPHN(σ, α) distribution can also be studied using the Moors kurtosis coefficient [21], usually given by:

$$K(\sigma, \alpha) = \frac{Q(7/8; \sigma, \alpha) - Q(5/8; \sigma, \alpha) + Q(3/8; \sigma, \alpha) - Q(1/8; \sigma, \alpha)}{Q(3/4; \sigma, \alpha) - Q(1/4; \sigma, \alpha)}. \tag{11}$$

It can be seen in [21] that for large values of (11), the distribution has heavy tails and for small values, the model has lighter tails. Figure 5 shows the behavior of the Moors kurtosis coefficient for the UPHN distribution.
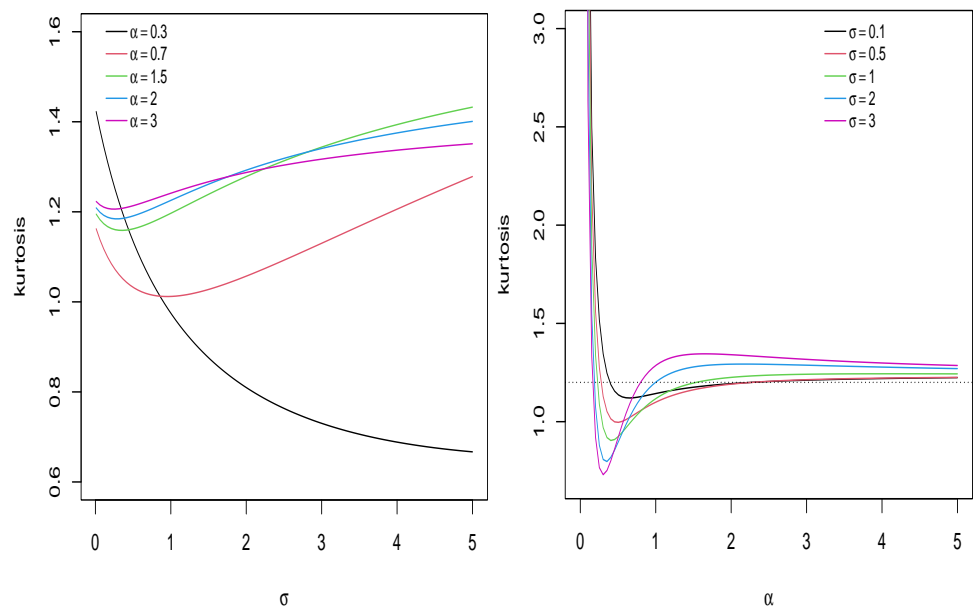
Figure 5. Plots of the Moors kurtosis coefficient for the UPHN model.

2.8. Order Statistics

Order statistics have a wide range of applications in physical and life sciences (see, for instance, Balakrishnan and Cohen [22]). From a statistical perspective, they allow the

computation of useful functions such as the sample range and the sample median. The following result states the pdf of k-th order statistic from a UPHN random sample of size n, which is arranged in a non-decreasing order.

Proposition 8. *Let* $X_1, X_2, \ldots, X_n$ *be independent and identically distributed UPHN(σ, α) random variables. Then, the pdf of the k-th order statistic, $X_{k:n}$, is given by:*

$$f_{X_{k:n}}(x) = \frac{2\alpha n!}{(k-1)!(n-k)!\sigma x^2} \phi\left(\frac{1-x}{\sigma x}\right) \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha(n-k+1)-1} \left[1 - \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha}\right]^{k-1}.$$

Corollary 3. *Let* $X_1, X_2, \ldots, X_n$ *be a random sample from a UPHN(σ, α) distribution. Then,*

1. *The pdf of the minimum, $X_{1:n}$, is*

$$f_{X_{1:n}}(x) = \frac{2\alpha n}{\sigma x^2} \phi\left(\frac{1-x}{\sigma x}\right) \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{n\alpha - 1},$$

2. *The pdf of the maximum, $X_{n:n}$, is*

$$f_{X_{n:n}}(x) = \frac{2\alpha n}{\sigma x^2} \phi\left(\frac{1-x}{\sigma x}\right) \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha - 1} \left[1 - \left(2\Phi\left(\frac{1-x}{\sigma x}\right) - 1\right)^{\alpha}\right]^{n-1}.$$

2.9. Bonferroni Curve

In various disciplines, such as socioeconomics and public health sciences, there is a need to compare and analyze inequality in distributions. Bonferroni curves are tools useful to reach this aim. It is worth mentioning that these curves have applications not only in economics to study income and poverty, but also in medicine or reliability. A detailed discussion can be seen in Bonferroni [23] or Arcagni and Porro [24]. The expressions of these curves for the UPHN model are presented below.

Proposition 9. *Let* $X \sim UPHN(\sigma, \alpha)$. *Then, for $0 < p < 1$, the Bonferroni curve, say $B(p)$, is given by:*

$$B(p) = \frac{\alpha}{\mu_X p} \sum_{j=0}^{\infty} (-1)^j \sigma^j \gamma_{a(q)}^j,$$

where $\mu_X = E(X)$.

Proof. The expression above is obtained by using the definition of the Bonferroni curve, that is,

$$B(p) = \frac{1}{\mu_X p} \int_0^p F^{-1}(t)dt = \frac{1}{\mu_X p} \int_0^q x f(x)dx, \quad 0 < p < 1,$$

where μ_X is the expected value of the corresponding non-negative random variable and $q = F^{-1}(p)$. $\square$

3. Inference

In this section, the inference about the parameters in the UPHN distribution is carried out from a classical point of view. Let us consider $X_1, X_2, \ldots, X_n$ a random sample of $X \sim UPHN(\sigma, \alpha)$. The maximum likelihood (ML) estimation method is discussed below. Given a random sample $x_1, x_2, \ldots, x_n$ of size n from UPHN(σ, α), the log-likelihood function is given by:

$$\ell(\boldsymbol{\theta}) = n \log\left(\frac{2\alpha}{\sigma}\right) - \frac{n}{2} \log(2\pi) - 2 \sum_{i=1}^{n} \log(x_i) - \frac{1}{2} \sum_{i=1}^{n} v_i^2 + (\alpha - 1) \sum_{i=1}^{n} \log(2\Phi(v_i) - 1), \tag{12}$$

where $v_i = \dfrac{1 - x_i}{\sigma x_i}$. Therefore, the score assumes the form $S(\boldsymbol{\theta}) = (S_\sigma(\boldsymbol{\theta}), S_\alpha(\boldsymbol{\theta}))$ where:

$$S_\sigma(\boldsymbol{\theta}) = -\frac{n}{\sigma^2} + \frac{1}{\sigma} \sum_{i=1}^{n} v_i^2 - \frac{2(\alpha - 1)}{\sigma} \sum_{i=1}^{n} \frac{\phi(v_i) v_i}{2\Phi(v_i) - 1}, \tag{13}$$

$$S_\alpha(\boldsymbol{\theta}) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log(2\Phi(v_i) - 1). \tag{14}$$

The ML estimators (MLEs) can be obtained by solving the likelihood Equations (13) and (14). Numerical methods, such as the Newton–Raphson procedure, must be used. Alternative maximization techniques could also be applied, for instance, those proposed in MacDonald [25].

The asymptotic variance of the MLEs, say $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}, \widehat{\alpha})$, can be estimated by the Fisher information matrix defined as $\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E}\left[\partial^2 \ell(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\right]$, where $\ell(\boldsymbol{\theta})$ is the log-likelihood function of the UPHN model given in (12). Recall that, under the regularity conditions,

$$\mathcal{I}(\boldsymbol{\theta})^{-1/2} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{D}} N_2(\boldsymbol{0}_2, \boldsymbol{I}_2), \quad \text{as } n \to +\infty,$$

where $\mathcal{D}$ stands for convergence in distribution and $N_2(\boldsymbol{0}_2, \boldsymbol{I}_2)$ denotes the standard bivariate normal distribution. The elements of the matrix $-\partial^2 \ell(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are given by $I_{\sigma\sigma} = -\partial^2 \ell(\boldsymbol{\theta}) / \partial\sigma^2$, $I_{\sigma\alpha} = -\partial^2 \ell(\boldsymbol{\theta}) / \partial\sigma\partial\alpha$ and $I_{\alpha\alpha} = -\partial^2 \ell(\boldsymbol{\theta}) / \partial\alpha^2$. Explicitly, we have:

$$I_{\sigma\sigma} = -\frac{n}{\sigma^2} + \frac{3}{\sigma^2} \sum_{i=1}^{n} v_i^2 - \frac{4(\alpha - 1)}{\sigma^2} \sum_{i=1}^{n} \frac{\phi(v_i) v_i}{2\Phi(v_i) - 1} + \frac{2(\alpha - 1)}{\sigma^2} \sum_{i=1}^{n} \frac{v_i^2 \phi(v_i)[v_i(2\Phi(v_i) - 1) + 2\phi(v_i)]}{(2\phi(v_i) - 1)^2},$$

$$I_{\sigma\alpha} = -\frac{2}{\sigma^2} \sum_{i=1}^{n} \frac{v_i \phi(v_i)}{2\Phi(v_i) - 1},$$

$$I_{\alpha\alpha} = \frac{n}{\alpha^2}.$$

In practice, it is not possible to obtain in a closed form the expected value of previous expressions. However, the covariance matrix of the MLEs, $\mathcal{I}(\boldsymbol{\theta})^{-1}$, can be estimated consistently by $I(\widehat{\boldsymbol{\theta}})^{-1}$, where $I(\widehat{\boldsymbol{\theta}})$ denotes the observed information matrix, which is obtained as

$$I(\widehat{\boldsymbol{\theta}}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

The asymptotic variances of $\widehat{\sigma}$ and $\widehat{\alpha}$ are estimated by the diagonal elements of $I(\widehat{\boldsymbol{\theta}})^{-1}$, and their standard errors by the square root of asymptotic variances.

4. Quantile Regression

We now introduce a regression study, in which a quantile regression model is proposed to describe the conditional quantile of the response variable. Given the quantile function of the *UPHN* distribution, the pdf of the UPHN distribution can be reparameterized in terms of its pth quantile, denoted as $\rho = Q(p; \sigma, \alpha)$.

Let $\sigma = \dfrac{1 - \rho}{\rho \Phi^{-1}\left(\frac{1}{2}((1 - p)^{1/\alpha} + 1)\right)}$. Then, the reparameterized pdf for the UPHN model is:

$$f_X(x; \rho, \alpha, p) = \frac{2\alpha}{(1 - \alpha)x^2} h(\alpha, p) \phi\left(\frac{(1 - x)\rho}{(1 - \rho)x} h(\alpha, p)\right) \left(2\Phi\left(\frac{(1 - x)\rho}{(1 - \rho)x} h(\alpha, p)\right) - 1\right)^{\alpha - 1}, \tag{15}$$

where $h(\alpha, p) = \Phi^{-1}\left(\frac{1}{2}\left((1 - p)^{1/\alpha} + 1\right)\right)$, $0 < x < 1$, $0 < p < 1$, $0 < \rho < 1$, and $\alpha > 0$.

The cdf of the reparameterized UPHN distribution is:

$$F(x; \rho, \alpha, p) = 1 - \left[2\Phi\left(\frac{(1-x)\rho\Phi^{-1}\left(\frac{1}{2}[(1-p)^{1/\alpha}+1]\right)}{(1-\rho)x} \right) - 1 \right]^{\alpha}.$$

4.1. Model Formulation

Let us suppose that $0 < p < 1$ is fixed and we have a random sample $x_i(p) \sim \text{UPHN}(\rho_i, \alpha)$, $i = 1, 2, \ldots, n$. Here, the parameters ρ_i and α are unknown. As a result, the suggested quantile regression model can be defined as:

$$g(\rho_i(p)) = z_i^\top \boldsymbol{\beta}(\boldsymbol{p}), \tag{16}$$

where $\boldsymbol{\beta}^\top(p) = (\beta_0(p), \beta_1(p), \ldots, \beta_k(p))$ and $z_i^\top = (1, z_{i1}, \ldots, z_{ik})$ act as the regression coefficient vector and the ith vector of covariate values, respectively. It is important to note that g is the link function that associates the covariates with the response variable. We can obtain a median quantile regression for a unit response variable when $p = 0.5$. Taking into account that the UPHN distribution is a probability model with support on $(0, 1)$, we use and define the logit-link function as:

$$g(\rho_i(p)) = \log\left(\frac{\rho_i(p)}{1 - \rho_i(p)} \right), \quad i = 1, \ldots, n. \tag{17}$$

4.2. ML Method for the Regression Coefficients

Now, our objective is to estimate the unknown parameters of the UPHN quantile regression model using the ML technique. For this purpose, the logit link function is considered:

$$g(\rho_i(p)) = \log\left(\frac{\rho_i(p)}{1 - \rho_i(p)} \right) = z_i^\top \boldsymbol{\beta}. \tag{18}$$

From the above formula, the inverse link function is:

$$\rho_i(p) = \frac{\exp(z_i^\top \boldsymbol{\beta}(p))}{1 + \exp(z_i^\top \boldsymbol{\beta}(p))}. \tag{19}$$

Let us now have a random sample $X_1, \ldots, X_n$ with $X_i \sim \text{UPHN}(\rho_i(p), \alpha)$, and observed values $x_1, \ldots, x_n$. Then, using the pdf introduced in (15), the associated log-likelihood function is:

$$\begin{aligned} l(\boldsymbol{\Theta}(p)) &= n[\log(2) + \log(\alpha)] + n\log[h(\alpha, p)] + \sum_{i=1}^{n} \log\left[\phi\left(\frac{[1-x_i]\rho_i}{[1-\rho_i]x_i} h(\alpha, p) \right) \right] \\ &+ \sum_{i=1}^{n} \log\left(\frac{\rho_i}{[1-\rho_i]x_i^2} \right) + (\alpha - 1) \sum_{i=1}^{n} \log\left[2\Phi\left(\frac{[1-x_i]\rho_i}{[1-\rho_i]x_i} h(\alpha, p) \right) - 1 \right], \end{aligned} \tag{20}$$

where $\boldsymbol{\Theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$ denotes the unknown vector of parameters. The ML estimators of $\boldsymbol{\Theta}$, denoted as $\widehat{\boldsymbol{\Theta}} = (\widehat{\alpha}, \widehat{\boldsymbol{\beta}}^\top)^\top$ is achieved by maximizing $l(\boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta}$. The `optim` function of the R software, 4.3.1 version, [26] can be used to maximize $l(\boldsymbol{\Theta})$.

4.3. Model Checking

Once we have fitted the model, it is crucial to assess whether the regression model is appropriate for the data. In this sense, the analysis of residuals is key to check or validate the fitted model. We focus on the Cox–Snell residuals (Cox and Snell 1968) [27], which are calculated for the *i*th observation as:

$$\widehat{\eta}_i = -\log[1 - F(x_i; \rho_i(p), \alpha)], \quad i = 1, \ldots, n.$$

Here, $F(x_i, \rho_i(p), \alpha)$ stands for the estimated cdf of the UPHN distribution reparameterized in terms of quantiles. If the fitted model provides a good fit for the data, the residual $\widehat{\eta}_i$ will be an observation of a random variable from an exponential distribution with a scale parameter of one.

5. Simulation

In this section, a simulation study is carried out to evaluate the performance of the ML estimators. We used R software 4.3.1 version, for our calculations, developing customized code that integrates the `optim` function from R package `stats` [28].

5.1. Simulation 1

Next, an algorithm to generate samples from the UPHN(σ, α) is provided. Algorithm 1 is based on the inverse transformation method, where the inverse of the cdf is used.

Algorithm 1 Simulating values from the UPHN(σ, α) distribution.

1: Simulate $U \sim U(0, 1)$.
2: Calculate $X = \left[\sigma \Phi^{-1}\left(\frac{(1-U)^{1/\alpha}}{2} + \frac{1}{2}\right) + 1\right]^{-1}$

As parameter values in our simulation, we consider $\sigma \in \{1, 2\}$ and $\alpha \in \{0.8, 1, 2, 3, 5\}$. For the sample size, we consider $n \in \{150, 300, 600\}$. For each sample size and every combination of σ, α, we perform 1000 repetitions and the corresponding ML estimates are calculated. The results are given in Table 3. As summaries, we provide the estimated bias (bias) for the ML estimators of σ and α, standard errors (SE), the root of the estimated mean square error (RMSE), and the empirical coverage probability (CP) for the asymptotic intervals based on MLEs to 95%. For the ML estimators of σ and α, it should be noted that, as the sample size increases, the bias, SE, and RMSE decrease. It should also be noted that as the sample size increases, the SE and RMSE are closer, which suggests that the standard errors of the ML estimators are well estimated. As for the CPs, we highlight that they are close to the nominal level 0.95.

Table 3. Estimated bias, SE, and RMSE for ML estimators in finite samples from the *UPHN* model.

True Value			$n = 150$				$n = 300$				$n = 600$			
σ	α	Estimator	Bias	se	RMSE	CP	bias	se	RMSE	CP	Bias	se	RMSE	CP
1	0.8	σ	−0.0075	0.0734	0.0747	0.930	−0.0036	0.0522	0.0520	0.948	−0.0013	0.0370	0.0364	0.947
		α	0.0160	0.0774	0.0793	0.953	0.0084	0.0541	0.0549	0.955	0.0037	0.0380	0.0375	0.954
	1	σ	−0.0084	0.0680	0.0690	0.932	−0.0048	0.0482	0.0479	0.944	−0.0012	0.0343	0.0335	0.956
		α	0.0181	0.0993	0.1028	0.946	0.0127	0.0697	0.0724	0.946	0.0017	0.0487	0.0490	0.949
	2	σ	−0.0027	0.0558	0.0554	0.944	0.0012	0.0397	0.0403	0.948	0.0001	0.0280	0.0293	0.933
		α	0.0441	0.2229	0.2312	0.953	0.0083	0.1542	0.1555	0.951	0.0080	0.1090	0.1135	0.944
	3	σ	−0.0038	0.0507	0.0506	0.951	−0.0022	0.0359	0.0360	0.940	−0.0012	0.0254	0.0263	0.931
		α	0.0718	0.3630	0.3601	0.971	0.0230	0.2514	0.2451	0.953	0.0211	0.1776	0.1811	0.942
	5	σ	−0.0050	0.0459	0.0463	0.945	−0.0020	0.0326	0.0334	0.938	−0.0004	0.0231	0.0240	0.937
		α	0.1446	0.6799	0.7206	0.958	0.0678	0.4715	0.4976	0.948	0.0285	0.3300	0.3399	0.948

Table 3. *Cont.*

True Value			$n = 150$				$n = 300$				$n = 600$			
σ	α	Estimator	Bias	se	RMSE	CP	bias	se	RMSE	CP	Bias	se	RMSE	CP
2	0.8	σ	−0.0089	0.1474	0.1461	0.954	−0.0080	0.1044	0.1044	0.948	−0.0029	0.0741	0.0748	0.941
		α	0.0124	0.0770	0.0773	0.952	0.0060	0.0539	0.0553	0.952	0.0020	0.0379	0.0367	0.956
	1	σ	−0.0194	0.1358	0.1320	0.949	−0.0061	0.0967	0.0989	0.946	−0.0011	0.0687	0.0689	0.944
		α	0.0169	0.0991	0.1021	0.953	0.0109	0.0696	0.0733	0.947	0.0014	0.0487	0.0491	0.948
	2	σ	−0.0050	0.1117	0.1097	0.946	−0.0049	0.0789	0.0771	0.953	−0.0022	0.0559	0.0570	0.942
		α	0.0335	0.2215	0.2244	0.957	0.0218	0.1554	0.1548	0.951	0.0092	0.1091	0.1121	0.949
	3	σ	−0.0125	0.1009	0.0997	0.946	−0.0015	0.0719	0.0760	0.934	0.0007	0.0509	0.0511	0.942
		α	0.0789	0.3638	0.3705	0.955	0.0238	0.2514	0.2534	0.955	0.0098	0.1767	0.1741	0.949
	5	σ	−0.0059	0.0921	0.0929	0.949	−0.0022	0.0652	0.0667	0.942	−0.0002	0.0462	0.0467	0.945
		α	0.1166	0.6755	0.7079	0.947	0.0580	0.4702	0.4761	0.958	0.0204	0.3293	0.3392	0.943

5.2. Simulation 2

In this section, data are generated from the quantile regression model with sample sizes $n \in \{100, 200, 300, 400\}$ and quantile levels $p \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ under two different scenarios. In the first scenario, three covariates are considered, that is,

$$logit(\rho_i(p)) = \beta_0(p) + \beta_1(p)z_{i1} + \beta_2(p)z_{i2} + \beta_3(p)z_{i3}, \quad i = 1, 2, \ldots, n, \tag{21}$$

where $\beta_0(p) = 0.275, \beta_1(p) = 0.098, \beta_2(p) = 0.015, \beta_3(p) = 0.397, z_{i1}, z_{i2}, z_{i3} \overset{\text{iid}}{\sim} N(0, 1)$, and $\alpha = 1.285$, whereas in the second scenario, five covariates are proposed

$$logit(\rho_i(p)) = \beta_0(p) + \beta_1(p)z_{i1} + \beta_2(p)z_{i2} + \beta_3(p)z_{i3} + \beta_4(p)z_{i4} + \beta_5(p)z_{i5}, \tag{22}$$

where $\beta_0(p) = -2.975, \beta_1(p) = 0.005, \beta_2(p) = 0.092, \beta_3(p) = -0.958, \beta_4(p) = -0.122, \beta_5(p) = -0.289, z_{i1}, z_{i2}, z_{i3}, z_{i4}, z_{i5} \overset{\text{iid}}{\sim} N(0, 1)$, and $\alpha = 7.425$.

For fixed values of n and p, the response variable is generated as

$$x_i = \left[\sigma_i \Phi^{-1} \left(\frac{(1 - u_i)^{1/\alpha}}{2} + \frac{1}{2} \right) + 1 \right]^{-1}, \tag{23}$$

where u_i is an observation generated from a continuous uniform distribution on $(0, 1)$, that is, $u_i \sim U(0, 1)$ and σ_i is calculated as

$$\sigma_i = \frac{1 - \rho_i}{\rho_i \Phi^{-1}\left(\frac{1}{2}[(1 - p)^{1/\alpha} + 1] \right)}. \tag{24}$$

For both scenarios, and fixed (n, p, α), $M = 500$ Monte Carlo replicates were performed. Then, the empirical bias, RMSE, and CP are calculated by

$$\text{Bias}(\widehat{\gamma}) = \frac{1}{M} \sum_{i=1}^{M} (\widehat{\gamma}_i - \gamma),$$

$$\text{RMSE}(\widehat{\gamma}) = \left[\frac{1}{M} \sum_{i=1}^{M} (\widehat{\gamma}_i - \gamma)^2 \right]^{1/2},$$

$$\text{CP}_{95\%} = \frac{1}{M} \sum_{i=1}^{M} 1_{[\widehat{\gamma}_i \pm 1.96 \times \text{SE}(\widehat{\gamma}_i)]}(\gamma),$$

respectively, where $\gamma = \alpha, \beta_j(p), j = 1, \ldots, 5, 1_{[\widehat{\gamma}_i \pm 1.96 \times \text{SE}(\widehat{\gamma}_i)]}(\gamma)$ is the indicator function defined by

$$1_{[\widehat{\gamma}_i \pm 1.96 \times \mathrm{SE}(\widehat{\gamma}_i)]}(\gamma) = \begin{cases} 1, & \gamma \in (\widehat{\gamma}_i \pm 1.96 \times \mathrm{SE}(\widehat{\gamma}_i)) \\ 0, & \gamma \notin (\widehat{\gamma}_i \pm 1.96 \times \mathrm{SE}(\widehat{\gamma}_i)) \end{cases},$$

and $\mathrm{SE}(\widehat{\gamma}_i)$ is the estimated standard error of $\widehat{\gamma}_i$.

These summaries are provided in Tables 4 and 5. In both settings, the good performance of our estimators is observed.

Table 4. Empirical bias, RMSE, and CP$_{95\%}$ for the true values $\beta_0(p) = 0.275$, $\beta_1(p) = 0.098$, $\beta_2(p) = 0.015$, $\beta_3(p) = 0.397$, and $\alpha = 1.285$ with simulated data.

p	n	Bias					RMSE					CP$_{95\%}$				
		$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	α	$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	α	$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	α
	100	0.018	0.000	0.000	−0.005	0.070	0.069	0.072	0.064	0.067	0.184	0.940	0.950	0.956	0.948	0.960
	200	0.011	−0.003	0.001	0.002	0.022	0.048	0.046	0.043	0.043	0.116	0.924	0.932	0.942	0.954	0.952
0.1	300	0.005	−0.002	0.003	−0.001	0.015	0.037	0.035	0.037	0.039	0.096	0.944	0.964	0.948	0.940	0.958
	400	0.005	0.000	0.000	−0.001	0.014	0.034	0.034	0.032	0.034	0.081	0.940	0.950	0.950	0.952	0.958
	100	0.012	−0.003	−0.004	0.000	0.073	0.072	0.060	0.059	0.071	0.190	0.918	0.942	0.936	0.944	0.952
0.25	200	0.006	−0.002	0.000	−0.001	0.026	0.046	0.046	0.044	0.044	0.118	0.938	0.946	0.956	0.948	0.956
	300	0.005	0.001	−0.001	0.002	0.016	0.037	0.040	0.036	0.039	0.090	0.938	0.956	0.954	0.944	0.962
	400	0.001	0.001	−0.001	0.002	0.021	0.032	0.033	0.034	0.031	0.084	0.944	0.942	0.940	0.960	0.956
	100	0.010	−0.006	−0.002	−0.006	0.057	0.079	0.065	0.064	0.075	0.188	0.940	0.944	0.944	0.934	0.950
0.5	200	0.004	−0.001	0.001	−0.003	0.038	0.051	0.048	0.048	0.046	0.118	0.956	0.946	0.930	0.940	0.962
	300	0.002	0.005	0.001	−0.002	0.019	0.046	0.037	0.035	0.036	0.097	0.926	0.956	0.948	0.956	0.952
	400	−0.001	−0.001	0.000	−0.001	0.022	0.035	0.03	0.034	0.031	0.081	0.956	0.960	0.940	0.954	0.954
	100	−0.007	0.001	−0.002	0.002	0.059	0.117	0.068	0.068	0.064	0.183	0.940	0.946	0.942	0.940	0.954
0.75	200	0.001	0.002	−0.001	−0.002	0.027	0.082	0.041	0.044	0.047	0.122	0.950	0.944	0.946	0.938	0.958
	300	−0.004	−0.001	0.000	−0.002	0.024	0.066	0.036	0.035	0.040	0.094	0.954	0.958	0.966	0.926	0.956
	400	0.002	0.001	0.001	0.001	0.014	0.053	0.033	0.032	0.031	0.079	0.966	0.936	0.958	0.950	0.964
	100	−0.031	0.000	0.001	0.004	0.067	0.196	0.070	0.066	0.065	0.198	0.920	0.944	0.952	0.936	0.940
0.9	200	−0.020	0.001	0.000	0.004	0.035	0.140	0.050	0.043	0.044	0.123	0.928	0.944	0.942	0.964	0.954
	300	−0.008	−0.001	−0.001	0.000	0.019	0.113	0.038	0.039	0.039	0.097	0.942	0.948	0.958	0.936	0.948
	400	−0.004	−0.001	0.000	0.002	0.013	0.095	0.034	0.031	0.033	0.083	0.952	0.942	0.944	0.948	0.942

Table 5. Empirical bias, RMSE, and CP$_{95\%}$ for the true values $\beta_0(p) = -2.975$, $\beta_1(p) = 0.005$, $\beta_2(p) = 0.092$, $\beta_3(p) = -0.958$, $\beta_4(p) = -0.122$, $\beta_5(p) = -0.289$, and $\alpha = 7.425$ with simulated data.

p	n	Bias							RMSE							CP$_{95\%}$						
		$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	$\beta_4(p)$	$\beta_5(p)$	α	$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	$\beta_4(p)$	$\beta_5(p)$	α	$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	$\beta_4(p)$	$\beta_5(p)$	α
	100	0.009	0.001	0.000	−0.003	−0.001	0.001	0.844	0.038	0.034	0.030	0.031	0.029	0.030	1.745	0.928	0.936	0.960	0.946	0.944	0.956	0.964
	200	0.006	0.001	0.000	0.001	0.000	0.001	0.344	0.027	0.020	0.021	0.021	0.020	0.022	1.069	0.926	0.944	0.948	0.950	0.968	0.938	0.952
0.1	300	0.003	0.000	0.000	0.000	0.000	0.000	0.210	0.020	0.018	0.019	0.018	0.017	0.017	0.829	0.942	0.936	0.924	0.964	0.958	0.954	0.958
	400	0.003	0.001	0.001	0.000	−0.001	−0.001	0.187	0.018	0.015	0.015	0.015	0.014	0.017	0.699	0.952	0.952	0.962	0.930	0.960	0.926	0.952
	100	0.006	0.002	0.002	0.002	0.001	0.000	0.851	0.035	0.035	0.032	0.03	0.031	0.032	1.806	0.926	0.930	0.952	0.958	0.938	0.948	0.956
0.25	200	0.002	0.000	0.000	0.001	0.001	0.001	0.341	0.023	0.021	0.024	0.024	0.023	0.023	1.053	0.946	0.944	0.926	0.942	0.938	0.942	0.968
	300	0.003	0.000	−0.001	−0.001	0.001	0.000	0.239	0.019	0.018	0.017	0.018	0.018	0.017	0.793	0.940	0.948	0.942	0.946	0.942	0.948	0.956
	400	0.001	0.000	0.001	0.000	−0.001	0.001	0.223	0.016	0.015	0.016	0.016	0.016	0.015	0.701	0.950	0.95	0.938	0.932	0.950	0.956	0.96
	100	0.003	0.001	0.000	0.000	0.000	−0.002	0.783	0.032	0.031	0.035	0.028	0.033	0.027	1.698	0.936	0.948	0.916	0.954	0.938	0.942	0.974
0.5	200	0.001	0.001	−0.001	0.002	0.001	−0.001	0.441	0.022	0.021	0.020	0.020	0.021	0.020	1.065	0.950	0.944	0.940	0.954	0.928	0.950	0.964
	300	0.001	0.000	0.000	0.000	−0.001	0.002	0.249	0.018	0.017	0.017	0.018	0.018	0.018	0.84	0.936	0.946	0.952	0.940	0.946	0.940	0.964
	400	−0.001	0.000	0.001	0.000	0.000	0.000	0.238	0.015	0.015	0.016	0.015	0.015	0.015	0.688	0.962	0.950	0.962	0.942	0.952	0.948	0.95
	100	−0.003	−0.001	0.001	0.002	0.004	0.000	0.730	0.037	0.030	0.031	0.029	0.033	0.030	1.741	0.930	0.942	0.946	0.956	0.926	0.930	0.956
0.750	200	−0.002	0.000	−0.001	−0.001	0.000	0.000	0.415	0.026	0.022	0.022	0.022	0.022	0.021	1.088	0.948	0.954	0.944	0.964	0.942	0.942	0.956
	300	−0.002	0.000	−0.002	0.000	0.001	0.001	0.302	0.021	0.017	0.017	0.017	0.017	0.020	0.844	0.946	0.954	0.958	0.960	0.962	0.932	0.948
	400	0.000	0.002	−0.001	0.001	0.000	−0.001	0.212	0.017	0.015	0.016	0.016	0.015	0.015	0.695	0.960	0.948	0.930	0.956	0.942	0.952	0.964
	100	−0.017	0.001	0.000	0.000	0.000	0.002	0.919	0.053	0.031	0.029	0.032	0.027	0.031	1.949	0.896	0.948	0.942	0.954	0.930	0.940	0.954
0.9	200	−0.006	0.001	0.000	0.000	0.000	−0.001	0.348	0.036	0.023	0.022	0.021	0.024	0.022	1.012	0.944	0.960	0.926	0.950	0.948	0.952	0.968
	300	−0.005	−0.001	0.000	−0.001	0.000	0.000	0.250	0.028	0.019	0.016	0.018	0.019	0.018	0.830	0.956	0.936	0.944	0.942	0.940	0.934	0.942
	400	−0.003	0.000	0.001	0.000	−0.001	0.000	0.179	0.025	0.015	0.015	0.016	0.017	0.016	0.706	0.938	0.952	0.950	0.944	0.930	0.960	0.948

6. Applications

In this section, the unit-power half-normal quantile regression model is applied to two real-world health data sets. The first one involves patient health outcomes and the second one examines survival rates under a specific treatment. We compare the results with those obtained from the Kumaraswamy (kum), the unit-Birnbaum–Saunders (ubs) [9],

the unit-generalized half-normal-X (ughnx) [13], the unit-Gompertz (ugompertz) [29], and the unit-Weibull (uweibull) [30] quantile regression models.

We performed our calculations by using the R software. To fit the unit-power half-normal quantile regression model, we implemented our own code, combining existing functions available in R packages. For the Kumaraswamy, the unit-Birnbaum–Saunders, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Weibull quantile regression models, we used the R package unitquantreg [31]. The codes are available in a public GitHub repository (https://github.com/DarlinSoto/UPHN (accessed on 12 August 2024)).

6.1. Application 1: Body Fat Data Set

This data set was reported and studied in [32] and consists of 298 observations about the body fat percentage of individuals treated at a public hospital in Curitiba, Paraná, Brazil. The data set includes the variables: fat percentage of arms, legs, body, android, gynecoid, the body mass index (kg/m^2), age (years), gender (female or male), and level of physical activity (sedentary, insufficiently active, or active). The goal is to explore the functional relationship between the covariates gender, age, body mass index, and the level of physical activity, with the response variable body fat in arms. The assumed regression model for ρ_i is

$$
\begin{aligned}
logit(\rho_i(p)) \quad = \quad & \beta_0(p) + \beta_1(p)\text{Age}_i + \beta_2(p)\text{BodyMassIndex}_i + \beta_3(p)\text{Gender\_M}_i \\
& + \beta_4(p)\text{IPAQ\_suffactive} + \beta_5(p)\text{IPAQ\_active}, \quad i = 1, \ldots, 298,
\end{aligned}
$$

where

$$
\text{Gender\_M}_i = \begin{cases} 1, & \text{if the person } i \text{ is a male} \\ 0, & \text{if the person } i \text{ is a female} \end{cases} \tag{25}
$$

$$
\text{IPAQ\_suffactive}_i = \begin{cases} 1, & \text{if the physical activity of the person } i \text{ is insufficiently active} \\ 0, & \text{if the physical activity of the person } i \text{ is sedentary or active} \end{cases} \tag{26}
$$

$$
\text{IPAQ\_active}_i = \begin{cases} 1, & \text{if the physical activity of the person } i \text{ is active} \\ 0, & \text{if the physical activity of the person } i \text{ is sedentary} \\ & \text{or insufficiently active.} \end{cases} \tag{27}
$$

To explore the relationship between gender, age, body mass index, and the level of physical activity with the body fat in arms, the unit-power half-normal, the Kumaraswamy, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Weibull quantile regression models were fitted to the data and the fits were compared by using goodness-of-fit measurements. Table 6 presents the maximum likelihood estimates and their standard errors, Table 7 shows their *p*-values, and Table 8 shows the negative value of the log-likelihood, Akaike information criterion, Bayesian information criterion, and Hannan and Quinn information criterion given by

$$
\text{Neg2LogLike} = -2\log(L) \tag{28}
$$
$$
\text{AIC} = -2\log(L) + 2k \tag{29}
$$
$$
\text{BIC} = -2\log(L) + k\log(n) \tag{30}
$$
$$
\text{HQIC} = -2\log(L) + 2k\log[log(n)] \tag{31}
$$

where *n* is the sample size, *k* is the number of parameters, and *L* is the likelihood function evaluated in the estimated parameters.

According to the results presented in Tables 6 and 7, and for all quantile levels *p*, the unit-power half-normal, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Weibull quantile regression models show that all covariates are statistically significant at the 0.05 level. On the other hand, the Kumaraswamy quantile regression

model shows that all variables are significant except for the `IPAQ_suffactive` variable. Upon analyzing the signs of $\widehat{\beta}_j$, we observe that an increase in covariates such as `Age` and `BodyMassIndex` corresponds to an increase in arm fat percentage. Moreover, males exhibit lower arm fat percentages than females, and individuals with a sedentary lifestyle demonstrate an increase in arm fat percentages compared to those with insufficiently active or active lifestyles.

Table 6. Maximum likelihood estimates and their standard errors for different distributions (uphn, kum, ughnx, ugompertz, and uweibull) in body fat data.

p	Distribution	$\widehat{\beta}_0(p)$	$SE(\widehat{\beta}_0(p))$	$\widehat{\beta}_1(p)$	$SE(\widehat{\beta}_1(p))$	$\widehat{\beta}_2(p)$	$SE(\widehat{\beta}_2(p))$	$\widehat{\beta}_3(p)$	$SE(\widehat{\beta}_3(p))$	$\widehat{\beta}_4(p)$	$SE(\widehat{\beta}_4(p))$	$\widehat{\beta}_5(p)$	$SE(\widehat{\beta}_5(p))$	$\widehat{\alpha}$	$SE(\widehat{\alpha})$	$\widehat{\theta}$	$SE(\widehat{\theta})$
	uphn	−3.315	0.143	0.005	0.001	0.091	0.006	−0.955	0.035	−0.118	0.052	−0.283	0.050	7.921	0.867		
	kum	−3.070	0.137	0.004	0.001	0.073	0.006	−0.730	0.033	−0.073	0.048	−0.196	0.047			4.705	0.210
0.1	ughnx	−3.294	0.155	0.004	0.001	0.080	0.006	−0.799	0.035	−0.099	0.049	−0.228	0.050			2.619	0.116
	ugompertz	−3.436	0.165	0.005	0.001	0.094	0.007	−0.967	0.041	−0.128	0.062	−0.352	0.06			3.657	0.155
	uweibull	−3.214	0.168	0.006	0.001	0.086	0.007	−0.966	0.042	−0.131	0.055	−0.355	0.055			5.976	0.243
	uphn	−3.174	0.147	0.005	0.001	0.092	0.007	−0.958	0.036	−0.122	0.053	−0.289	0.051	7.425	0.789		
	kum	−2.894	0.142	0.004	0.001	0.077	0.006	−0.771	0.035	−0.074	0.050	−0.206	0.050			4.715	0.210
0.25	ughnx	−2.939	0.149	0.004	0.001	0.080	0.006	−0.799	0.035	−0.099	0.049	−0.228	0.050			2.619	0.116
	ugompertz	−3.327	0.169	0.006	0.001	0.098	0.007	−1.000	0.042	−0.136	0.064	−0.364	0.062			3.674	0.155
	uweibull	−2.960	0.159	0.005	0.001	0.082	0.007	−0.919	0.040	−0.127	0.053	−0.339	0.053			5.980	0.243
	uphn	−2.975	0.146	0.005	0.001	0.092	0.007	−0.958	0.036	−0.122	0.053	−0.289	0.051	7.425	0.789		
	kum	−2.754	0.149	0.004	0.001	0.082	0.006	−0.820	0.037	−0.076	0.054	−0.216	0.053			4.724	0.210
0.5	ughnx	−2.653	0.146	0.004	0.001	0.080	0.006	−0.799	0.035	−0.099	0.049	−0.228	0.050			2.619	0.116
	ugompertz	−3.190	0.178	0.006	0.001	0.104	0.008	−1.057	0.044	−0.148	0.068	−0.385	0.066			3.700	0.155
	uweibull	−2.640	0.148	0.005	0.001	0.077	0.006	−0.863	0.037	−0.121	0.050	−0.319	0.050			5.985	0.243
	uphn	−2.772	0.147	0.005	0.001	0.093	0.007	−0.958	0.036	−0.122	0.053	−0.287	0.051	7.440	0.791		
	kum	−2.658	0.157	0.004	0.001	0.087	0.006	−0.870	0.039	−0.078	0.057	−0.227	0.057			4.731	0.210
0.75	ughnx	−2.449	0.144	0.004	0.001	0.080	0.006	−0.799	0.034	−0.099	0.049	−0.228	0.050			2.619	0.116
	ugompertz	−3.040	0.191	0.006	0.002	0.113	0.008	−1.151	0.048	−0.167	0.074	−0.417	0.071			3.737	0.154
	uweibull	−2.272	0.138	0.005	0.001	0.072	0.006	−0.803	0.035	−0.115	0.047	−0.298	0.046			5.992	0.243
	uphn	−2.538	0.148	0.005	0.001	0.091	0.007	−0.957	0.036	−0.121	0.053	−0.285	0.051	7.501	0.799		
	kum	−2.599	0.165	0.005	0.001	0.092	0.007	−0.915	0.042	−0.078	0.06	−0.236	0.060			4.736	0.210
0.9	ughnx	−2.312	0.143	0.004	0.001	0.08	0.006	−0.800	0.034	−0.099	0.049	−0.228	0.050			2.619	0.116
	ugompertz	−2.902	0.212	0.007	0.002	0.127	0.009	−1.296	0.054	−0.195	0.083	−0.466	0.079			3.789	0.153
	uweibull	−1.896	0.130	0.004	0.001	0.067	0.006	−0.746	0.032	−0.109	0.044	−0.278	0.043			6.000	0.243

Table 7. *p*-values of Wald test for $\beta_j(p), j = 0, \dots, 5$, coefficients for different distributions (uphn, kum, ughnx, ugompertz, and uweibull) in body fat data.

p	Distribution	*p*-Value					
		$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$	$\beta_4(p)$	$\beta_5(p)$
	uphn	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.023	$<10^{-3}$
	kum	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.127	$<10^{-3}$
0.1	ughnx	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.046	$<10^{-3}$
	ugompertz	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.038	$<10^{-3}$
	uweibull	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.017	$<10^{-3}$
	uphn	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.021	$<10^{-3}$
	kum	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.139	$<10^{-3}$
0.25	ughnx	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.045	$<10^{-3}$
	ugompertz	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.034	$<10^{-3}$
	uweibull	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.016	$<10^{-3}$
	uphn	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.021	$<10^{-3}$
	kum	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.154	$<10^{-3}$
0.5	ughnx	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.045	$<10^{-3}$
	ugompertz	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.029	$<10^{-3}$
	uweibull	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.015	$<10^{-3}$
	uphn	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.022	$<10^{-3}$
	kum	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.173	$<10^{-3}$
0.75	ughnx	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.045	$<10^{-3}$
	ugompertz	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.023	$<10^{-3}$
	uweibull	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.014	$<10^{-3}$
	uphn	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.022	$<10^{-3}$
	kum	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.193	$<10^{-3}$
0.9	ughnx	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.044	$<10^{-3}$
	ugompertz	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.019	$<10^{-3}$
	uweibull	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.012	$<10^{-3}$

Based on the results given in Table 8, the unit-power half-normal quantile regression model consistently shows higher log-likelihood values across all quantile levels *p* compared

to the Kumaraswamy, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Weibull quantile regression models. Consequently, all goodness-of-fit measurements indicate that the unit-power half-normal quantile regression model provides the best fit to these data, while the unit-Gompertz model provides the worst one.

Table 8. Model selection criteria for different distributions (uphn, kum, ughnx, ugompertz, and uweibull) for body fat data.

p	Distribution	Neg2LogLike	AIC	BIC	HQIC
	uphn	−876.795	−862.795	−836.916	−852.436
	kum	−853.660	−839.660	−813.781	−829.301
0.1	ughnx	−858.131	−844.131	−818.251	−833.771
	ugompertz	−763.903	−749.903	−724.023	−739.544
	uweibull	−852.151	−838.151	−812.271	−827.791
	uphn	−877.216	−863.216	−837.336	−852.856
	kum	−856.224	−842.224	−816.344	−831.865
0.25	ughnx	−858.131	−844.131	−818.251	−833.771
	ugompertz	−766.845	−752.845	−726.965	−742.486
	uweibull	−853.364	−839.364	−813.484	−829.004
	uphn	−877.215	−863.215	−837.336	−852.856
	kum	−859.03	−845.03	−819.151	−834.671
0.5	ughnx	−858.131	−844.131	−818.251	−833.771
	ugompertz	−771.300	−757.300	−731.420	−746.941
	uweibull	−854.932	−840.932	−815.052	−830.573
	uphn	−877.222	−863.222	−837.343	−852.863
	kum	−861.657	−847.657	−821.778	−837.298
0.75	ughnx	−858.131	−844.131	−818.251	−833.771
	ugompertz	−777.425	−763.425	−737.546	−753.066
	uweibull	−856.757	−842.757	−816.877	−832.397
	uphn	−877.185	−863.185	−837.306	−852.826
	kum	−863.813	−849.813	−823.933	−839.454
0.9	ughnx	−858.131	−844.131	−818.251	−833.771
	ugompertz	−784.259	−770.259	−744.379	−759.899
	uweibull	−858.602	−844.602	−818.723	−834.243

To validate these inferences, we conducted a diagnostic analysis for the fitted models. The Cox–Snell quantile residuals are considered, QQ plots are carried out by using their simulated envelopes. The Cramér–von Mises test is applied, where the null hypothesis is that the residuals come from an exponential distribution with a scale parameter of one. These plots are given in Figure 6. This figure suggests that the unit-power half-normal and Kumaraswamy quantile regression models provide a more substantial agreement with the data set compared to the unit-generalized half-normal-X, unit-Gompertz, and unit-Weibull quantile regression models, mainly because most of the residuals envelopes fail to cover the grey points, and the null hypothesis that the residuals come from an exponential distribution is rejected at the 5% significance level.

Therefore, the unit-power half-normal quantile regression model provides the best fit for the data, as it consistently shows the minimum goodness-of-fit measures, with all coefficients being statistically significant. The model's assumptions are validated through diagnostic analyses, which confirm its reliability to describe the relationships between the covariates and arm fat percentage.
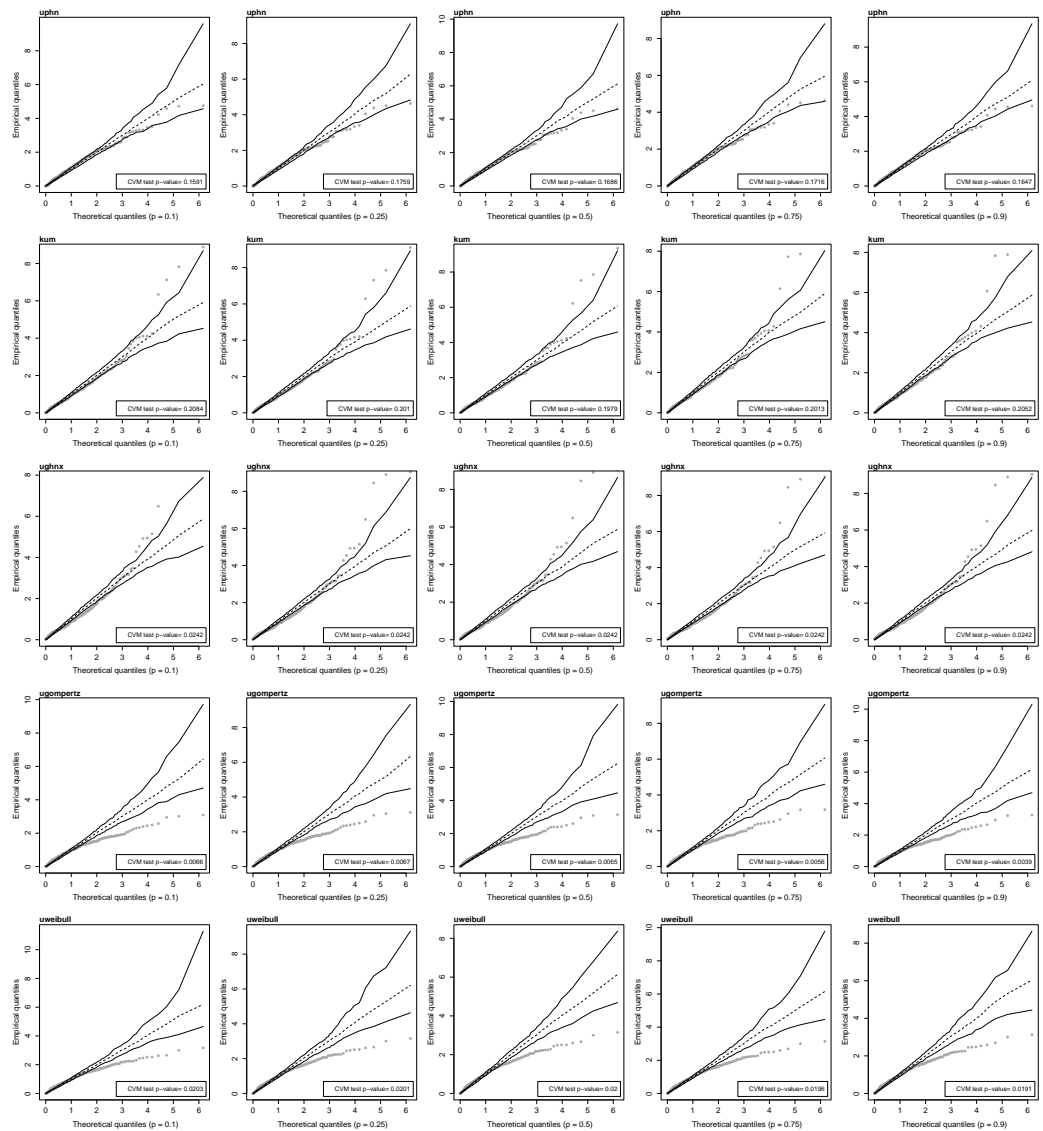
Figure 6. QQ plots with envelopes of Cox–Snell residuals for the indicated 100 *p*-th quantile level and Cramér-von Mises test *p*-value for residuals obtained from different distributions (uphn, kum, ughnx, ugompertz, and uweibull) with body fat data.

6.2. *Application 2: Autologous Stem Cell Transplants Data Set*

We consider another application linked to autologous peripheral blood stem cell (PBSC) transplants, which have been extensively utilized to accelerate hematologic recovery after myeloablative therapy for diverse malignant hematological disorders. The data set comprises a study involving 239 patients who agreed to undergo autologous PBSC transplant following myeloablative chemotherapy doses between 2003 and 2008 at the Edmonton Hematopoietic Stem Cell Lab within the Cross Cancer Institute of Alberta Health Services. This data set has been studied in [33] and contains information about the patients' age, gender, and clinical characteristics. We aim to explain the response variable recovery rates for viable CD34+ cells (`rcd`) in terms of age, gender, and chemotherapy. The regression model assumed for ρ_i is given by

$$logit(\rho_i(p)) = \beta_0(p) + \beta_1(p)\text{Sexmale\_M}_i + \beta_2(p)\text{Age}_i + \beta_3(p)\text{Chemo}_i, \quad i = 1, \ldots, 239, \quad (32)$$

where

$$\text{Sexmale\_M}_i = \begin{cases} 1, & \text{if the gender of the patient } i \text{ is male} \\ 0, & \text{if the gender of the patient } i \text{ is female} \end{cases} \tag{33}$$

$$\text{Chemo}_i = \begin{cases} 1, & \text{if a patient } i \text{ receives a chemotherapy on a 3-day protocol} \\ 0, & \text{if a patient } i \text{ receives a chemotherapy on a 1-day protocol.} \end{cases} \tag{34}$$

We examined the relationship between gender, age, chemotherapy, and the recovery rate for viable CD34+ cells by fitting the unit-power half-normal, the Kumaraswamy, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Birnbaum–Saunders quantile regression models, and evaluated the fits using goodness-of-fit measures.

Tables 9 and 10 present the maximum likelihood estimates, their standard errors, and their *p*-values. For the unit-power half-normal and the unit-Gompertz quantile regression models, all quantile levels show statistically significant regression coefficients for the variables age and chemo at the 5% level. On the other hand, for the unit-generalized half-normal-X and the unit-Birnbaum–Saunders quantile regression models, the variable chemo is the only one statistically significant at the 5% level, while for the Kumaraswamy quantile regression, none of the three variables is statistically significant at the 5% level. Thus, based on the unit-power half-normal and the unit-Gompertz quantile regressions, we may conclude that the covariates `Age` and `Chemo` significantly affect the response variable recovery rate for viable CD34+ cells. Specifically, as `Age` increases, the recovery rate for viable CD34+ cells also increases. Additionally, patients receiving chemotherapy on a three-day protocol exhibit a higher recovery rate for viable CD34+ cells than those on a one-day protocol.

Table 9. Maximum likelihood estimates with their standard errors across different distributions (uphn, kum, ughnx, ugompertz, and ubs) with PBSC data.

p	Distribution	$\hat{\beta}_0$	$SE(\hat{\beta}_0(p))$	$\hat{\beta}_1$	$SE(\hat{\beta}_1(p))$	$\hat{\beta}_2$	$SE(\hat{\beta}_2(p))$	$\hat{\beta}_3$	$SE(\hat{\beta}_3(p))$	$\hat{\alpha}$	$SE(\hat{\alpha})$	$\hat{\theta}$	$SE(\hat{\theta})$
	uphn	−0.498	0.211	0.097	0.090	0.015	0.003	0.397	0.094	1.284	0.104		
	kum	0.281	0.156	0.023	0.055	0.003	0.002	0.085	0.065			6.707	0.454
0.1	ughnx	−1.122	0.457	−0.152	0.167	0.001	0.007	0.380	0.192			0.645	0.028
	ugompertz	−0.456	0.231	0.094	0.098	0.014	0.004	0.337	0.106			2.789	0.342
	ubs	−0.150	0.350	−0.005	0.143	0.008	0.006	0.344	0.155			0.883	0.040
	uphn	−0.180	0.210	0.097	0.090	0.015	0.003	0.397	0.094	1.285	0.104		
	kum	0.653	0.191	0.029	0.071	0.004	0.003	0.113	0.085			6.708	0.454
0.25	ughnx	0.320	0.422	−0.152	0.160	0.001	0.007	0.380	0.179			0.645	0.029
	ugompertz	−0.169	0.252	0.102	0.106	0.015	0.004	0.365	0.115			2.789	0.342
	ubs	0.491	0.314	−0.006	0.130	0.007	0.005	0.316	0.140			0.883	0.040
	uphn	0.275	0.211	0.098	0.090	0.015	0.003	0.397	0.094	1.285	0.104		
	kum	1.053	0.253	0.038	0.096	0.006	0.004	0.159	0.115			6.708	0.454
0.5	ughnx	1.482	0.401	−0.152	0.158	0.001	0.006	0.38	0.174			0.645	0.029
	ugompertz	0.277	0.285	0.112	0.118	0.017	0.005	0.401	0.126			2.788	0.342
	ubs	1.179	0.291	−0.006	0.122	0.007	0.005	0.298	0.131			0.883	0.040
	uphn	0.888	0.218	0.097	0.090	0.015	0.003	0.397	0.094	1.285	0.104		
	kum	1.443	0.343	0.050	0.132	0.008	0.006	0.226	0.158			6.708	0.454
0.75	ughnx	2.309	0.393	−0.152	0.157	0.001	0.006	0.380	0.172			0.645	0.029
	ugompertz	0.945	0.323	0.122	0.129	0.019	0.005	0.436	0.138			2.785	0.342
	ubs	1.820	0.282	−0.007	0.117	0.007	0.005	0.289	0.126			0.883	0.040
	uphn	1.626	0.238	0.097	0.09	0.015	0.003	0.397	0.094	1.285	0.104		
	kum	1.786	0.452	0.063	0.175	0.012	0.007	0.313	0.208			6.708	0.454
0.9	ughnx	2.864	0.389	−0.152	0.157	0.001	0.006	0.380	0.171			0.645	0.029
	ugompertz	1.825	0.350	0.128	0.137	0.020	0.005	0.457	0.146			2.781	0.341
	ubs	2.338	0.279	−0.007	0.115	0.007	0.005	0.284	0.123			0.882	0.040

Table 11 displays the model selection criteria for the unit-power half-normal, the Kumaraswamy, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Birnbaum–Saunders quantile regressions. From this table, it can be said that the log-likelihood values of the unit-power half-normal quantile regression surpass those of the Kumaraswamy, the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Birnbaum–Saunders quantile regressions. Therefore, based on these criteria, we may conclude that the unit-power half-normal model has a superior performance.

Table 10. *p*-values of Wald test for coefficients $\beta_j(p)$, $j = 0, 1, 2, 3$, across different distributions (uphn, kum, ughnx, ugompertz, and ubs) for PBSC data.

p	Distribution	*p*-Value			
		$\beta_0(p)$	$\beta_1(p)$	$\beta_2(p)$	$\beta_3(p)$
	uphn	0.018	0.277	$<10^{-3}$	$<10^{-3}$
	kum	0.071	0.677	0.164	0.189
0.1	ughnx	0.014	0.365	0.932	0.048
	ugompertz	0.048	0.338	$<10^{-3}$	0.001
	ubs	0.669	0.974	0.165	0.027
	uphn	0.392	0.277	$<10^{-3}$	$<10^{-3}$
	kum	$<10^{-3}$	0.685	0.157	0.181
0.25	ughnx	0.448	0.343	0.928	0.034
	ugompertz	0.502	0.338	$<10^{-3}$	0.001
	ubs	0.118	0.965	0.156	0.024
	uphn	0.193	0.276	$<10^{-3}$	$<10^{-3}$
	kum	$<10^{-3}$	0.695	0.147	0.168
0.5	ughnx	$<10^{-3}$	0.337	0.926	0.028
	ugompertz	0.331	0.341	$<10^{-3}$	0.001
	ubs	$<10^{-3}$	0.958	0.15	0.023
	uphn	$<10^{-3}$	0.278	$<10^{-3}$	$<10^{-3}$
	kum	$<10^{-3}$	0.706	0.134	0.151
0.75	ughnx	$<10^{-3}$	0.335	0.926	0.027
	ugompertz	0.003	0.346	$<10^{-3}$	0.002
	ubs	$<10^{-3}$	0.954	0.146	0.022
	uphn	$<10^{-3}$	0.277	$<10^{-3}$	$<10^{-3}$
	kum	$<10^{-3}$	0.718	0.119	0.132
0.9	ughnx	$<10^{-3}$	0.335	0.925	0.026
	ugompertz	$<10^{-3}$	0.350	$<10^{-3}$	0.002
	ubs	$<10^{-3}$	0.951	0.144	0.022

Finally, Figure 7 displays the QQ plots of the Cox–Snell quantile residuals and the *p*-value of the Cramér–von Mises test for the null hypothesis that the residuals come from an exponential distribution with a scale parameter of one. This figure indicates that most of the Cox–Snell quantile residuals obtained from the unit-power half-normal and the Kumaraswamy quantile regressions are inside the simulated envelopes. However, this is not the case for the Cox–Snell quantile residuals obtained from the unit-generalized half-normal-X, the unit-Gompertz, and the unit-Birnbaum–Saunders quantile regressions. These facts suggest that the assumption about Cox–Snell quantile residuals may not be right in those cases. This is validated by the *p*-value of the Cramér–von Mises test, where for the unit-generalized half-normal, the unit-Gompertz, and the unit-Birnbaum–Saunders quantile regressions, the *p*-value is less than 0.05, rejecting the null hypothesis that residuals come from an exponential distribution.

Therefore, the unit-power half-normal quantile regression model is the best fit for the autologous stem cell transplants data. It consistently shows higher log-likelihood values and better model selection criteria compared to the other models. The significant coefficients for age and chemotherapy at the 5% level highlight their strong association with recovery rates for viable CD34+ cells. Furthermore, diagnostic analyses, including QQ plots and the Cramér–von Mises test, validate the model's assumptions, confirming its superior performance.

Table 11. Model selection criteria for different distributions (uphn, kum, ughnx, ugompertz, and ubs) for PBSC data.

p	Distribution	Neg2LogLike	AIC	BIC	HQIC
	uphn	−384.643	−374.643	−357.26	−367.638
	kum	−384.207	−374.207	−356.824	−367.202
0.1	ughnx	−190.055	−180.055	−162.673	−173.051
	ugompertz	−369.921	−359.921	−342.539	−352.916
	ubs	−313.824	−303.824	−286.442	−296.82
	uphn	−384.643	−374.643	−357.26	−367.638
	kum	−384.234	−374.234	−356.851	−367.229
0.25	ughnx	−190.055	−180.055	−162.673	−173.051
	ugompertz	−369.908	−359.908	−342.526	−352.904
	ubs	−313.861	−303.861	−286.479	−296.856
	uphn	−384.643	−374.643	−357.26	−367.638
	kum	−384.271	−374.271	−356.889	−367.267
0.5	ughnx	−190.055	−180.055	−162.673	−173.051
	ugompertz	−369.865	−359.865	−342.483	−352.861
	ubs	−313.889	−303.889	−286.507	−296.885
	uphn	−384.643	−374.643	−357.26	−367.638
	kum	−384.318	−374.318	−356.936	−367.313
0.75	ughnx	−190.055	−180.055	−162.673	−173.051
	ugompertz	−369.793	−359.793	−342.41	−352.788
	ubs	−313.906	−303.906	−286.524	−296.902
	uphn	−384.643	−374.643	−357.26	−367.638
	kum	−384.367	−374.367	−356.985	−367.362
0.9	ughnx	−190.055	−180.055	−162.673	−173.051
	ugompertz	−369.728	−359.728	−342.345	−352.723
	ubs	−313.915	−303.915	−286.532	−296.910
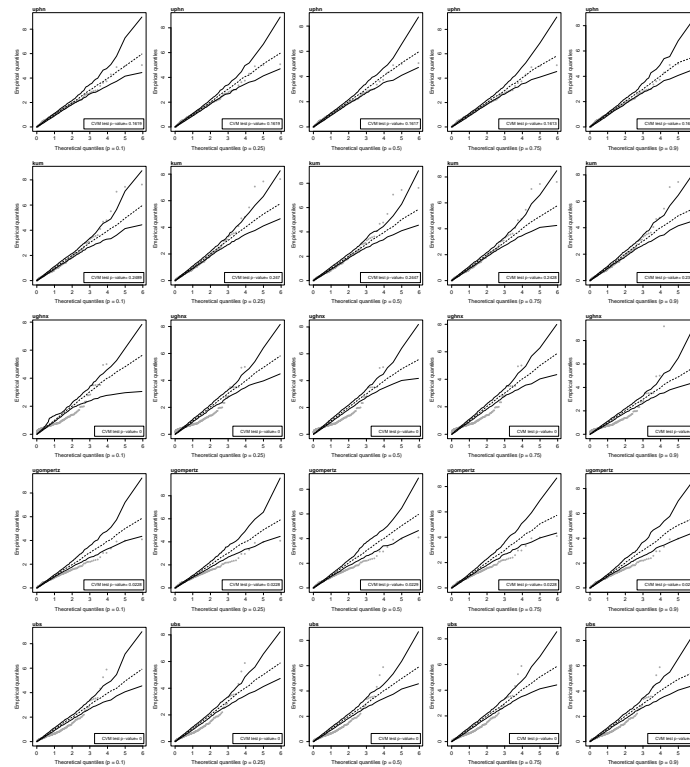
Figure 7. QQ plots with envelopes of Cox–Snell residuals for the proposed 100 p-th quantile level and Cramér–von Mises test *p*-value for residuals obtained from different distributions (uphn, kum, ughnx, ugompertz, and uweibull) with PBSC data.

7. Conclusions

In this paper, we have put forth a new quantile regression model. The predictor covariates were linked to the quantile of the dependent variable through the logit link function. We examined various aspects such as estimation, inference relying on the maximum likelihood method, and diagnoses via the quantile residual. We evaluated two real data sets drawn from medical data using our proposed quantile regression model and the previously introduced Kumaraswamy quantile regression models. The outcomes showed that our proposed quantile regression fitted well to the two real-world data sets. In the first one, the covariates age, body mass index, gender, and level of physical activity are statistically significant to evaluate the level of physical activity with the body fat in arms. For the second data set, the covariates age and chemo are statistically significant to explain the response variable recovery rates for viable CD34+ cells (`rcd`), using our proposed quantile regression model as an alternative to other rival models outlined in the related literature. We carried out a comparison based on different information criteria and the model's goodness of fit, showing its suitability.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data sets are available in the references given in Section 6.

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

References

1. Jones, M.C. Kumaraswamy's Distribution: A Beta-Type Distribution with Some Tractability Advantages. *Stat. Methodol.* **2009**, *6*, 70–81. [CrossRef]
2. Gómez-Déniz, E.; Sordo, M.A.; Calderín-Ojeda, E. The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insur. Math. Econ.* **2014**, *54*, 49–57. [CrossRef]
3. Mazucheli, J.; Bapat, S.R.; Menezes, A.F.B. A new one-parameter unit Lindley distribution. *Chil. J. Stat.* **2019**, *1*, 53–67.
4. Abd El-Monsef, M.M.E.; El-Awady, M.M.; Seyam, M.M. A new quantile regression model for modelling child mortality. *Int. J. Biomath.* **2022**, *10*, 142–149.
5. Altun, E.; Cordeiro, G.M. The unit-improved second-degree Lindley distribution: Inference and regression modelling. *Comput. Stat.* **2020**, *35*, 259–279. [CrossRef]
6. Ahmad, H.H.; Almetwally, E.M.; Elgarhy, M.; Ramadan, D.A. On Unit Exponential Pareto Distribution for Modeling the Recovery Rate of COVID-19. *Processes* **2023**, *11*, 232. [CrossRef]
7. Grassia, A. On a Family of Distributions with Argument between 0 and 1 Obtained by Transformation of the Gamma Distribution and Derived Compound Distributions. *J. Stat.* **1977**, *19*, 108–114. [CrossRef]
8. Ghitany, M.; Mazucheli, J.; Menezes, A.; Alqallaf, F. The unit-inverse Gaussian distribution: A new alternative to two-parameter distributions on the unit interval. *Commun. Stat. Theory Methods* **2018**, *48*, 3423–3438. [CrossRef]
9. Mazucheli, J.; Menezes, A.F.B.; Dey, S. The unit-Birnbaum-Saunders distribution with applications. *Chil. J. Stat.* **1977**, *1*, 47–57.
10. Modi, K.; Gill, V. Unit Burr III distribution with application. *J. Stat. Manag. Syst.* **2019**, *23*, 579–592. [CrossRef]
11. Korkmaz, M.; Chesneau, C. On the unit Burr-XII distribution with the quantile regression modeling and applications. *Comput. Appl. Math.* **2021**, *40*, 29. [CrossRef]
12. Haq, M.; Hashmi, S.; Aidi, K.; Ramos, P.F.L. Unit Modified Burr-III Distribution: Estimation, Characterizations and Validation Test. *Ann. Data Sci.* **2023**, *10*, 415–449. [CrossRef]
13. Bakouch, H.S.; Nikb, A.S.; Asgharzadehb, A.; Salinas, H.S. A flexible probability model for proportion data: Unit-half-normal distribution. *Commun. Stat. Case Stud. Data Anal.* **2021**, *7*, 271–288. [CrossRef]
14. Gómez, Y.M.; Bolfarine, H. Likelihood-based inference for the power half-normal distribution. *J. Stat. Theory Appl.* **2015**, *14*, 383–398. [CrossRef]

15. Barrios, L.; Gómez, Y.M.; Venegas, O.; Barranco-Chamorro, I.; Gómez, H.W. The Slashed Power Half-Normal Distribution with Applications. *Mathematics* **2022**, *10*, 1528. [CrossRef]

16. Pallini, A. The discrete Powe Half-Normal Distribution. *Statistica* **2022**, *82*, 229–242.

17. Lorenz, M.O. Methods of measuring the concentration of wealth. *J. Am. Stat. Assoc.* **1905**, *9*, 209–219. [CrossRef]

18. Gini, C. On the measurement of concentration and variability of characters. *Metron* **2005**, *1*, 1–38.

19. Gini, C. Measurement of inequality of incomes. *Econ. J.* **1921**, *31*, 124–126. [CrossRef]

20. MacGillivray, H.L. Skewness and Asymmetry: Measures and Orderings. *Ann. Stat.* **1986**, *14*, 994–1011. [CrossRef]

21. Moors, J.J. A quantile alternative for kurtosis. *J. R. Stat. Soc. Ser. D* **1988**, *37*, 25–32. [CrossRef]

22. Balakrishnan, N.; Cohen, C.A. *Order Statistics and Inference: Estimation Methods*; Statistical Modeling and Decision Science; Elsevier Science: Amsterdam, The Netherlands, 1991.

23. Bonferroni, C.E. *Elementi di Statistica Generale*; Libreria Seber: Firenze, Italy, 1930.

24. Arcagni, A.; Porro, F. The Graphical Representation of Inequality. *Rev. Colomb. Estad.* **2014**, *37*, 419–436. [CrossRef]

25. MacDonald, I.L. Does Newton-Raphson really fail? *Stat. Methods Med Res.* **2014**, *23*, 308–311. [CrossRef] [PubMed]

26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: https://www.R-project.org/ (accessed on 31 July 2024).

27. Cox, D.R.; Snell, E.J. A general definition of residuals. *J. Royal Stat. Soc. Ser. B* **1968**, *30*, 248–265. [CrossRef]

28. Bolar, K. STAT: Interactive Document for Working with Basic Statistical Analysis. 2019. Available online: https://cran.r-project.org/web/packages/STAT/index.html (accessed on 1 April 2024).

29. Mazucheli, J.; Menezes, A.F.; Dey, S. Unit-Gompertz Distribution with Applications. *Statistica* **2019**, *79*, 25–43.

30. Mazucheli, J.; Menezes, A.F.B.; Alqallaf, F.; Ghitany, M.E. Bias-Corrected Maximum Likelihood Estimators of the Parameters of the Unit-Weibull Distribution. *Austrian J. Stat.* **2021**, *50*, 41–53.

31. Menezes, A.F.B. unitquantreg: Parametric Quantile Regression Models for Bounded Data. 2023. Available online: https://cran.r-project.org/web/packages/unitquantreg/index.html (accessed on 15 April 2024).

32. Petterle, R.; Bonat, W.; Scarpin, C.; Jonasson, T.; Borba, V. Multivariate quasi-beta regression models for continuous bounded data. *Int. J. Biostat.* **2021**, *17*, 39–53. [CrossRef]

33. Yang, H.; Acker, J.P.; Cabuhat, M.; Letcher, B.; Larratt, L.; McGann, L.E. Association of post-thaw viable CD34+ cells and CFU-GM with time to hematopoietic engraftment. *Bone Marrow Transplant.* **2005**, *35*, 881–887. [CrossRef]