

Article

Power Truncated Positive Normal Distribution: A Quantile Regression Approach Applied to Health Databases

Karol I. Santoro ¹, Héctor J. Gómez ^{2,*}, Isaac E. Cortés ³, Tiago M. Magalhães ⁴ and Diego I. Gallardo ⁵

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 Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile

³ Facultad de Ciencias, Universidad Arturo Prat, Avenida Arturo Prat 2120, Iquique 1110939, Chile; isacortes@unap.cl

⁴ Department of Statistics, Institute of Exact Sciences, Federal University of Juiz de Fora, Juiz de Fora 36036-900, Brazil; tiago.magalhaes@ufjf.br

⁵ Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile; dgallardo@ubiobio.cl

* Correspondence: hgomez@uct.cl

Abstract: In this paper we present a new extension of the truncated positive normal (TPN) model, called power truncated positive normal. This extension incorporates a shape parameter that provides more flexibility to the model. In addition, this new extension was reparameterized based on the p -th quantile of the distribution in order to perform quantile regression. The initial values were calculated from a modification of the moment estimators, which allowed the maximum likelihood estimators to be obtained. A simulation study was carried out which suggests good behavior of the maximum likelihood estimators in finite samples. Finally, two applications using health databases are presented.

Keywords: power distribution; maximum likelihood estimators; quantile regression; truncated distribution

MSC: 62E10; 62E15; 62E17



Citation: Santoro, K.I.; Gómez, H.J.; Cortés, I.E.; Magalhães, T.M.; Gallardo, D.I. Power Truncated Positive Normal Distribution: A Quantile Regression Approach Applied to Health Databases. *Axioms* **2024**, *13*, 811. <https://doi.org/10.3390/axioms13120811>

Academic Editor: Hari Mohan Srivastava

Received: 14 October 2024

Revised: 15 November 2024

Accepted: 19 November 2024

Published: 21 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding the nutritional status of a population has become increasingly important. Some studies provide background information on poor dietary quality worldwide, leading to an increase in chronic diseases such as diabetes, cardiovascular disease and obesity. Identifying risk factors through the population's nutritional status allows for the design of public strategies to address this problem. The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutrition status of adults and children in the United States. Several reports on this topic can be found, including one by Fryar et al. [1], who studied the prevalence, treatment and control of hypertension in adults in Los Angeles County, California, and another report by Fryar et al. [2] on the prevalence of overweight, obesity and severe obesity among children and adolescents aged 2–19 years in the United States from 1963–1965 through 2017–2018. Additionally, there are numerous articles in the literature that address this issue, such as those by Steinberg [3], Vale et al. [4] and Livingstone et al. [5], among others.

Further analysis of this data is needed, necessitating the generation of new statistical models for more accurate studies. For example, Yoo [6] proposes a discrete Weibull regression model, while Parker et al. [7] develop a Bayesian model for functional covariates using NHANES data, among many others.

In this paper, we aim to introduce a new model with positive support to help analyze this type of data. The previous literature includes various works on elliptic distributions,

such as the family of spherical distributions, which is a specific case of elliptic distributions. The univariate case falls into the category of symmetric distributions. In 2018, Gómez et al. [8] introduced a positively supported model derived from the normal distribution, known as the truncated positive normal (TPN). This distribution features a shape parameter and a scale parameter. For certain parameter values, the distribution exhibits a bell-shaped behavior. The probability density function (pdf) for the TPN model is given by

$$f(z; \sigma, \lambda) = \frac{1}{\sigma \Phi(\lambda)} \phi\left(\frac{z}{\sigma} - \lambda\right), \quad z, \sigma \in \mathbb{R}^+, \lambda \in \mathbb{R},$$

where σ and λ are the scale and shape parameters, respectively, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cumulative distribution function (cdf) of the standard normal distribution, respectively. The cdf for the TPN model is

$$F_Z(z; \sigma, \lambda) = \frac{\Phi\left(\frac{z}{\sigma} - \lambda\right) + \Phi(\lambda) - 1}{\Phi(\lambda)}, \quad z, \sigma \in \mathbb{R}^+, \lambda \in \mathbb{R}.$$

Recently, there have been several extensions made to the TPN model. For instance, Salinas et al. [9] explored the unit-TPN model, while Gómez et al. [10] extended the model based on the Cooray and Ananda [11] extension. Other extensions of the model, including the slash methodology and a bimodal version, have also been conducted, see [12–14]. Additionally, a family of truncated distributions was introduced and, to facilitate further study, the truncated positive normal model and extensions package [15] was created.

One of the extensions that has been widely studied by various researchers is the one introduced by Lehmann [16] and Durrans [17]. They propose a family of distributions called exponential distributions, where their $F(z)$ corresponds to a cdf. This function is defined as

$$\varphi_F(z; \gamma) = F(z)^\gamma, \quad z \in \mathbb{R}, \gamma > 0.$$

In this way, the pdf is obtained by deriving the new cdf, i.e.,

$$\varphi_f(z; \gamma) = \gamma f(z) F(z)^{\gamma-1}, \quad z \in \mathbb{R}, \gamma > 0.$$

The distribution is primarily derived from the distribution of the maximum. This model is also referred to as Lehmann's model in the literature and has been extensively discussed by Gupta and Gupta [18], as well as by other researchers such as Segovia et al. [19], who introduced the exponential power Maxwell distribution (PM) model, Martinez-Flores et al. [20], who presented the flexible power-normal (FPN) model, and Tovar-Falon and Martinez-Flores [21], who proposed a new class of exponential beta-skew-Laplace distributions, among others.

On the other hand, reparameterized models are useful to study the influence of one or more variables in the response variable. This topic has attracted much attention in recent years. For instance, Ferrari and Cribari-Neto [22] reparameterized the beta model based on a reparameterization of the beta distribution in terms of the mean, which has also attracted the interest of several researchers, such as Ospina and Ferrari [23], Bayer et al. [24], Migliorati et al. [25] and Pereira et al. [26]. Nowadays, there is a need for more general statistical models compared to those that only predict means and modes, as highlighted by Chahuan et al. [27]. In contrast to regression models through the mean, quantile regression, introduced by Koenker and Bassett [28], allows modeling the effect of covariates on the entire response distribution. The recent literature approaches to this problem include those by Lemonte and Moreno-Arenas [29], Korkmaz et al. [30], He et al. [31], Cordeiro et al. [32], Alfó et al. [33], Peng [34], Gómez et al. [35] and Cortés et al. [36], to name a few. Quantile regression allows us to quantify how much any quantile of the response variable is modified by an increase of one unit in any of the covariates.

The first goal of this work is to introduce an extension of the TPN model using the approach developed by Lehmann [16] and Durrans [17]. This model, named power truncated positive normal, provides an alternative to three-parameter distributions for modeling nutrition data obtained from health or other fields with continuous positive observations. Subsequently, the model will be reparameterized based on the p -th quantile of the distribution with the aim of introducing quantile regression in this model. We will apply this method using data on serum immunoglobulin concentrations in preschool children (see [37] for details).

The paper is organized as follows. In Section 2, we introduce a new positively supported model and compute its basic properties such as risk function, quantiles and moments. Section 3 covers the modified moments and maximum likelihood (ML) estimators. In Section 4, we perform a quantile regression approach for our proposal. Section 5 presents a simulation study in order to assess the performance of the ML estimators in finite samples for different parameter combinations. Section 6 presents two applications, one without covariates and the other with covariates, to analyze the performance of the new distribution against other competing models. Finally, conclusions are given in Section 7.

2. Power Truncated Positive Normal Distribution

In this section, we introduce the power truncated positive normal (PTPN) distribution. The pdf and cdf are provided along with some properties.

2.1. Pdf, Cdf and Hazard Functions

Proposition 1. Let $Z \sim \text{PTPN}(\sigma, \lambda, \gamma)$. Then, the cdf is given by

$$F(z; \sigma, \lambda, \gamma) = \left[\frac{\Phi\left(\frac{z}{\sigma} - \lambda\right) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right]^\gamma, \quad z > 0, \quad (1)$$

where $\sigma > 0$ is a scale parameter, and $\lambda \in \mathbb{R}$ and $\gamma > 0$ are shape parameters.

Proposition 2. Let $Z \sim \text{PTPN}(\sigma, \lambda, \gamma)$. Then, the pdf is given by

$$f(z; \sigma, \lambda, \gamma) = \frac{\gamma}{\sigma \Phi(\lambda)^\gamma} \phi\left(\frac{z}{\sigma} - \lambda\right) \left(\Phi\left(\frac{z}{\sigma} - \lambda\right) + \Phi(\lambda) - 1\right)^{\gamma-1}, \quad z > 0, \quad (2)$$

Proof. The demonstration is straightforward by deriving Equation (1). $\square$

Proposition 3. Let $Z \sim \text{PTPN}(\sigma, \lambda, \gamma)$. Then, the hazard function of Z is given by

$$h(z; \sigma, \lambda, \gamma) = \frac{\gamma \phi\left(\frac{z}{\sigma} - \lambda\right) \left(\Phi\left(\frac{z}{\sigma} - \lambda\right) + \Phi(\lambda) - 1\right)^{\gamma-1}}{\sigma \left(\Phi(\lambda)^\gamma - \left(\Phi\left(\frac{z}{\sigma} - \lambda\right) + \Phi(\lambda) - 1\right)^\gamma\right)}, \quad z > 0, \quad (3)$$

Proof. The result is obtained from the definition $h(z) = \frac{f(z)}{1 - F(z)}$. $\square$

Figure 1 shows the pdf, cdf and hazard function for the $\text{PTPN}(\sigma = 1, \lambda, \gamma)$ model, considering some values for λ and γ . It is observed that the pdf for the PTPN model can exhibit decreasing and uni-modal forms, whereas the hazard function is strictly increasing.

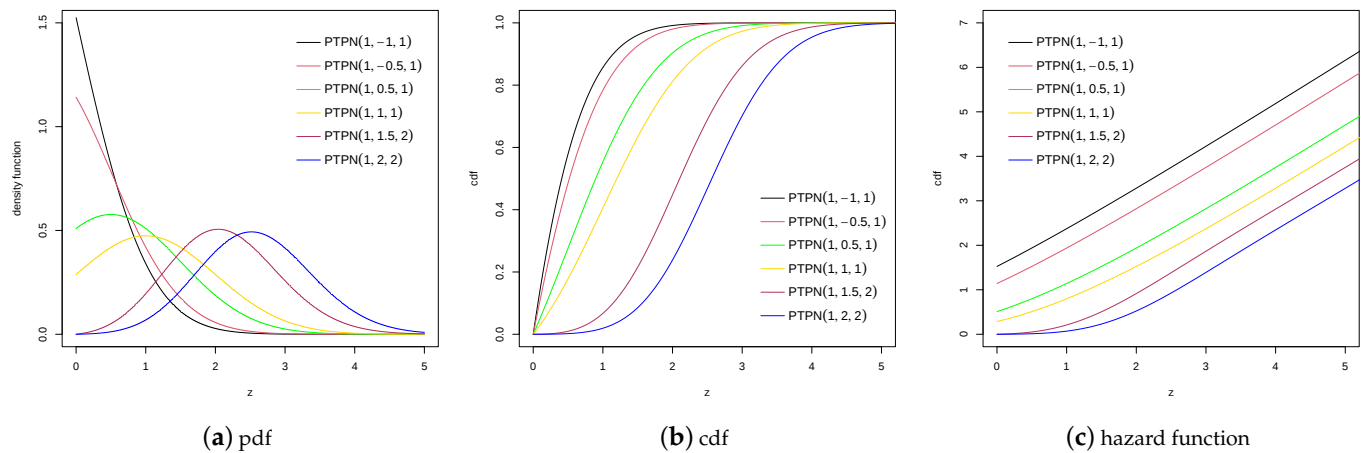


Figure 1. Pdf, cdf and hazard function for the $PTPN(\sigma = 1, \lambda, \gamma)$ model with different combinations for λ and α .

Observation 1. Let $Z \sim PTPN(\sigma, \lambda, \gamma)$. The following distributions are special cases of the PTPN distribution.

1. If $\gamma = 1$, then $PTPN(\sigma, \lambda, \gamma = 1)$ reduces to the $TPN(\sigma, \lambda)$ distribution.
2. If $\lambda = 0$ and $\gamma = 1$, then $PTPN(\sigma, \lambda = 0, \gamma = 1)$ reduces to the half-normal distribution (we denote $HN(\sigma)$).
3. If $\lambda = 0$, then $PTPN(\sigma, \lambda = 0, \gamma)$ reduces to the power half-normal distribution, $PHN(\sigma, \gamma)$, introduced in [38].

Figure 2 illustrates the connections between the distribution of PTPN and the previously mentioned special cases.

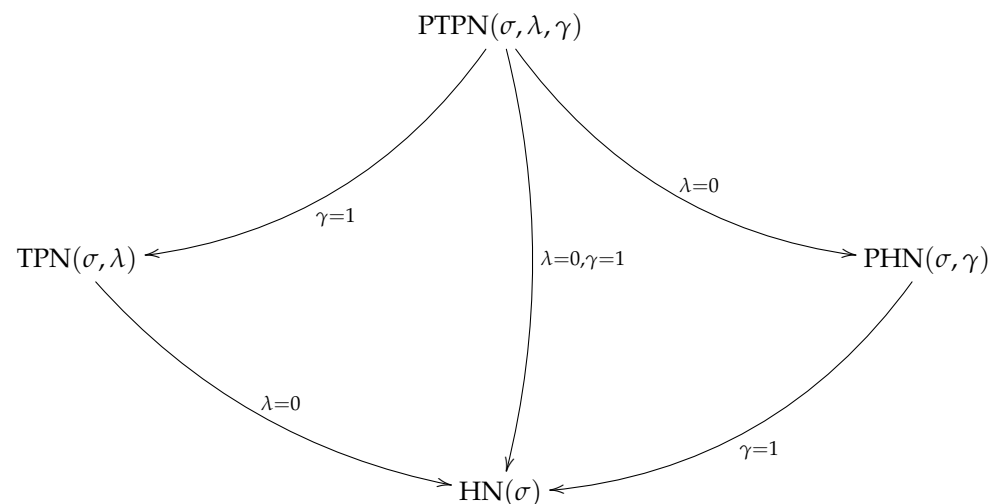


Figure 2. Particular cases of the PTPN distribution.

2.2. Modes

The shape of the pdf of $Z \sim PTPN(\sigma, \lambda, \gamma)$ can be analyzed by identifying its inflection points. To find this value, we can take two paths; we will develop one of them as follows: We calculate the first derivative of $\log f(z)$, where $f(z)$ represents the pdf for the PTPN model, providing

$$\frac{\partial \log(f(z))}{\partial z} = -\frac{v}{\sigma} + \frac{(\gamma - 1)\phi(v)}{\sigma(\Phi(v) + \Phi(\lambda) - 1)},$$

where $v = \frac{z}{\sigma} - \lambda$. To find the solution, we can set the previous expression equal to zero, resulting in

$$\phi(v) = \frac{v(\Phi(v) + \Phi(\lambda) - 1)}{(\gamma - 1)}, \quad (4)$$

from which the mode of Z can be obtained numerically. The nature of the points are determined by $u(z) = \partial^2 \log(f(z))/\partial z^2$, where $u(z)$ is given by the following:

$$u(z) = -\frac{1}{\sigma^2} - \frac{(\gamma - 1)\phi(v)}{\sigma^2} \left(\frac{v(\Phi(v) + \Phi(\lambda) - 1) + \Phi(v)}{(\Phi(v) + \Phi(\lambda) - 1)^2} \right). \quad (5)$$

Depending on whether $u(z_0) < 0$ or $u(z_0) > 0$, where $z = z_0$ is a solution of Equation (4), the inflection points can be local maxima or minima. Figure 3 shows the shape of $u(z)$ for $\sigma = 1$ and selected values of λ and γ . From here, we observe that the pdf of the PTPN distribution has a local maximum.

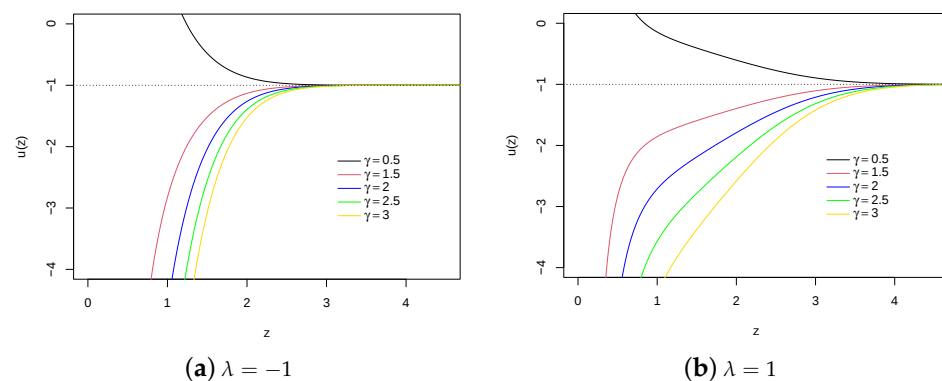


Figure 3. Shape of $u(z)$ for $\sigma = 1$ and some selected values of α and λ .

The second way we will approach it is through its definition, i.e., the mode is the value at which the pdf reaches its maximum. As studied above, the calculation of the mode is not straightforward; however, it is of interest to observe the values that the mode takes for different parameters. Therefore, an expression for the first derivative of the pdf is introduced in Proposition 4 and the values of the mode for various parameter values are presented in Table 1.

Proposition 4. Let $Z \sim \text{PTPN}(\sigma, \lambda, \gamma)$. The mode of Z is obtained as the solution for the following non-linear equation for z

$$(\gamma - 1)\sigma\phi(v)(\Phi(v) + \Phi(\lambda) - 1)^{-1} - (z - \lambda\sigma) = 0. \quad (6)$$

Proof. The result is obtained by deriving the pdf of PTPN and making it equal to zero, as shown below:

$$\frac{\partial f(z)}{\partial z} = \frac{\gamma}{\sigma\Phi(\lambda)^\gamma} \left[-\frac{\phi(v)}{\sigma} \left(\frac{z}{\sigma} - \lambda \right) (\Phi(v) - \Phi(\lambda) - 1)^{\gamma-1} + \frac{\phi(v)^2}{\sigma} (\gamma - 1)(\Phi(v) + \Phi(\lambda) - 1)^{\gamma-2} \right].$$

Therefore, $\partial f(z)/\partial z = 0$ is equivalent to

$$(\gamma - 1)\sigma\phi(v)(\Phi(v) + \Phi(\lambda) - 1)^{-1} - (z - \lambda\sigma) = 0.$$

□

Remark 1. The mode of the PTPN(σ, λ, γ) model when $\gamma = 1$ is as follows:

1. If $\lambda \geq 0$, then the model will be attained at $z = \sigma\lambda$.
2. If $\lambda < 0$, the pdf of the PTPN model will be strictly decreasing in $(0, \infty)$ and, then, the mode of the model is 0.

For different values of the parameters in Equation (6), different values of the mode are obtained. Table 1 shows some of these values for $\sigma = 1$, $\lambda = 1$, and $\gamma = 1, 2, 3, 5, 9$ and 12.

Table 1. Mode for $\sigma = 1$ and $\lambda = \{-1, 1\}$ associated with the PTPN distribution.

$\lambda = 1$		$\lambda = -1$	
γ	Mode	γ	Mode
1	1	1.5	0.313
2	1.593	2	0.483
3	1.857	3	0.690
5	2.150	5	0.921
9	2.451	9	1.160
12	2.587	12	1.270

2.3. Moments

The moments for $Z \sim \text{PTPN}(\sigma, \lambda, \gamma)$ can be expressed using a generic expression, as shown in Gupta and Gupta [18], given by

$$E(Z^n) = \gamma \int_{-\infty}^{\infty} z^n f(z) \{F(z)\}^{\gamma-1} dz = \gamma \int_0^1 [F^{-1}(u)]^n u^{\gamma-1} du, \quad (7)$$

where $u = F(z)$. The following proposition is between the moments of the PTPN distribution.

Proposition 5. Let $Z \sim \text{PTPN}(\sigma, \lambda, \gamma)$ and n be a positive integer. Then, the n -th moment of Z is given by

$$\mu_n = E[Z^n] = \sum_{k=0}^n \binom{n}{k} \gamma \sigma^n \lambda^k I_{n,k}, \quad (8)$$

$$\text{where } I_{n,k} = \int_0^1 [\Phi^{-1}(1 + (u-1)\Phi(\lambda))]^{n-k} u^{\gamma-1} du.$$

Proof. By Equation (7), we obtain:

$$E[Z^n] = \gamma \sigma^n \int_0^1 u^{\gamma-1} [\Phi^{-1}(1 + (u^{1/\gamma} - 1)\Phi(\lambda)) + \lambda]^n du.$$

Therefore, by the binomial theorem, $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$ and then

$$\begin{aligned} E[Z^n] &= \gamma \sigma^n \int_0^1 \sum_{k=0}^n \binom{n}{k} u^{\gamma-1} \lambda^k [\Phi^{-1}(1 + (u^{1/\gamma} - 1)\Phi(\lambda)) + \lambda]^{n-k} du, \\ &= \sum_{k=0}^n \binom{n}{k} \gamma \sigma^n \lambda^k \int_0^1 u^{\gamma-1} [\Phi^{-1}(1 + (u^{1/\gamma} - 1)\Phi(\lambda)) + \lambda]^{n-k} du. \end{aligned}$$

□

Corollary 1. Let $Z \sim PTPN(\sigma, \lambda, \gamma)$. Then, the skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) coefficients are given by

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

respectively, where $b_n = b_n(\lambda) = \sum_{k=0}^n \binom{n}{k} \lambda^k I_{n,k}$.

Table 2 presents the values of the mean, standard deviation, kurtosis coefficient and skewness for various combinations of λ, γ with $\sigma = 1$ fixed. It is observed that, for negative values of λ and as γ decreases, the kurtosis is increased. This behavior is illustrated in Figure 4.

Table 2. The table shows the mean, standard deviation, skewness and kurtosis for different parameter combinations.

λ	γ	Mean	sd	Skewness	Kurtosis
−2.00	0.50	0.23	0.29	2.12	8.76
	1.00	0.37	0.34	1.54	6.02
	3.00	0.66	0.37	1.04	4.49
1.00	0.50	0.86	0.78	1.02	3.60
	1.00	1.29	0.79	0.59	3.00
	3.00	1.98	0.69	0.37	3.12
1.50	0.50	1.13	0.91	0.76	3.01
	1.00	1.64	0.88	0.39	2.80
	3.00	2.40	0.72	0.29	3.10

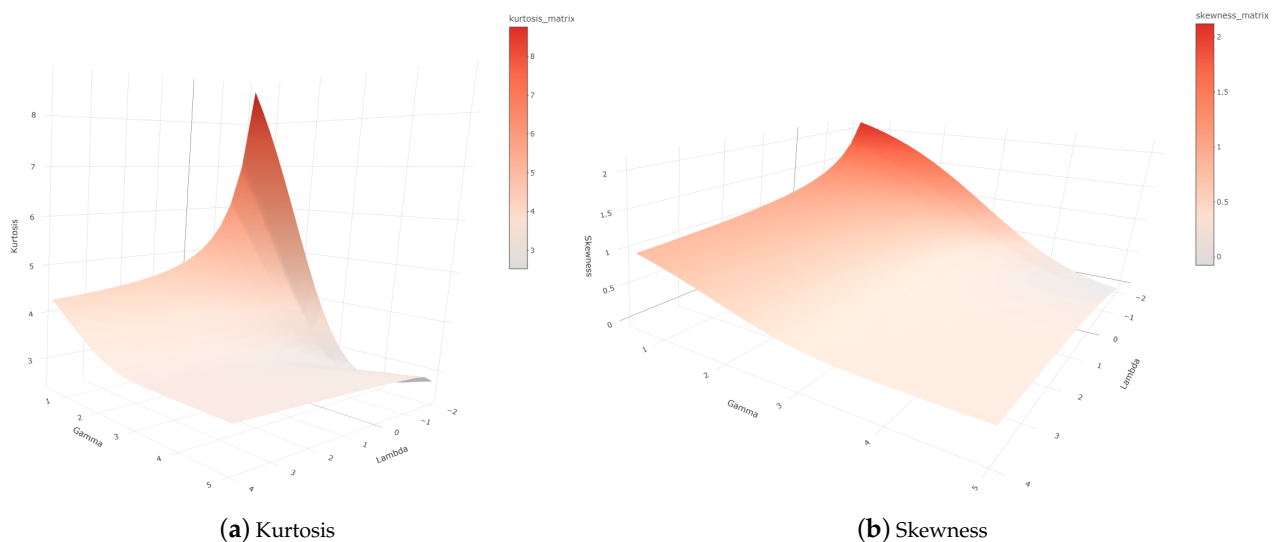


Figure 4. (a) Plots of the kurtosis for PTPN ($\sigma = 1, \lambda, \gamma$). (b) Plots of the skewness for PTPN ($\sigma = 1, \lambda, \gamma$).

2.4. Quantile Function

Proposition 6. Let $Z \sim PTPN(\sigma, \lambda, \gamma)$. Then, the quantile function of Z is given by the following:

$$Q(p) = \sigma[\Phi^{-1}(\Phi(\lambda)(p^{1/\gamma} - 1) + 1) + \lambda]. \quad (9)$$

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

Corollary 2. The quartiles for the PTPN distribution are as follows:

1. (First quartile) $Q(0.25; \sigma, \lambda, \gamma) = \sigma[\Phi^{-1}(\Phi(\lambda)(0.25^{1/\gamma} - 1) + 1) + \lambda]$.
2. (Median) $Q(0.5; \sigma, \lambda, \gamma) = \sigma[\Phi^{-1}(\Phi(\lambda)(0.5^{1/\gamma} - 1) + 1) + \lambda]$.
3. (Third quartile) $Q(0.75; \sigma, \lambda, \gamma) = \sigma[\Phi^{-1}(\Phi(\lambda)(0.75^{1/\gamma} - 1) + 1) + \lambda]$.

2.5. Order Statistics

The order statistics have various applications in the physical and life sciences (see Balakrishnan and Cohen [39]). From a statistical standpoint, they enable the computation of useful functions such as the sample range and the sample median. The following result states the probability density function of the k -th order statistic from a PTPN random sample of size n , which is arranged in non-decreasing order.

Proposition 7. Suppose the random variables $z_1, z_2, \dots, z_n$ are independent and identically distributed PTPN random variables. Then, the pdf of the k -th order is given by the following:

$$f_{X;k:n}(x) = \frac{\gamma n!}{(k-1)!(n-k)! \sigma \Phi^\gamma(\lambda)} \phi(v) [\Phi(v) + \Phi(\lambda) - 1]^{\gamma-1} \left(\frac{\Phi(v) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right)^{\gamma(k-1)} \left[1 - \left(\frac{\Phi(v) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right)^\gamma \right]^{n-k}.$$

Corollary 3. Suppose $z_1, z_2, \dots, z_n$ are random samples from the PTPN distribution. Then, the density function

1. $f_{X;1:n}(x)$ of the 1-st order statistics is given by the following:

$$f_{X;1:n}(x) = \frac{\gamma n}{\sigma \Phi^\gamma(\lambda)} \phi(v) [\Phi(v) + \Phi(\lambda) - 1]^{\gamma-1} \left[1 - \left(\frac{\Phi(v) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right)^\gamma \right]^{n-1}.$$

2. $f_{X;n:n}(x)$ of the n -th order statistics is given by the following:

$$f_{X;n:n}(x) = \frac{\gamma n}{\sigma \Phi^\gamma(\lambda)} \phi(v) [\Phi(v) + \Phi(\lambda) - 1]^{\gamma-1} \left(\frac{\Phi(v) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right)^{\gamma(n-1)}.$$

2.6. Shannon Entropy

The Shannon entropy of a random variable measures the uncertainty associated with its possible values and their probabilities, which allows optimizing the representation and transmission of that information. It is defined as follows:

$$S(Z) = -\mathbb{E}(\log f(Z)).$$

Using the definition in the PTPN we obtain the following:

$$S(Z) = \log \left(\frac{\sqrt{2\pi\sigma\Phi(\lambda)^\gamma}}{\gamma} \right) - (\gamma-1)\mu^* + \frac{\mu_2}{2\sigma^2} - \frac{\lambda\mu}{\sigma} + \frac{\lambda^2}{2},$$

where $\mu^* = \mathbb{E}(\log(\Phi(\frac{z}{\sigma} - \lambda) + \Phi(\lambda) - 1))$. Figure 5 shows different shapes for the entropy, including its particular cases TPN and HN.

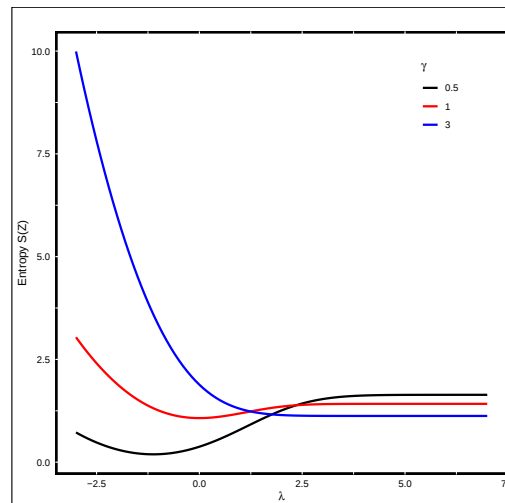


Figure 5. Entropy for the $PTPN(\sigma = 1, \lambda, \gamma)$ model.

3. Inference

In this section, we will discuss the estimation procedure for the parameters of the PTPN distribution from a classical point of view. For this, let us consider $z_1, z_2, \dots, z_n$ as a random sample from $Z \sim PTPN(\sigma, \lambda, \gamma)$.

3.1. Modified Moment Estimators

It is well known that, for any random variable X with cdf $F(\cdot)$, we have $F(x) \sim U(0, 1)$ and $-\log(F(x)) \sim \text{Exp}(1)$ (i.e., the standard exponential model). Therefore, $E(F(x)) = 1/2$ and $E(-\log(F(x))) = 1$. Fixing $\gamma = \hat{\gamma}_M$ at a certain value and using a modified version of the moment estimators, we can find that the estimators $\hat{\sigma}_M$ and $\hat{\lambda}_M$ are given by the solution of the equations

$$\frac{1}{n} \sum_{i=1}^n \left[\frac{\Phi\left(\frac{z_i}{\sigma} - \lambda\right) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right]^{\hat{\gamma}_M} = \frac{1}{2}, \quad \text{and} \quad (10)$$

$$\frac{1}{n} \sum_{i=1}^n -\log \left(\left[\frac{\Phi\left(\frac{z_i}{\sigma} - \lambda\right) + \Phi(\lambda) - 1}{\Phi(\lambda)} \right]^{\hat{\gamma}_M} \right) = 1. \quad (11)$$

Equations (10) and (11) provide a system of non-linear equations. The `nleqslv` function of R 4.4.2 software [40] can be used to solve the system and obtain the values as initial estimates for the calculation of the ML estimators.

3.2. Maximum Likelihood Estimators

Given $z_1, z_2, \dots, z_n$, a random sample of size n from $PTPN(\sigma, \lambda, \gamma)$, the log-likelihood function is given by the following:

$$\ell(\theta) = n \left(\log(\gamma) - \log(\sigma) - \gamma \log[\Phi(\lambda)] - \frac{\log(2\pi)}{2} \right) - \frac{1}{2} \sum_{i=1}^n v_i^2 + (\gamma - 1) \sum_{i=1}^n \log[\Phi(v_i) + \Phi(\lambda) - 1], \quad (12)$$

where $v_i = \frac{z_i}{\sigma} - \lambda$. Therefore, the score assumes the form $S(\theta) = (S_\sigma(\theta), S_\lambda(\theta), S_\gamma(\theta))$ where

$$S_{\sigma}(\theta) = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n z_i v_i + \frac{(1-\gamma)}{\sigma^2} \sum_{i=1}^n z_i G(v_i), \quad (13)$$

$$S_{\lambda}(\theta) = -\frac{n\gamma\phi(\lambda)}{\Phi(\lambda)} + \sum_{i=1}^n v_i + (1-\gamma) \sum_{i=1}^n G(v_i) \left[1 - \frac{\phi(\lambda)}{\phi(v_i)} \right], \quad \text{and} \quad (14)$$

$$S_{\gamma}(\theta) = \frac{n}{\gamma} - n \log [\Phi(\lambda)] + \sum_{i=1}^n \log [\Phi(v_i) + \Phi(\lambda) - 1], \quad (15)$$

where $G(v_i) = \frac{\phi(v_i)}{\Phi(v_i) + \Phi(\lambda) - 1}$. The ML estimators (MLEs) can be obtained by solving the likelihood equations $S(\theta) = \mathbf{0}_3$, where $\mathbf{0}_3$ is a vector of length 3 with zeros. Numerical methods, like the Newton–Raphson procedure, can be used to solve these equations. Alternatively, other maximization techniques, such as the one proposed by MacDonald [41], could also be applied.

3.3. Observed Fisher Information Matrix

The asymptotic variance of the MLEs, say $\hat{\theta} = (\hat{\sigma}, \hat{\lambda}, \hat{\gamma})$, can be estimated by the Fisher information matrix defined as $\mathcal{I}(\theta) = -\mathbb{E}[\partial^2 \ell(\theta) / \partial \theta \partial \theta^T]$, where $\ell(\theta)$ is the log-likelihood function of the UPHN model given in (12). Under the regularity conditions,

$$\mathcal{I}(\theta)^{-1/2} (\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_3(\mathbf{0}_3, \mathbf{I}_3), \quad \text{as } n \rightarrow +\infty, \quad (16)$$

where $\mathcal{D}$ denotes the convergence in the distribution and $N_3(\mathbf{0}_3, \mathbf{I}_3)$ denotes the standard bivariate normal distribution. The elements of the matrix $\partial^2 \ell(\theta) / \partial \theta \partial \theta^T$ are given by $I_{\sigma\sigma} = \partial^2 \ell(\theta) / \partial \sigma^2$, $I_{\sigma\lambda} = \partial^2 \ell(\theta) / \partial \sigma \partial \lambda$ and so on. Explicitly, we have the following:

$$\begin{aligned} I_{\sigma\sigma} &= \frac{n}{\sigma^2} - 2 \sum_{i=1}^n \frac{z_i v_i}{\sigma^3} - \sum_{i=1}^n \frac{z_i^2}{\sigma^4} + \frac{2(\gamma-1)}{\sigma^3} \sum_{i=1}^n z_i G(v_i) - \frac{(\gamma-1)}{\sigma^4} \sum_{i=1}^n z_i^2 G(v_i) [v_i + G_i(v_i)], \\ I_{\sigma\lambda} &= -\sum_{i=1}^n \frac{z_i}{\sigma^2} + \frac{(1-\gamma)}{\sigma^2} \sum_{i=1}^n v_i G(v_i) + \frac{(1-\gamma)}{\sigma^2} \sum_{i=1}^n z_i G^2(v_i) \left[1 - \frac{\phi(\lambda)}{\phi(v_i)} \right], \\ I_{\sigma\gamma} &= -\frac{1}{\sigma^2} \sum_{i=1}^n z_i G(v_i), \\ I_{\lambda\lambda} &= -n + n\gamma \frac{\phi(\lambda)}{\Phi(\lambda)} \left[\lambda + \frac{\phi(\lambda)}{\Phi(\lambda)} \right] + (1-\gamma) \sum_{i=1}^n G(v_i) \left[v_i + \lambda \frac{\phi(\lambda)}{\phi(v_i)} \right] + (1-\gamma) \sum_{i=1}^n G^2(v_i) \left[1 - \frac{\phi(\lambda)}{\phi(v_i)} \right]^2, \\ I_{\lambda\gamma} &= -\frac{n\phi(\lambda)}{\Phi(\lambda)} - \sum_{i=1}^n G(v_i) \left[1 - \frac{\phi(\lambda)}{\phi(v_i)} \right], \quad \text{and} \\ I_{\gamma\gamma} &= -\frac{n}{\gamma^2}. \end{aligned}$$

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

$$I(\hat{\theta}) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}}.$$

The asymptotic variances of $\hat{\sigma}$, $\hat{\lambda}$ and $\hat{\gamma}$ are estimated by the diagonal elements of $I(\hat{\theta})^{-1}$.

To ensure the existence and uniqueness of maximum likelihood estimators, it is necessary to demonstrate that the log-likelihood function $\ell(\theta)$ is strictly concave over the

parameter space. This can be verified by checking that the matrix $\partial^2 \ell(\theta) / \partial \theta \partial \theta^\top$ is negative definite as follows:

- $I_{\sigma\sigma} < 0$;
- $I_{\sigma\sigma} I_{\lambda\lambda} > I_{\sigma\lambda}^2$;
- $I_{\sigma\sigma} I_{\lambda\lambda} I_{\gamma\gamma} + I_{\sigma\lambda} I_{\lambda\gamma} I_{\sigma\gamma} + I_{\sigma\gamma} I_{\sigma\lambda} I_{\lambda\gamma} < I_{\lambda\lambda} I_{\sigma\gamma}^2 + I_{\sigma\sigma} I_{\lambda\gamma}^2 + I_{\gamma\gamma} I_{\sigma\lambda}^2$.

4. Quantile Regression Model

The objective of this section is to formulate the pdf, cdf, quantile regression model and log-likelihood function associated with the reparameterized PTPN distribution.

For the PTPN model, the mean is a function of an integral that is solved numerically, so it is not possible to consider a mean-parameterized version of the model. On the other hand, in a context of heterogeneous observations, quantile regression is a more appropriate approach for analyzing data in the presence of covariates because it allows a complete description of the distribution of the response variable, not only of a specific measure as it occurs when regression on the mean is used.

Specifically, for the PTPN model and considering that $\tau = \tau_p$ represents the p -th quantile of the distribution, we obtain the equation $\tau = Q(p; \sigma, \lambda, \gamma)$, $\tau = Q(0, \infty)$. Solving this equation, we obtain $\sigma = \tau(h(\lambda, \gamma, p))^{-1}$, where $h(\lambda, \gamma, p) = \Phi^{-1}[\Phi(\lambda)(p^{1/\gamma} - 1) + 1]$. Thus, we can reparameterize the pdf and cdf of the PTPN model as

$$\begin{aligned} f(z; \tau, \lambda, \gamma) &= \frac{\gamma h(\lambda, \gamma, p)}{\tau [\Phi(\lambda)]^\gamma} \phi\left(\frac{zh(\lambda, \gamma, p)}{\tau} - \lambda\right) \left(\Phi\left(\frac{zh(\lambda, \gamma, p)}{\tau} - \lambda\right) + \Phi(\lambda - 1)\right)^{\gamma-1} \\ F(z; \tau, \lambda, \gamma) &= \left[\frac{\Phi\left(\frac{zh(\lambda, \gamma, p)}{\tau} - \lambda\right) \Phi(\lambda) - 1}{\Phi(\lambda)}\right]^\gamma \end{aligned}$$

respectively, where $z, \gamma, \tau > 0$, $\lambda \in \mathbb{R}$ and $0 < p < 1$ is fixed. We refer to this model as the reparameterized PTPN (RPTPN) model.

Considering $\mathbf{z}_i^\top = (z_{i1}, z_{i2}, \dots, z_{iq})$, a set of q known covariates related to the p -th quantile of the i -th individual, it can be introduced in the model as follows:

$$\psi(\tau_i(p)) = \mathbf{z}_i^\top \boldsymbol{\beta}(p), \quad (17)$$

where $\boldsymbol{\beta}(p) = (\beta_1(p), \beta_2(p), \dots, \beta_q(p))^\top$ is a q -dimensional vector of unknown regression parameters ($q < n$) and $\psi(\cdot)$ is a link function, which is continuous, invertible and at least twice differentiable. A natural choice in this context is the logarithm link, i.e., $\psi(u) = \log(u)$. With this framework, the corresponding log-likelihood function for the RPTPN quantile regression model is given by

$$\begin{aligned} \ell(\theta(p)) &= n \left(\log(\gamma(p)) - \log(\tau_i(p)) - \log(h(\lambda(p), \gamma(p), p)) - \gamma(p) \log[\Phi(\lambda(p))] - \frac{\log(2\pi)}{2} \right) \\ &+ \frac{1}{2} \sum_{i=1}^n v_{i,h} + (\gamma(p) - 1) \sum_{i=1}^n \log[\Phi(v_{i,h}) + \Phi(\lambda(p)) - 1], \end{aligned}$$

where $v_{i,h} = \frac{z_i(h(\lambda(p), \gamma(p), p))}{\sigma(p)} - \lambda(p)$. The estimation of the regression parameters are obtained by directly maximizing this function.

5. Simulation

In this section, a Monte Carlo simulation study [42] is carried out to evaluate the performance of the ML estimators using R software [40].

Without Covariates

An algorithm to generate samples from the $PTPN(\sigma, \lambda, \gamma)$ is provided. This algorithm is based on the inverse transform sampling method, which is detailed in Algorithm 1.

Algorithm 1 Simulating values from the $PTPN(\sigma, \lambda, \gamma)$ distribution.

- 1: Fix the values for σ , λ and γ .
 - 2: Simulate $U_i \sim U(0, 1)$, where $U(0, 1)$ denotes the continuous uniform distribution over the interval $(0, 1)$.
 - 3: Calculate $Z_i = Q(U_i)$, where $Z_i \sim PTPN(\sigma, \lambda, \gamma)$ and Q is the quantile function as defined in Equation (9).
 - 4: Repeat the previous steps for $i = 1, \dots, n$.
-

As parameter values in our simulation, we consider $\sigma \in \{1, 2, 3\}$, $\lambda \in \{2, 3\}$ and $\gamma \in \{0.75, 2.5\}$. For the sample size, we consider $n \in \{150, 300, 600, 1000\}$. For each of the 48 combinations of sample size, and σ , λ and γ , we perform 1000 replicates and the corresponding ML estimates are calculated. To assess the performance of the estimators, we provide the estimated bias (bias), the mean of the standard errors obtained in each replicate (SE) and the root of the estimated mean square error (RMSE). These terms are defined as

$$\text{bias}(\xi) = \hat{\xi}_{MC} - \xi, \quad \text{SE}(\xi) = \sqrt{\frac{1}{999} \sum_{i=b}^{1,000} (\hat{\xi}_b - \hat{\xi}_{MC})^2},$$

$$\text{RMSE}(\xi) = \sqrt{\frac{1}{1,000} \sum_{b=1}^{1,000} (\hat{\xi}_b - \xi)^2} \quad \text{and} \quad \text{CP}(\xi) = \frac{1}{1,000} \sum_{b=1}^{1,000} I_{CI_b(\hat{\xi})}(\xi),$$

where $\xi \in \{\sigma, \lambda, \gamma\}$, $\hat{\xi}_b$ denotes the estimates obtained for ξ at the b -th replicate, $\hat{\xi}_{MC} = 1,000^{-1} \sum_{b=1}^{1,000} \hat{\xi}_b$, $I_A(\cdot)$ denotes the indicator function of the event A and $CI_b(\hat{\xi}) = \hat{\xi}_b \pm 1.96 \times \sqrt{\widehat{\text{Var}}_b(\xi)}$ is the 95% approximated confidence interval for ξ based on the asymptotic distribution in (16). Table 3 summarizes the results. For the ML estimators of σ , λ and γ , note that, as the sample size increases, the bias, SE and RMSE decrease. Also note that, as the sample size increases, the SE and RMSE are closer, suggesting that the standard errors of the ML estimators are well estimated.

The code necessary to generate random samples, as shown in Table 3, can be accessed in the repository at <https://github.com/isaaccortes1989/PTPN-SIMULATION> (accessed on 13 October 2024).

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

True Value			Estimator	$n = 150$				$n = 300$				$n = 600$				$n = 1000$			
σ	λ	γ		Bias	se	RMSE	cp	Bias	se	RMSE	cp	Bias	se	RMSE	cp	Bias	se	RMSE	cp
1	1	0.75	σ	0.0286	0.1800	0.1826	0.939	0.0089	0.1183	0.1190	0.956	0.0036	0.0818	0.0815	0.956	0.0042	0.0632	0.0638	0.958
			λ	−0.0579	0.5340	0.5337	0.975	−0.0102	0.3629	0.3518	0.970	0.0007	0.2532	0.2573	0.948	−0.0104	0.1958	0.2004	0.953
			γ	0.0239	0.1472	0.1504	0.956	0.0064	0.1014	0.0996	0.950	0.0010	0.0710	0.0726	0.951	0.0046	0.0552	0.0571	0.944
		2.5	σ	0.0502	0.2421	0.2377	0.970	0.0281	0.1537	0.1554	0.969	0.0141	0.1032	0.1081	0.950	0.0038	0.0774	0.0745	0.966
			λ	−0.0948	0.9525	0.9239	0.999	−0.0704	0.6152	0.6037	0.989	−0.0321	0.4195	0.4309	0.962	0.0002	0.3190	0.3029	0.964
			γ	0.3756	1.2994	1.4241	0.961	0.2042	0.8495	0.8743	0.949	0.0998	0.5749	0.6097	0.946	0.0236	0.4321	0.4154	0.960
	2	0.75	σ	0.0219	0.1503	0.1525	0.965	0.0118	0.1024	0.1032	0.966	0.0042	0.0708	0.0705	0.956	0.0069	0.0548	0.0541	0.971
			λ	−0.0323	0.6026	0.6037	0.964	−0.0252	0.4151	0.4106	0.960	−0.0051	0.2900	0.2894	0.960	−0.0222	0.2228	0.2206	0.964
			γ	0.0480	0.2288	0.2433	0.948	0.0289	0.1579	0.1628	0.951	0.0130	0.1093	0.1104	0.948	0.0138	0.0843	0.0844	0.964
		2.5	σ	0.0156	0.2653	0.2675	0.959	0.0176	0.1754	0.1779	0.969	0.0031	0.1185	0.1228	0.960	0.0062	0.0901	0.0933	0.951
			λ	0.2093	1.4855	1.5418	0.986	0.0351	0.9465	0.9859	0.969	0.0465	0.6366	0.6739	0.952	0.0000	0.4742	0.4909	0.950
			γ	0.6902	2.3674	2.6968	0.885	0.4527	1.5601	1.6894	0.923	0.1720	1.0093	1.0802	0.932	0.1447	0.7717	0.8226	0.942
2	1	0.75	σ	0.0294	0.3501	0.3578	0.920	0.0189	0.2374	0.2377	0.955	0.0109	0.1643	0.1655	0.956	0.0063	0.1260	0.1273	0.948
			λ	−0.0155	0.5295	0.5246	0.972	−0.0190	0.3640	0.3569	0.974	−0.0095	0.2537	0.2537	0.963	−0.0036	0.1955	0.1963	0.954
			γ	0.0176	0.1470	0.1454	0.958	0.0118	0.1023	0.1039	0.952	0.0042	0.0713	0.0726	0.956	0.0024	0.0551	0.0558	0.939
		2.5	σ	0.1054	0.4872	0.4886	0.974	0.0336	0.3021	0.3050	0.953	0.0336	0.2068	0.2082	0.960	0.0031	0.1545	0.1508	0.951
			λ	−0.1076	0.9432	0.9604	0.998	−0.0208	0.6154	0.6142	0.981	−0.0446	0.4186	0.4182	0.969	0.0076	0.3194	0.3127	0.957
			γ	0.3910	1.2938	1.4140	0.955	0.1495	0.8383	0.8844	0.948	0.1120	0.5761	0.5924	0.954	0.0201	0.4323	0.4296	0.953
	2	0.75	σ	0.0391	0.3002	0.3014	0.973	0.0031	0.2016	0.2029	0.964	0.0070	0.1412	0.1458	0.945	0.0056	0.1089	0.1058	0.961
			λ	−0.0273	0.6030	0.5990	0.972	0.0099	0.4154	0.4104	0.958	−0.0006	0.2893	0.2959	0.952	−0.0079	0.2230	0.2207	0.953
			γ	0.0504	0.2307	0.2438	0.953	0.0172	0.1559	0.1587	0.954	0.0086	0.1082	0.1129	0.930	0.0087	0.0839	0.0851	0.944
		2.5	σ	0.0110	0.5286	0.5444	0.947	0.0195	0.3494	0.3451	0.966	0.0282	0.2374	0.2367	0.971	0.0077	0.1805	0.1775	0.960
			λ	0.2868	1.5206	1.6366	0.988	0.0646	0.9505	0.9464	0.969	−0.0214	0.6219	0.6198	0.963	0.0160	0.4773	0.4771	0.967
			γ	0.6870	2.3807	2.8987	0.873	0.3922	1.5425	1.7309	0.921	0.2668	1.0383	1.1091	0.938	0.0987	0.7629	0.7658	0.951
3	1	0.75	σ	0.0946	0.5480	0.5919	0.949	0.0476	0.3600	0.3610	0.957	0.0134	0.2458	0.2363	0.962	0.0005	0.1882	0.1876	0.953
			λ	−0.0582	0.5379	0.5538	0.979	−0.0270	0.3647	0.3659	0.962	−0.0066	0.2536	0.2499	0.959	0.0006	0.1953	0.1988	0.957
			γ	0.0262	0.1479	0.1580	0.957	0.0100	0.1017	0.1020	0.943	0.0041	0.0713	0.0716	0.953	0.0028	0.0552	0.0550	0.958
		2.5	σ	0.1485	0.7217	0.8283	0.9686	0.0799	0.4594	0.4587	0.9696	0.0022	0.3023	0.2988	0.959	0.0053	0.2320	0.2272	0.958
			λ	−0.0796	0.9458	0.9616	0.996	−0.0620	0.6145	0.6020	0.987	0.0197	0.4179	0.4157	0.964	0.0059	0.3197	0.3166	0.961
			γ	0.3378	1.2724	1.4121	0.946	0.2008	0.8491	0.8831	0.961	0.0304	0.5631	0.5713	0.944	0.0268	0.4338	0.4382	0.946
	2	0.75	σ	0.0669	0.4529	0.4756	0.963	0.0134	0.3041	0.3089	0.953	0.0039	0.2115	0.2128	0.947	0.0015	0.1629	0.1578	0.959
			λ	−0.0327	0.6045	0.6150	0.973	0.0009	0.4157	0.4178	0.958	0.0106	0.2904	0.2941	0.953	0.0017	0.2235	0.2176	0.958
			γ	0.0566	0.2326	0.2603	0.955	0.0195	0.1561	0.1641	0.952	0.0060	0.1083	0.1109	0.932	0.0067	0.0839	0.0847	0.956
		2.5	σ	0.0930	0.8050	0.7983	0.969	0.0247	0.5230	0.5418	0.960	0.0180	0.3557	0.3549	0.972	0.0118	0.2707	0.2742	0.953
			λ	0.1132	1.4586	1.5120	0.990	0.0916	0.9565	1.0230	0.970	0.0282	0.6320	0.6346	0.961	0.0123	0.4767	0.4877	0.955
			γ	0.9086	2.5240	3.0342	0.910	0.3755	1.5276	1.6826	0.914	0.1797	1.0118	1.0667	0.947	0.1253	0.7706	0.8070	0.936

6. Applications

In this section, two applications will be presented to illustrate the performance of the PTPN distribution compared to other distributions. An application without covariates and an application with covariates will be presented.

6.1. Application 1

The first data set is from the National Health and Nutrition Examination Surveys (NHANES) from 2017–2020, <https://www.cdc.gov/nchs/nhanes/index.htm> (accessed on 13 October 2024). Table 4 summarizes the variable refrigerated serum iron (Iron) measured in $\mu\text{Mol/L}$. The importance of this variable lies in its use for monitoring various health conditions, such as anemia or excess iron. Note that Iron has a high kurtosis value.

Table 4. Mean, median, standard deviation (SD), skewness and kurtosis of the Iron variable.

n	Mean	Median	SD	Skewness	Kurtosis
9473	15.389	14.700	6.470	1.157	7.598

The proposed PTPN model is compared with some distributions from the literature, specifically, with the TPN and PHN distributions from the literature.

The pdf of the PHN distribution is given by

$$f(z; \sigma, \alpha) = \frac{2\alpha}{\sigma} \phi\left(\frac{z}{\sigma}\right) \left(2\Phi\left(\frac{z}{\sigma}\right) - 1\right)^{\alpha-1}, \quad z, \sigma, \alpha \in \mathbb{R}^+.$$

We fitted the PTPN, TPN and PHN models. The initial values used to obtain the ML estimators were the modified moment estimators discussed in Section 3.1. Considering $\hat{\gamma}_M = 2$, we obtained $\hat{\sigma}_M = 7.23$ and $\hat{\lambda}_M = 1.44$. The results are presented in Table 5. Note that the AIC [43] and BIC [44] criteria are the smallest values for the PTPN model among the considered models, suggesting that this distribution is more appropriate for these data. Additionally, the histogram and estimated pdf for the different models are presented in Figure 6, whereas QQ-plots [45] are presented in Figure 7, confirming that the PTPN model is preferable for this data set.

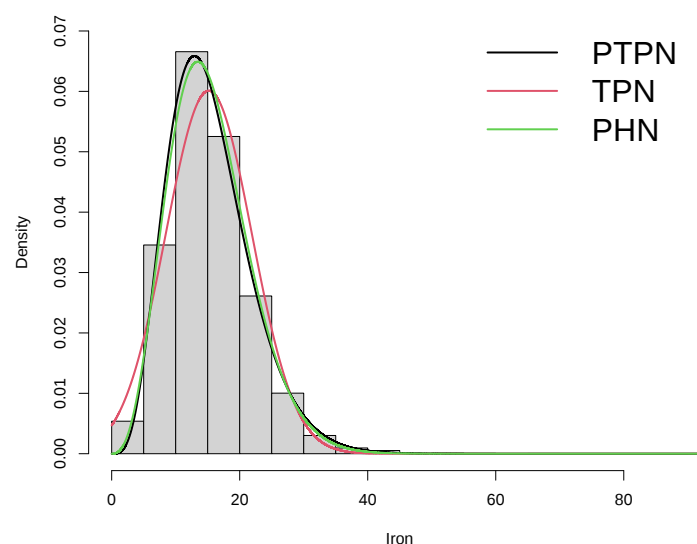
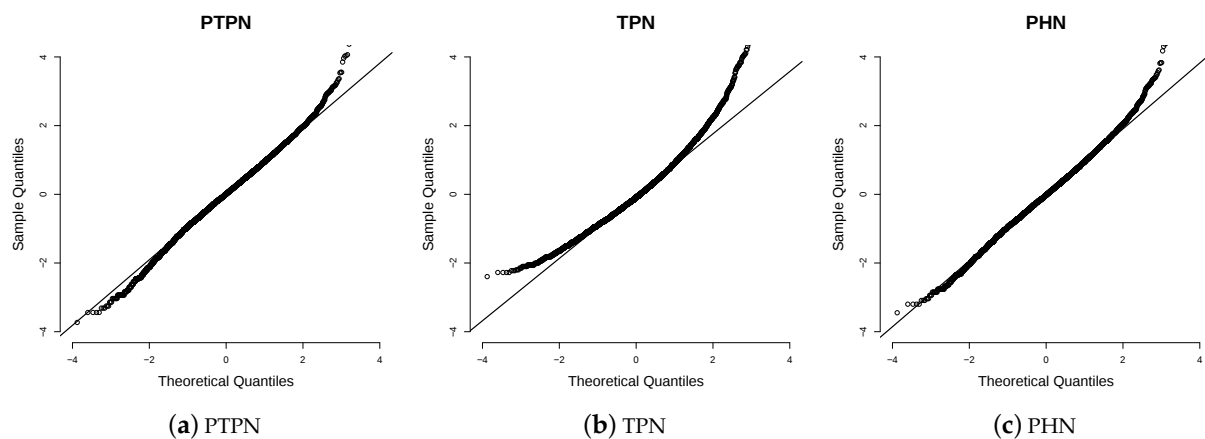


Figure 6. Histogram of the variable Iron and the estimated pdf: PTPN (black line), TPN (red line) and PHN (green line).

Table 5. Maximum likelihood estimates of the parameters with their respective standard errors (in parentheses), and the values of the information criteria.

Parameter	PTPN	TPN	PHN
$\hat{\lambda}$	−0.821 (0.147)	2.260 (0.023)	–
$\hat{\sigma}$	13.464 (0.511)	6.716 (0.056)	10.926 (0.068)
$\hat{\gamma}$	4.666 (0.196)	–	3.579 (0.055)
Log-likelihood	−30,572.79	−31,033.3	−30,594.13
AIC	61,151.58	62,070.61	61,192.25
BIC	61,173.05	62,084.92	61,206.56

**Figure 7.** QQ-plots of (a) PTPN, (b) TPN and (c) PHN models, in Iron data.

6.2. Application 2

The second application involves the study of the quantiles of serum immunoglobulin G (IgG) concentrations in 298 children aged between 6 months and 6 years. The IgG variable is an important measure in immunology as it is one of the most abundant antibodies in the human immune system and plays a fundamental role in defending against infections. This data set has been previously analyzed by [37,46,47].

We assume that $\text{IgG}_i \sim \text{RPTPN}(\tau_i(p), \lambda(p), \gamma(p))$ with the following structure:

$$\log \tau_i(p) = \beta_1(p) + \beta_2(p) \times \text{Age} \quad \text{and} \quad \log \gamma(p) = \nu_1(p), \quad (18)$$

where $\beta_1(p)$ and $\beta_2(p)$ are the regression parameters. Specifically, we focus on $p = 0.3$ because the 30% with lower IgG are more prone to suffering from various diseases. For comparative purposes, we also include the estimates for the SKD (see [48]) and RGTG (see [35]) regression models. The comparison of the models is performed using the AIC and BIC. Table 6 presents the ML estimates and their respective standard errors for the three models. Note that, in the three models, all estimates are significant and the smallest criterion values are those of the RPTPN model. This result indicates that the RPTPN model is the most suitable for fitting the IgG data. On the other hand, as for the PTPN model $\hat{\beta}_1(p) = 1.042$, the estimated 30-th quantile of IgG for a newborn is $\exp(1.042) \approx 2.8$. In addition, as $\hat{\beta}_2(p) = 0.129$, for each year of age the 30-th quantile of IgG is increased 14% (because $\exp(0.129) \approx 1.14$).

Finally, we compute the quantile residuals (QRs, see [49]) along with their respective envelopes, as well as the likelihood displacement (LD) and generalized Cook's distance (GCD) measures. The purpose of the residuals is to detect outliers and evaluate the suitability of the RPTPN model. If the model were appropriated, the residuals should approximately follow a standard normal distribution. In addition, the LD and GCD measures are presented in Figure 8.

Figure 8a shows that the residuals lie within the bounds, suggesting that there are no outliers and that the RPTN model is suitable for fitting the IgG data. Additionally, the Kolmogorov–Smirnov test was performed to verify the normality assumption, and the null hypothesis was not rejected at the 5% significance level. Finally, Figure 8b,c indicate that observations #20, #25, #94 and #180 are potentially influential in the fit of the RPTN model.

Table 6. ML estimates and their respective standard errors, along with the information criterion values for each of the fitted models, considering $p = 0.3$.

Model	$\hat{\beta}_1(p)$	$\hat{\beta}_2(p)$	$\hat{\lambda}(p)$	$\hat{v}_1(p)$	$\hat{\sigma}(p)$	AIC	BIC
RPTN	1.042 (0.046)	0.129 (0.013)	0.529 (0.248)	1.231 (0.124)	-	1228.243	1243.032
SKD	2.484 (0.414)	0.564 (0.110)	-	3.975	1.399	1240.918	1252.001
RGTG	1.054 (0.048)	0.128 (0.013)	1.633 (0.318)	0.444 (0.156)	-	1231.515	1246.304

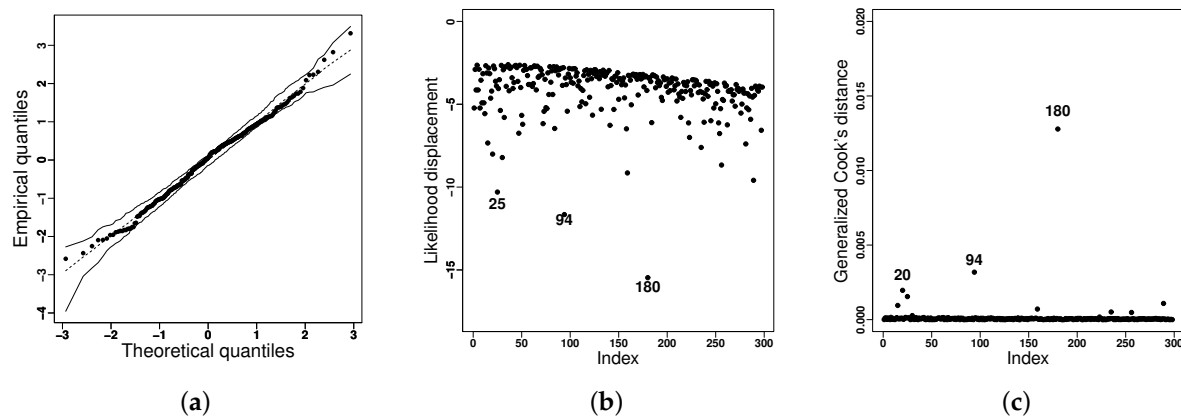


Figure 8. Quantile residuals (a), likelihood displacement (b) and generalized Cook's distance (c) resulting from fitting the RPTN model at $p = 0.3$.

7. Conclusions

In this paper, we introduce a new distribution called the power truncated positive normal. This model incorporates a new shape parameter, providing more flexibility to accommodate various data sets. Based on the TPN distribution, it inherits many of its properties. One of its key features is that its density, distribution and quantile functions can be expressed in closed form, making it easier to generate random numbers and conduct simulations with this new distribution. Our application of the model showed that applying the power to the TPN model resulted in greater flexibility. Additionally, the model was parameterized in terms of the p -th quantile, and its effectiveness was verified against other regression models. Future extensions for this model can involve random effects and measurement error-in-variable, to name a couple.

Author Contributions: Conceptualization, H.J.G., K.I.S. and I.E.C.; formal analysis, K.I.S., T.M.M. and D.I.G.; funding acquisition, H.J.G.; investigation, H.J.G., K.I.S. and I.E.C.; methodology, H.J.G., K.I.S. and D.I.G.; software, I.E.C.; writing—original draft, H.J.G., K.I.S., D.I.G. and T.M.M.; writing—review and editing, D.I.G. and T.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data are available for Application 1 (<https://www.cdc.gov/nchs/nhanes/index.htm> (accessed on 13 October 2024)).

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

References

1. Fryar, C.D.; Kit, B.; Carroll, M.D.; Afful, J.; Kuo, T. Hypertension Prevalence, Treatment, and Control Among Adults: Los Angeles County and the United States, 2015–2018. National Center for Health Statistics. 2023. Available online: <https://www.cdc.gov/nchs/data/hestat/hypertension-15-18/hypertension-15-18.htm> (accessed on 13 October 2024).
2. Fryar, C.D.; Carroll, M.D.; Afful, J. Prevalence of Overweight, Obesity, and Severe Obesity Among Children and Adolescents Aged 2–19 Years: United States, 1963–1965 Through 2017–2018. National Center for Health Statistics. 2020. Available online: <https://www.cdc.gov/nchs/data/hestat/obesity-child-17-18/obesity-child.htm#table1> (accessed on 13 October 2024).
3. Steinberg, F.M. Advancing the Use of Evidence-Based Practice in Nutrition and Dietetics. *J. Nutr.* **2024**, *154*, 1065–1066. [CrossRef] [PubMed]
4. do Vale, M.R.L.; Johnsen, J.T.; Laur, C.; Lepre, B.; Ray, S. Advancing research, policy and practice to promote resilient and sustainable food and health systems in the year of action on nutrition: Proceedings of the 7th annual International Summit on Nutrition and Health. *BMJ Nutr. Prev. Health* **2022**, *5*. [CrossRef]
5. Livingstone, K.M.; Ramos-López, O.; Pérusse, L.; Kato, H.; Ordovás, J.M.; Martínez, J.A. Precision nutrition: A review of current approaches and future endeavors. *Trends Food Sci. Technol.* **2022**, *128*, 253–264. [CrossRef]
6. Yoo, H. Modeling clustered count data with discrete weibull regression model. *Commun. Stat. Appl. Methods* **2022**, *29*, 413–420. [CrossRef]
7. Parker, P.A.; Holan, S.H. A Bayesian functional data model for surveys collected under informative sampling with application to mortality estimation using NHANES. *Biometrics* **2023**, *79*, 1397–1408. [CrossRef]
8. Gómez, H.J.; Olmos, N.M.; Varela, H.; Bolfarine, H. Inference for a truncated positive normal distribution. *Appl. Math. J. Chin. Univ.* **2018**, *33*, 163–176. [CrossRef]
9. Salinas, H.S.; Bakouch, H.S.; Almuhaifith, F.E.; Caimanque, W.E.; Barrios-Blanco, L.; Albalawi, O. Statistical Advancement of a Flexible Unitary Distribution and Its Applications. *Axioms* **2024**, *13*, 397. [CrossRef]
10. Gómez, H.J.; Gallardo, D.I.; Venegas, O. Generalized Truncation Positive Normal Distribution. *Symmetry* **2019**, *11*, 1361. [CrossRef]
11. Cooray, K.; Ananda, M.M. A generalization of the half-normal distribution with applications to lifetime data. *Commun. Stat. Theory Methods* **2008**, *10*, 195–224. [CrossRef]
12. Gómez, H.J.; Gallardo, D.I.; Santoro, K.I. Slash Truncation Positive Normal Distribution and Its Estimation Based on the EM Algorithm. *Symmetry* **2020**, *13*, 2164. [CrossRef]
13. Gómez, H.J.; Caimanque, W.E.; Gómez, Y.M.; Magalhães, T.M.; Concha, M.; Gallardo, D.I. A Bimodal Model Based on Truncation Positive Normal with Application to Height Data. *Symmetry* **2020**, *14*, 665. [CrossRef]
14. Gómez, H.J.; Santoro, K.I.; Barranco-Chamorro, I.; Venegas, O.; Gallardo, D.I.; Gómez, H.W. A Family of Truncated Positive Distributions. *Mathematics* **2023**, *11*, 4431. [CrossRef]
15. Gallardo, D.I.; Gómez, H.J.; Gómez, Y.M. tpn: Truncated Positive Normal Model and Extensions. R Package Version 1.6. Available online: <https://CRAN.R-project.org/package=tpn> (accessed on 13 October 2024).
16. Lehmann, E. The power of rank tests. *Ann. Math. Stat. Stat.* **1953**, *24*, 23–43. [CrossRef]
17. Durrans, R. Distributions of fractional order statistics in hydrology. *Water Resour. Res.* **1992**, *28*, 1649–1655. [CrossRef]
18. Gupta, D.; Gupta, R.C. Analyzing skewed data by power normal model. *Test* **2008**, *17*, 197–210. [CrossRef]
19. Segovia, F.A.; Gomez, Y.M.; Gallardo, D.I. Exponentiated power Maxwell distribution with quantile regression and applications. *Sort-Stat. Oper. Res. Trans.* **2021**, *45*, 181–200.
20. Martínez-Florez, G.; Gallardo, D.I.; Venegas, O.; Bolfarine, H.; Gomez, H.W. Flexible Power-Normal Models with Applications. *Mathematics* **2021**, *9*, 3183. [CrossRef]
21. Tovar-Falon, R.; Martínez-Flórez, G. A New Class of Exponentiated Beta-Skew-Laplace Distribution. *An. Acad. Bras. Cienc.* **2022**, *94*, e20191597. [CrossRef]
22. Ferrari, S.L.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
23. Ospina, R.; Ferrari, S.L.P. Inflated beta distributions. *Stat. Pap.* **2008**, *51*, 111–126. [CrossRef]
24. Bayes, C.L.; Bazán, J.L.; García, C. A new robust regression model for proportions. *Bayesian Anal.* **2012**, *7*, 841–866. [CrossRef]
25. Migliorati, S.; Di Brisco, A.M.; Ongaro, A. A New Regression Model for Bounded Responses. *Bayesian Anal.* **2018**, *13*, 845–872. [CrossRef]
26. Pereira, G.H.A.; Botter, D.A.; Sandoval, M.C. A regression model for special proportions. *Stat. Model.* **2013**, *13*, 125–151. [CrossRef]
27. Chahuan-Jimenez, K.; Rubilar, R.; de la Fuente-Mella, H.; Leiva, V. Breakpoint analysis for the COVID-19 pandemic and its effect on the stock markets. *Entropy* **2021**, *32*, 100. [CrossRef]
28. Koenker, R.; Basset, G. Cuantiles de regresión. *Econometría* **1978**, *46*, 33–50. [CrossRef]
29. Lemonte, A.J.; Moreno-Arenas, G. On a heavy-tailed parametric quantile regression model for limited range response variables. *Comput. Stat.* **2020**, *35*, 379–398. [CrossRef]
30. Korkmaz, M.Ç.; Chesneau, C.; Korkmaz, Z.S. On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications. *Symmetry* **2021**, *13*, 117. [CrossRef]
31. He, X.M.; Pan, X.O.; Tan, K.M.; Zhou, W.X. Smoothed quantile regression with large-scale inference. *J. Econom.* **2023**, *232*, 367–388. [CrossRef]

32. Cordeiro, G.M.; Rodrigues, G.M.; Prata, F.; Ortega, E.M.M. A new quantile regression model with application to human development index. *Comput. Stat.* **2024**, *39*, 2925–2948. [\[CrossRef\]](#)
33. Alfò, M.; Salvati, N.; Ranalli, M.G. Finite mixtures of quantile and M-quantile regression models. *Stat. Comput.* **2024**, *27*, 547–570. [\[CrossRef\]](#)
34. Peng, L.M. Quantile Regression for Survival Data. *Annu. Rev. Stat. Its Appl.* **2021**, *8*, 413–437. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Gómez, H.J.; Santoro, K.I.; Ayma, D.; Cortés, I.E.; Gallardo, D.I.; Magalhães, T.M. A New Generalization of the Truncated Gumbel Distribution with Quantile Regression and Applications. *Mathematics* **2024**, *12*, 1762. [\[CrossRef\]](#)
36. Cortés, I.E.; de Castro, M.; Gallardo, D.I. A new family of quantile regression models applied to nutritional data. *J. Appl. Stat.* **2023**, *51*, 1378–1398. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Isaacs, D.; Altman, D.G.; Tidmarsh, C.E.; Valman, H.B.; Webster, A.D. Serum immunoglobulin concentrations in preschool children measured by laser nephelometry: Reference ranges for IgG, IgA, IgM. *J. Clin. Pathol.* **1983**, *36*, 1193–1196. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Gómez, Y.M.; Bolfarine, H. Likelihood-based inference for the power half-normal distribution. *J. Stat. Theory Appl.* **2015**, *14*, 383–398. [\[CrossRef\]](#)
39. Balakrishnan, N.; Cohen, C.A. Order Statistics and Inference: Estimation Methods. In *Statistical Modeling and Decision Science*; Elsevier Science: Amsterdam, The Netherlands, 1991.
40. 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 March 2024).
41. MacDonald, I.L. Does Newton-Raphson really fail? *Stat. Methods Med Res.* **2014**, *23*, 308–311. [\[CrossRef\]](#)
42. Kroese, D.P.; Brereton, T.; Taimre, T.; Botev, Z.I. Why the Monte Carlo method is so important today. *WIREs Comput. Stat.* **2014**, *6*, 386–392. [\[CrossRef\]](#)
43. Akaike, H. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*; Petrov, B.N., Csáki, F., Eds.; Akadémiai Kiadó: Budapest, Hungary, 1973; pp. 267–281.
44. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [\[CrossRef\]](#)
45. Wilk, M.B.; Gnanadesikan, R. Probability plotting methods for the analysis of data. *Biometrika* **1968**, *55*, 1–17. [\[CrossRef\]](#)
46. Noufaily, A.; Jones, M.C. Parametric quantile regression based on the generalized gamma distribution. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2013**, *62*, 723–740. [\[CrossRef\]](#)
47. Royston, P.; Wright, E.M. A method for estimating age-specific reference intervals (‘normal ranges’) based on fractional polynomials and exponential transformation. *J. R. Stat. Soc. Ser. A Stat. Soc.* **1998**, *161*, 79–101. [\[CrossRef\]](#)
48. Galarza Morales, C.; Lachos, D.V.; Barbosa, C.C.; Castro, C.L. Robust quantile regression using a generalized class of skewed distributions. *Stat* **2017**, *6*, 113–130. [\[CrossRef\]](#)
49. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.