

Article

A New Generalization of the Truncated Gumbel Distribution with Quantile Regression and Applications

Héctor J. Gómez ^{1,*}, Karol I. Santoro ², Diego Ayma ³, Isaac E. Cortés ⁴, Diego I. Gallardo ⁵
and Tiago M. Magalhães ⁶

- ¹ Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile
- ² Departamento de Estadística y Ciencia de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; karol.santoro@uantof.cl
- ³ Departamento de Matemáticas, Facultad de Ciencias, Universidad Católica del Norte, Antofagasta 1240000, Chile; diego.ayma@ce.ucn.cl
- ⁴ Facultad de Ciencias, Universidad Arturo Prat, Avenida Arturo Prat 2120, Iquique 1110939, Chile; isacortes@unap.cl
- ⁵ Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile; dgallardo@ubiobio.cl
- ⁶ Department of Statistics, Institute of Exact Sciences, Federal University of Juiz de Fora, Juiz de Fora 36036-900, MG, Brazil; tiago.magalhaes@ufjf.edu.br
- * Correspondence: hgomez@uct.cl

Abstract: In this article, we introduce a new model with positive support. This model is an extension of the truncated Gumbel distribution, where a shape parameter is incorporated that provides greater flexibility to the new model. The model is parameterized in terms of the p -th quantile of the distribution to perform quantile regression in this model. An extensive simulation study demonstrates the good performance of the maximum likelihood estimators in finite samples. Finally, two applications to real datasets related to the level of beta-carotene and body mass index are presented.



Citation: 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. <https://doi.org/10.3390/math12111762>

Academic Editors: Jinfeng Xu and Elvira Di Nardo

Received: 8 April 2024
Revised: 20 May 2024
Accepted: 29 May 2024
Published: 5 June 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/>).

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

MSC: 62E10; 62E15; 62E17

1. Introduction

The Gumbel distribution, also known as the type-I generalized extreme value distribution, is commonly used to model data with extreme observations. This distribution and its extensions have a wide range of applications in several disciplines such as hydrology, economics, finance, climatology and seismology. The probability density function (pdf), the cumulative distribution function (cdf) and the quantile function of a random variable X that follows the Gumbel distribution are given by

$$g(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\left(\frac{x-\mu}{\sigma}\right) - \exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right), \quad x \in \mathbb{R},$$
$$G(x; \mu, \sigma) = \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right), \quad x \in \mathbb{R},$$
$$Q(p) = \mu - \sigma \log(-\log(p)), \quad p \in (0, 1),$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. Applications of this model in different scenarios can be found in Bhaskaran et al. [1], Gurung et al. [2], Purohit et al. [3], Li et al. [4] and Kang et al. [5].

Several extensions of the Gumbel distribution have been recently proposed in the literature. Hossam et al. [6] presented a statistical model that combines the new alpha power transformation method and Gumbel distribution. Watthanawisut and Bodhisuwan [7] proposed a new extension of the so called Topp-Leone Gumbel distribution that is used to model minimum flow data. Fayomi et al. [8] presented the exponentiated Gumbel-G family of distributions and explored a special case called EGuNH. Nagode et al. [9] introduced a three-parameter Gumbel distribution, which was applied to rope failure data. Oseni and Okasha [10] derived the Gumbel-geometric distribution, which was applied to precipitation and maximum annual wind speed data. Note that all these extensions do not consider a regression framework, and their main objectives rely on the fit of univariate data.

It is evident that regression models have become relevant tools in the era of Data Science. Among them, the so-called quantile regression models (introduced by Koenker and Bassett [11]) are an alternative to the usual regression techniques where the mean response conditional to values of covariates (or explanatory variables) is estimated. The quantile regression models allow us to measure the effects of covariates at different quantiles of the response variable distribution. Thus, they provide an analysis across the entire conditional distribution, as can be seen in Cade et al. [12], Koenker [13] and Wei et al. [14]. The mean, as the only summary measure, is generally quite poor for assessing risk, as it is greatly affected by the presence of outlier observations. Outlier data can be quite strange, but at the same time, these can be enough to cause serious problems when analyzing the information obtained; see, for example, Gómez-Déniz et al. [15], who analyzed extreme values in insurance companies. To our knowledge, there are no studies on quantile regression models based on the Gumbel distribution. Thus, the objectives of this work were to introduce a new generalization of the truncated Gumbel distribution and then establish a quantile regression model based on this novel generalization. To do this, a reparametrization was obtained of the new truncated Gumbel generalization by incorporating a parameter that represents the quantile. We should note that the proposed generalization was achieved by considering the work of Neamah and Qasim [16] and the transformation provided by Cooray and Ananda [17]. The latter authors developed an extension of the half-normal (HN) distribution through the relation $Y = \beta X^{1/\alpha}$, where $X \sim \text{HN}(1)$.

The rest of the paper is organized as follows. In Section 2, we introduce our proposal, the generalized truncated Gumbel (GTG) distribution, and several important properties of this new model are presented. In Section 3, inference is performed, including some initial points to obtain maximum likelihood (ML) estimators and present the observed Fisher information matrix for the proposed model. In Section 4, the reparametrized model in terms of a quantile is presented. In Section 5, the simulation study carried out to analyze the performances of the ML estimators in finite samples for the proposed model without and with covariates is discussed. In Section 6, two real-data applications are presented to illustrate the proposed models, without and with covariates. Finally, in Section 7, some concluding comments are presented.

2. Generalized Truncated Gumbel Distribution

Neamah and Qasim [16] derived a new model with positive support for the Gumbel distribution by truncating its pdf from the left. We will refer to the resulting model of these authors as the truncated Gumbel (TG) distribution, which is defined in the interval $(0, \infty)$. In considering the reparametrization $\lambda = \mu/\sigma$, the pdf of the TG distribution can be written as follows:

$$f(y; \beta, \lambda, \alpha) = \frac{1}{\beta(1 - G(-\lambda))} g\left(\frac{y}{\beta} - \lambda\right), \quad y > 0, \quad (1)$$

where $\beta > 0$ is a scale parameter, $\lambda \in \mathbb{R}$ is a shape parameter, and $g(u) = \exp(-u - \exp(-u))$ and $G(u) = \exp(-\exp(-u))$ are the pdf and cdf for the standard Gumbel distribution, respectively.

In this work, we considered the transformation developed by Cooray and Ananda [17] to extend the TG distribution. That is, we considered the transformation $Z = \beta Y^{1/\alpha}$, where $Y \sim TG(1, \lambda)$. We will refer to this extension as the generalized truncated Gumbel (GTG) distribution. Important functions, such as the pdf, cdf, hazard and quantile functions of the GTG distribution are provided below.

2.1. Pdf, cdf and Hazard Function

Proposition 1. Let $Z \sim GTG(\beta, \lambda, \alpha)$. Then, the pdf of Z is given by

$$f(z; \beta, \lambda, \alpha) = \frac{\alpha z^{\alpha-1}}{\beta^\alpha (1 - G(-\lambda))} g\left(\left(\frac{z}{\beta}\right)^\alpha - \lambda\right), \quad z > 0, \tag{2}$$

where $\beta > 0, \lambda \in \mathbb{R}$ and $\alpha > 0$.

Proof. Considering the stochastic representation of Z , we have that $z = h(y) = \beta y^{1/\alpha}$. Then,

$$f_Z(z) = f_Y(h^{-1}(z)) \left| \frac{dh^{-1}(z)}{dz} \right| = \frac{1}{1 - G(-\lambda)} g(h^{-1}(z) - \lambda) \frac{\alpha z^{\alpha-1}}{\beta^\alpha}.$$

Therefore, the result is obtained by replacing $h^{-1}(z) = \left(\frac{z}{\beta}\right)^\alpha$ in $g(\cdot)$. \square

Remark 1. We previously mentioned that $Z \sim GTG(\beta, \lambda, \alpha)$ if $Z = \beta Y^{1/\alpha}$, where $Y \sim TG(1, \lambda)$. Thus, when $\alpha = 1$, we obtain the TG distribution; that is, $GTG(\beta, \lambda, 1) \equiv TG(\beta, \lambda)$.

Proposition 2. Let $Z \sim GTG(\beta, \lambda, \alpha)$. Then, the cdf and hazard function of Z are given by

$$F(z; \beta, \lambda, \alpha) = \frac{G\left(\left(\frac{z}{\beta}\right)^\alpha - \lambda\right) - G(-\lambda)}{1 - G(-\lambda)} \tag{3}$$

and

$$h(z; \beta, \lambda, \alpha) = \frac{\alpha z^{\alpha-1} g\left(\left(\frac{z}{\beta}\right)^\alpha - \lambda\right)}{\beta^\alpha \left[1 - G\left(\left(\frac{z}{\beta}\right)^\alpha - \lambda\right)\right]}, \tag{4}$$

respectively, for all $z > 0$.

Proof. Both functions are obtained immediately from their definitions. \square

Figure 1 shows the pdf, cdf and hazard function for the $GTG(1, \lambda, \alpha)$ model, considering some combinations for λ and α . We observe that the GTG model can have decreasing or unimodal shapes for the pdf, whereas for the hazard function, we can have decreasing or increasing shapes. Also, we observe that for some combinations of λ and α , the cdf rapidly increases, although all of them tend to 1 when z increases.

2.2. Mode

The shape of the pdf of $Z \sim GTG(\beta, \lambda, \alpha)$ can be examined based on its inflection points. By computing the first derivative of $\log(f(z))$ with respect to z , where $f(z)$ is the pdf for the GTG model, we obtain that

$$\frac{\partial \log(f(z))}{\partial z} = \frac{\alpha - 1}{z} - \frac{\alpha z^{\alpha-1}}{\beta^\alpha} [1 - \exp(-v)],$$

where $v = \left(\frac{z}{\beta}\right)^\alpha - \lambda$. By equating the previous expression to 0, we obtain that

$$\frac{\beta^\alpha(\alpha - 1)}{\alpha} = z^\alpha(1 - \exp(-v)), \tag{5}$$

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

$$u(z) = -\frac{\alpha - 1}{z^2} - \frac{(1 - \exp(-v))\alpha(\alpha - 1)z^{\alpha-2}}{\beta^\alpha} - \frac{\exp(-v)\alpha^2 z^{2(\alpha-1)}}{\beta^{2\alpha}}.$$

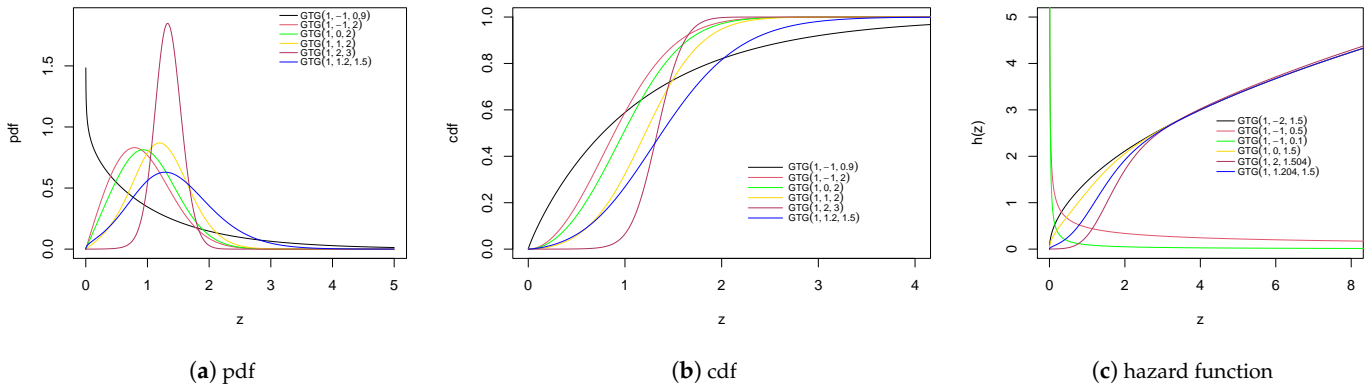


Figure 1. Pdf, cdf and hazard function for the $GTG(\beta = 1, \lambda, \alpha)$ model with different value combinations for λ and α .

Depending on whether $u(z_0) < 0$ or $u(z_0) > 0$, where $z = z_0$ is a solution of Equation (5), the inflection points can be local maxima or minima. Figure 2 shows the shape of $u(z)$ for $\beta = 1$ and selected values of λ and α . From here, we observe that the pdf of the GTG distribution is zero when $z \rightarrow \infty$, both for when λ takes a positive or negative value.

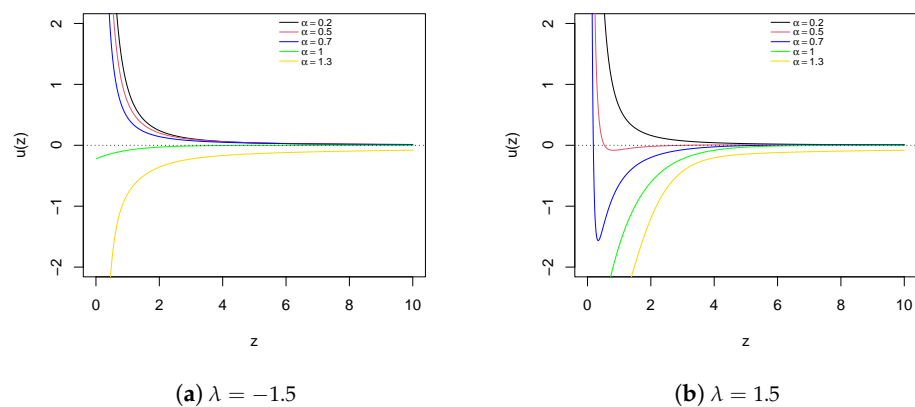


Figure 2. Shape of $u(z)$ for $\beta = 1$ and some value selections of α and λ .

2.3. Quantiles

Proposition 3. If $Z \sim GTG(\beta, \lambda, \alpha)$, then the quantile function of Z is given by

$$Q(p) = \beta \left(\lambda + G^{-1}[p + (1 - p)G(-\lambda) + G(-\lambda)] \right)^{\frac{1}{\alpha}}, \quad 0 < p < 1. \tag{6}$$

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

Corollary 1. The quartiles of the GTG distribution are as follows:

1. (First quartile) $Q(0.25) = \beta \left(\lambda + G^{-1}[0.25 + 0.75G(-\lambda)] \right)^{\frac{1}{\alpha}}$.
2. (Median) $Q(0.5) = \beta \left(\lambda + G^{-1}[0.5(1 + G(-\lambda))] \right)^{\frac{1}{\alpha}}$.
3. (Third quartile) $Q(0.75) = \beta \left(\lambda + G^{-1}[0.75 + 0.25G(-\lambda)] \right)^{\frac{1}{\alpha}}$.

Proof. It is immediate from Proposition 3. \square

2.4. Moments

Proposition 4. Let $Z \sim GTG(\beta, \lambda, \alpha)$ and n be a positive integer. Then, the n -th moment of Z is given by

$$\mu_n = E[Z^n] = \frac{\beta^n}{1 - G(-\lambda)} b_n(\lambda, \alpha), \tag{7}$$

where $b_n(\lambda, \alpha) = \sum_{k=0}^{\infty} (-1)^k \binom{n/\alpha}{k} \lambda^{n/\alpha - k} \gamma_k$, $\gamma_k = \int_0^{\lambda} (\ln(t))^k e^{-t} dt$, and $\binom{n/\alpha}{k} = \frac{1}{k!} \prod_{i=0}^{k-1} (n/\alpha - i)$ is the generalized binomial coefficient. When $n/\alpha \in \mathbb{N}$, the sum in $b_n(\lambda, \alpha)$ stops at n/α .

Proof. Given the stochastic representation of Z , it is immediate that $E[Z^n] = \beta^n E[Y^{n/\alpha}]$, where $Y \sim TG(1, \lambda)$. Then, the (n/α) moment of Y can be computed by following the properties presented in Neamah and Qasim [16]. \square

Corollary 2. If $Z \sim GTG(\beta, \lambda, \alpha)$, then the first four moments and the variance of Z are obtained as follows:

1. $\mu_1 = E(Z) = \frac{\beta}{1 - G(-\lambda)} b_1(\lambda, \alpha);$
2. $\mu_2 = E(Z^2) = \frac{\beta^2}{1 - G(-\lambda)} b_2(\lambda, \alpha);$
3. $\mu_3 = E(Z^3) = \frac{\beta^3}{1 - G(-\lambda)} b_3(\lambda, \alpha);$
4. $\mu_4 = E(Z^4) = \frac{\beta^4}{1 - G(-\lambda)} b_4(\lambda, \alpha);$
5. $Var(Z) = \frac{\beta^2}{(1 - G(-\lambda))^2} [(1 - G(-\lambda)) b_2(\lambda, \alpha) - b_1^2(\lambda, \alpha)].$

Proof. It is immediate from Proposition 4. \square

Corollary 3. Let $Z \sim GTG(\beta, \lambda, \alpha)$. Then, the skewness coefficient ($\sqrt{\beta_1}$) and the kurtosis coefficient (β_2) are given by

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

where $\overline{G}^k = (1 - G(-\lambda))^k$, and $b_i = b_i(\lambda, \alpha)$, for $i = 1, 2, 3, 4$.

Proof. The expressions above are obtained using the definitions of the skewness and kurtosis coefficients; that is,

$$\sqrt{\beta_1} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}} \quad \text{and} \quad \beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}.$$

where $\mu_n = E[X^n]$, for $n = 1, 2, 3, 4$, are given in Corollary 2. \square

Remark 2. Proposition 4 shows that the moments of the GTG distribution basically depend on the moments of the $TG(1, \lambda)$ model. Plots for the expected value, variance, skewness and kurtosis coefficients of the $GTG(1, \lambda, \alpha)$ model are given in Figure 3 for different values of the λ and α parameters. The bottom plots in Figure 3 reflect the effect of the α parameter: a lower value of α

produces higher values of the skewness and kurtosis coefficients. This fact can also be appreciated in Tables 1 and 2.

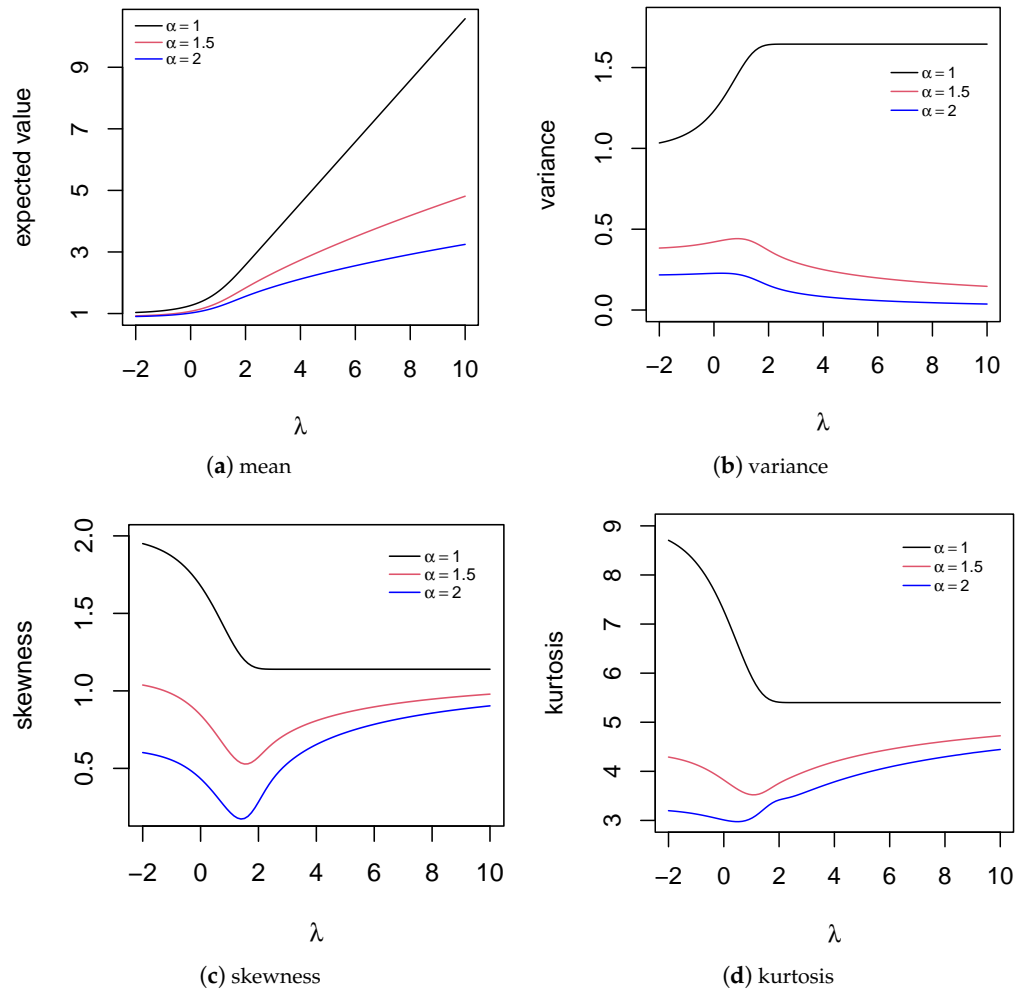


Figure 3. Plots of the (a) mean, (b) variance, (c) skewness and (d) kurtosis coefficients for GTG ($\beta = 1, \lambda, \alpha$) for $\alpha \in \{1, 1.5, 2\}$ as a function of $\lambda \in \mathbb{R}$.

Table 1. Skewness coefficient of the GTG($\beta = 1, \lambda, \alpha$) model for different values of λ and α .

λ	α				
	0.5	1	1.5	2	2.5
-2	6.4487	1.9502	1.0377	0.1428	-0.3964
-1.5	6.3433	1.9190	1.0161	0.1265	-0.4114
0	5.5570	1.6782	0.8441	-0.0092	-0.5421
1.5	3.9215	1.2019	0.5286	-0.2238	-0.7331
2	3.4630	1.1457	0.5728	-0.0296	-0.3987
2.5	3.1208	1.1396	0.6610	0.2027	-0.0336

Table 2. Kurtosis coefficient of the GTG($\beta = 1, \lambda, \alpha$) model for different values of λ and α .

λ	α				
	0.5	1	1.5	2	2.5
-2	83.4229	8.7063	4.2903	2.7226	3.0664
-1.5	80.8181	8.5270	4.2297	2.7201	3.0884
0	62.8258	7.2726	3.8231	2.7589	3.3450
1.5	33.3157	5.5217	3.5897	3.4425	4.6563

Table 2. Cont.

λ	0.5	1	α 1.5	2	2.5
2	26.8341	5.4081	3.7574	3.4504	4.1342
2.5	22.4625	5.4001	3.8929	3.2871	3.3145

2.5. Bonferroni Curves

In different disciplines, such as socio-economics and public health sciences, there is a necessity to compare and analyze the inequality of non-negative distributions. Generally, Bonferroni curves are used as graphical methods to achieve the required comparison/analysis (see Bonferroni [18], and Arcagni and Porro [19] for a further discussion about these curves). The following result shows the expressions of these curves for the GTG model.

Proposition 5. If $Z \sim GTG(\beta, \lambda, \alpha)$, then the Bonferroni curves, say $B(p)$, are given by

$$B(p) = \frac{\beta}{\mu_1 p [1 - G(-\lambda)]} \sum_{k=0}^{\infty} (-1)^k \binom{1/\alpha}{k} \lambda^{1/\alpha - k} \gamma_k^k, \quad 0 < p < 1,$$

where $\mu_1 = E(Z)$, $\gamma_k = \int_v^{e^\lambda} (\ln(t))^k e^{-t} dt$, and $v = \exp(-(q/\beta)^\alpha + \lambda)$.

Proof. The expression above is obtained using the definition of the Bonferroni curves; that is,

$$B(p) = \frac{1}{\mu p} \int_0^p F^{-1}(t) dt = \frac{1}{\mu p} \int_0^q z f(z) dz, \quad 0 < p < 1,$$

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

Figure 4 shows the Bonferroni curves for the $GTG(1, \lambda, \alpha)$ model, considering different values for λ and α .

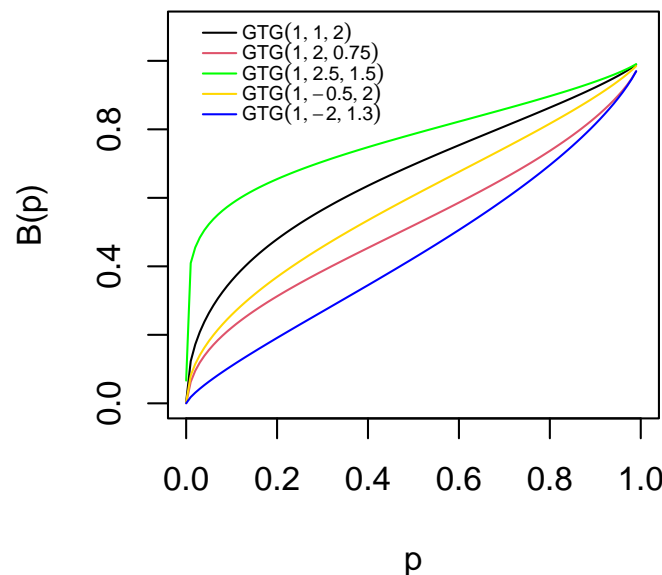


Figure 4. Bonferroni curves for the $GTG(\beta, \lambda, \alpha)$ model, with $\beta = 1$ and for selected values of λ and α .

3. Inference for the GTG Distribution

In this section, we discuss the maximum likelihood (ML) approach for parameter estimation in the GTG model.

3.1. Maximum Likelihood Estimators

Let z_1, z_2, \dots, z_n be a random sample of size n from the GTG(β, λ, α) model. Then, the log-likelihood function for $\theta = (\beta, \lambda, \alpha)$ is given by

$$\ell(\theta) = n \log(\alpha) + (\alpha - 1) \sum_{i=1}^n \log(z_i) - n\alpha \log(\beta) - n \log(1 - G(-\lambda)) - \sum_{i=1}^n \left(\frac{z_i}{\beta}\right)^\alpha + n\lambda - \sum_{i=1}^n \exp(-v_i), \quad (8)$$

where $v_i = \left(\frac{z_i}{\beta}\right)^\alpha - \lambda$. Therefore, the score assumes the form $S(\theta) = (S_\beta(\theta), S_\lambda(\theta), S_\alpha(\theta))$, where

$$S_\beta(\theta) = -\frac{n\alpha}{\beta} + \sum_{i=1}^n \frac{\alpha z_i^\alpha}{\beta^{\alpha+1}} (1 - \exp(-v_i)), \quad (9)$$

$$S_\lambda(\theta) = \frac{ng(-\lambda)}{1 - G(-\lambda)} + n - \sum_{i=1}^n \exp(-v_i), \quad (10)$$

and

$$S_\alpha(\theta) = \frac{n}{\alpha} + \sum_{i=1}^n \log(z_i) - n \log(\beta) + \sum_{i=1}^n \left(\frac{z_i}{\beta}\right)^\alpha \log\left(\frac{z_i}{\beta}\right) (\exp(-v_i) - 1). \quad (11)$$

The ML estimators are then obtained by numerically solving the equation $S(\theta) = \mathbf{0}_3$, where $\mathbf{0}_p$ denotes a vector of zeros with length p . Solutions for Equations (9)–(11) can be obtained using numerical procedures in R [20], such as the Newton–Raphson method. To initialize the numerical algorithm that solves $S(\theta) = \mathbf{0}_3$, in the next subsection, we propose an initial point for the vector θ .

3.2. Initial Points

In this subsection, we propose estimators based on the quantiles for the GTG distribution, and these estimators are an alternative to the moment estimators, which meets the objective of using them as initial values to calculate the maximum likelihood estimators of the GTG distribution.

Let q_1, q_2 and q_3 be the sample quartiles that are based on z_1, z_2, \dots, z_n . Initial values for θ can be obtained by equating the sample quartiles with the theoretical quartiles. The resulting equations are given by

$$q_1 = \beta \left[G^{-1} \left(\frac{1}{4} + \frac{3}{4} G(-\lambda) \right) + \lambda \right]^{\frac{1}{\alpha}},$$

$$q_2 = \beta \left[G^{-1} \left(\frac{1}{2} + \frac{1}{2} G(-\lambda) \right) + \lambda \right]^{\frac{1}{\alpha}},$$

and

$$q_3 = \beta \left[G^{-1} \left(\frac{3}{4} + \frac{1}{4} G(-\lambda) \right) + \lambda \right]^{\frac{1}{\alpha}}.$$

The solutions for β and α , say $\tilde{\beta}$ and $\tilde{\alpha}$, can be expressed in terms of $\tilde{\lambda}$ (the solution for λ) as follows:

$$\tilde{\beta} = \tilde{\beta}(\tilde{\lambda}) = \frac{q_1}{\left[G^{-1} \left(\frac{1}{4} + \frac{3}{4} G(-\tilde{\lambda}) \right) + \tilde{\lambda} \right]^{1/\tilde{\alpha}(\tilde{\lambda})}} \quad \text{and} \quad \tilde{\alpha} = \tilde{\alpha}(\tilde{\lambda}) = \frac{\log \left(\frac{G^{-1} \left(\frac{3}{4} + \frac{1}{4} G(-\tilde{\lambda}) \right) + \tilde{\lambda}}{G^{-1} \left(\frac{1}{4} + \frac{3}{4} G(-\tilde{\lambda}) \right) + \tilde{\lambda}} \right)}{\log \left(\frac{q_3}{q_1} \right)},$$

whereas $\tilde{\lambda}$ is obtained from the non-linear equation

$$\tilde{\beta}(\tilde{\lambda}) \left[G^{-1} \left(\frac{1}{2} + \frac{1}{2} G(-\tilde{\lambda}) \right) + \tilde{\lambda} \right]^{1/\tilde{\alpha}(\tilde{\lambda})} = q_2.$$

Therefore, the initial point based on this method is given by $\theta_{\text{quart}} = (\tilde{\beta}, \tilde{\lambda}, \tilde{\alpha})$.

3.3. Observed Fisher Information Matrix

The asymptotic variance of the ML estimators, say $\hat{\theta} = (\hat{\beta}, \hat{\lambda}, \hat{\alpha})$, can be estimated from the observed Fisher information matrix defined as $\mathcal{I}(\theta) = -E[\partial^2 \ell(\theta) / \partial \theta \partial \theta^T]$, with $\ell(\theta)$ given in Equation (8). Under regularity conditions,

$$\mathcal{I}(\theta)^{-1/2} (\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_3(\mathbf{0}_3, \mathbf{I}_3), \quad \text{as } n \rightarrow +\infty, \tag{12}$$

where \mathcal{D} stands for convergence in distribution, and $N_3(\mathbf{0}_3, \mathbf{I}_3)$ denotes the standard tri-variate normal distribution (see Wang et al. [21]). Moreover, $\mathcal{I}(\theta)$ can be estimated from the matrix $-\partial^2 \ell(\theta) / \partial \theta \partial \theta^T$, whose elements are given by $I_{\beta\beta} = -\partial^2 \ell(\theta) / \partial \beta^2$, $I_{\beta\lambda} = -\partial^2 \ell(\theta) / \partial \beta \partial \lambda$, and so on. Explicitly, we have that

$$\begin{aligned} I_{\beta\beta} &= -\frac{n\alpha}{\beta^2} + \frac{(\alpha+1)\alpha}{\beta^2} \sum_{i=1}^n \left(\frac{z_i}{\beta}\right)^\alpha (1 - \exp(-v_i)) + \frac{\alpha^2}{\beta^{2(\alpha+1)}} \sum_{i=1}^n z_i^{2\alpha} \exp(-v_i), \\ I_{\beta\lambda} &= \frac{\alpha}{\beta^{\alpha+1}} \sum_{i=1}^n z_i^\alpha \exp(-v_i), \\ I_{\beta\alpha} &= \frac{n}{\beta} + \sum_{i=1}^n \frac{z_i^\alpha}{\beta^{\alpha+1}} [1 - \exp(-v_i)] \left[1 + \alpha \log\left(\frac{z_i}{\beta}\right) \right] - \frac{\alpha}{\beta^{\alpha+1}} \sum_{i=1}^n z_i^\alpha \left(\frac{z_i}{\beta}\right)^\alpha \log\left(\frac{z_i}{\beta}\right) \exp(-v_i) \\ I_{\lambda\lambda} &= \frac{ng^2(-\lambda)}{(1-G(-\lambda))^2} - \frac{ng(-\lambda)(1-\exp(-\lambda))}{1-G(-\lambda)} + \sum_{i=1}^n \exp(-v_i), \\ I_{\lambda\alpha} &= -\sum_{i=1}^n \left(\frac{z_i}{\beta}\right)^\alpha \log\left(\frac{z_i}{\beta}\right) \exp(-v_i), \\ I_{\alpha\alpha} &= \frac{n}{\alpha^2} - \sum_{i=1}^n \log^2\left(\frac{z_i}{\beta}\right) \left(\frac{z_i}{\beta}\right)^\alpha [\exp(-v_i) - 1] + \sum_{i=1}^n \log^2\left(\frac{z_i}{\beta}\right) \left(\frac{z_i}{\beta}\right)^{2\alpha} \exp(-v_i), \end{aligned}$$

where $v_i = \left(\frac{z_i}{\beta}\right) - \lambda$.

4. GTG Quantile Regression Model

For the GTG model, the mean has a complicated form, and then, it is not recommendable to consider a mean-parameterized version of the model. On the other hand, and thinking in a context of heterogeneous observations, quantile regression is a more appropriate tool for analyzing data in presence of covariates because they allow for a complete description of the distribution of the response variable (not just a particular measure as is the case when regression on the mean is conducted).

Specifically, for the GTG model and considering that τ represents the p th quantile of the distribution, we obtain the equation $\tau = Q(p; \beta, \alpha, \lambda)$, $\tau \in (0, \infty)$. By solving such an equation, we obtain

$$\beta = \tau(h(\lambda, p))^{-1/\alpha},$$

where $h(\lambda, p) = \lambda + G^{-1}[p(1 - G(-\lambda)) + G(-\lambda)]$.

Thus, we can reparameterize the pdf and cdf of the GTG model as

$$f(z; \tau, \lambda, \alpha) = \frac{\alpha z^{\alpha-1} h(\lambda, p)}{\tau^\alpha (1 - G(-\lambda))} g\left(\left(\frac{z}{\tau}\right)^\alpha h(\lambda, p) - \lambda\right) \tag{13}$$

and

$$F(z; \tau, \lambda, \alpha) = \frac{G\left(\left(\frac{z}{\tau}\right)^\alpha h(\lambda, p) - \lambda\right) - G(-\lambda)}{1 - G(-\lambda)},$$

respectively, where $z > 0, \alpha > 0, \tau \in (0, \infty)$, and $0 < p < 1$ is fixed. We refer to this model as the reparameterized GTG (RGTG) model.

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

$$\psi(\tau_i(p)) = \mathbf{z}_i^\top \boldsymbol{\beta}(p), \tag{14}$$

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 RGTG quantile regression model is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}(p)) &= k_n(p) + (\alpha(p) - 1) \sum_{i=1}^n \log(z_i) - \alpha(p) \sum_{i=1}^n \log(\tau_i(p)) - h(\lambda(p), p) \sum_{i=1}^n \left(\frac{z_i}{\tau_i(p)}\right)^{\alpha(p)} \\ &+ \sum_{i=1}^n \exp\left(-\left(\frac{z_i}{\tau_i(p)}\right)^{\alpha(p)} h(\lambda(p), p) + \lambda(p)\right), \end{aligned} \tag{15}$$

where $k_n(p) = n(\log(\alpha(p)) + h(\lambda(p), p) - \log(1 - G(-\lambda(p))) + \lambda(p))$. The estimation of the regression parameters is obtained by directly maximizing this function.

5. Simulation Study

In this section, we present two simulation studies related to assessing the performances of the ML estimators for the GTG model and the RGTG quantile regression model.

5.1. Without Covariates

In this study, we carried out a simulation study to evaluate the performances of the ML estimators given in Section 3.1. We generated random values from the GTG(β, λ, α) distribution with Algorithm 1.

Algorithm 1 Simulating values from the GTG(β, λ, α) distribution

- 1: Simulate $U \sim Uniform(0, 1)$
 - 2: Compute $Z = \beta(\lambda + G^{-1}[U(1 - G(-\lambda)) + G(-\lambda)])^{\frac{1}{\alpha}}$
-

We used the following sequence to perform a simulation study to evaluate the behavior in finite samples of the MLEs of the GTG model. For β , we fixed three values: 1, 2 and 3; for λ , we fixed two values: 2 and 3; for α , we fixed two values: 1 and 2; and for the sample size n we fixed four values: 150, 300, 600 and 1000. For each combination of β, λ, α and n , we simulated 1000 replicates of that size and calculated ML estimators and their standard errors. Table 3 summarizes the mean of the estimated biases (bias), the mean of the estimated standard errors (SE), and the squared root of the estimated mean squared errors (RMSE), and each estimated coverage probability (CP) was obtained by taking into account the asymptotic distribution of the ML estimator using a 95% confidence level. Note

that as the sample size increases, the bias, SE and RMSE decrease, which suggests that the ML estimators of the GTG model have an acceptable behavior even in finite samples. Moreover, the SE and RMSE terms tend to become closer as the sample size increases, suggesting that the variance of the estimators is well estimated. Finally, the CP terms come closer to the nominal value as n increases, which suggests that the asymptotic approach to the normal of the ML estimators of the GTG model is reasonable, even in finite samples.

5.2. With Covariates

Under this condition, we generated 3000 random samples with a sample size ranging from $n \in \{100, 150, \dots, 1000\}$ from the RGTG quantile regression model using three different values for $p : \{0.1, 0.5, 0.9\}$. The structure of the scenarios is defined as follows:

$$\log(\rho_i(p)) = 2.071 + 0.013z_{1i} + 0.034z_{2i}, \quad \lambda(p) = 3.881 \quad \text{and} \quad \log(\alpha(p)) = 1.220, \quad (16)$$

where $z_{1ip} \sim \text{Logis}(0, 1)$ and $z_{2ip} \sim \text{Bern}(0.6)$. Here, Logis and Bern denote the standard logistic distribution and the Bernoulli distribution, respectively. The code for reproducing the simulations is available from the following repository <https://github.com/isaaccortes1989/RGTG/tree/main/Simulation%20Study> (accessed on 5 April 2024).

Figures A1–A5 in Appendix A.1 display the standard deviation (SD), mean of SE, RMSE, mean of the relative bias (RB) and CP of the 95% asymptotic confidence intervals of the estimates, under different sample sizes. It can be observed in Figures A1–A4 that the first four measures decrease as the sample size n increases, as expected in standard asymptotic theory. Finally, the CPs in Figure A5 indicate convergence to the nominal values.

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

β	True Value		Estimator	$n = 150$				$n = 300$				$n = 600$				$n = 1000$			
	λ	α		bias	SE	RMSE	CP	bias	SE	RMSE	CP	bias	SE	RMSE	CP	bias	SE	RMSE	CP
1	2	1	$\hat{\beta}$	0.0037	0.2919	0.3391	0.940	0.0072	0.2009	0.2236	0.947	-0.0006	0.1381	0.1451	0.940	-0.0012	0.1062	0.1072	0.956
			$\hat{\lambda}$	-0.0199	0.4139	11.213	0.992	0.0027	0.2479	0.5446	0.974	0.0123	0.1647	0.1746	0.952	0.0065	0.1255	0.1272	0.962
			$\hat{\alpha}$	0.0160	0.1669	0.2052	0.952	0.0110	0.1128	0.1290	0.952	0.0021	0.0765	0.0796	0.950	0.0003	0.0586	0.0587	0.950
		2	$\hat{\beta}$	-0.0073	0.1499	0.1587	0.972	-0.0085	0.1013	0.1068	0.962	-0.0020	0.0695	0.0720	0.952	-0.0013	0.0532	0.0541	0.950
			$\hat{\lambda}$	-0.0191	0.4371	10.033	0.995	0.0089	0.2605	0.5956	0.976	0.0095	0.1650	0.1716	0.959	0.0057	0.1254	0.1283	0.955
			$\hat{\alpha}$	0.0344	0.3351	0.3904	0.962	0.0049	0.2226	0.2500	0.962	0.0070	0.1536	0.1622	0.939	0.0035	0.1173	0.1184	0.950
	3	1	$\hat{\beta}$	-0.0012	0.4218	0.4101	0.928	-0.0243	0.3005	0.2890	0.942	-0.0294	0.2138	0.2170	0.935	-0.0059	0.1659	0.1705	0.943
			$\hat{\lambda}$	0.1398	0.7379	0.8553	0.955	0.1023	0.4818	0.4865	0.965	0.0765	0.3289	0.3461	0.957	0.0296	0.2454	0.2577	0.946
			$\hat{\alpha}$	0.0089	0.2096	0.2148	0.969	-0.0076	0.1436	0.1390	0.966	-0.0115	0.1005	0.1014	0.947	-0.0014	0.0776	0.0798	0.943
		2	$\hat{\beta}$	-0.0344	0.2294	0.2225	0.946	-0.0221	0.1582	0.1651	0.942	-0.0150	0.1103	0.1094	0.954	-0.0019	0.0838	0.0862	0.931
			$\hat{\lambda}$	0.1880	0.7545	0.7776	0.958	0.1042	0.4859	0.5398	0.955	0.0599	0.3258	0.3287	0.965	0.0187	0.2433	0.2525	0.935
			$\hat{\alpha}$	-0.0027	0.4186	0.3975	0.963	-0.0097	0.2879	0.2913	0.951	-0.0116	0.2015	0.1968	0.954	0.0085	0.1554	0.1608	0.931
2	2	1	$\hat{\beta}$	-0.0249	0.5893	0.7102	0.939	-0.0043	0.4028	0.4603	0.939	0.0001	0.2765	0.2915	0.948	-0.0018	0.2115	0.2021	0.947
			$\hat{\lambda}$	-0.0104	0.5128	11.682	0.995	0.0018	0.2526	0.6524	0.971	0.0131	0.1649	0.1731	0.955	0.0061	0.1250	0.1195	0.963
			$\hat{\alpha}$	0.0068	0.1681	0.2191	0.949	0.0054	0.1127	0.1346	0.952	0.0032	0.0767	0.0807	0.951	0.0016	0.0585	0.0550	0.959
		2	$\hat{\beta}$	-0.0167	0.3002	0.3378	0.965	-0.0104	0.2012	0.2031	0.971	-0.0096	0.1391	0.1437	0.954	-0.0027	0.1066	0.1079	0.949
			$\hat{\lambda}$	-0.0277	0.5436	11.101	0.994	0.0184	0.2422	0.2441	0.984	0.0155	0.1650	0.1708	0.964	0.0066	0.1256	0.1290	0.949
			$\hat{\alpha}$	0.0461	0.3383	0.4160	0.953	0.0113	0.2232	0.2278	0.966	0.0007	0.1529	0.1576	0.940	0.0026	0.1174	0.1177	0.947
	3	1	$\hat{\beta}$	-0.0513	0.8424	0.8094	0.920	-0.0578	0.6030	0.5686	0.947	-0.0114	0.4286	0.4123	0.958	-0.0226	0.3320	0.3224	0.951
			$\hat{\lambda}$	0.1962	0.7572	0.7638	0.955	0.1102	0.4844	0.4806	0.962	0.0411	0.3219	0.3260	0.969	0.0351	0.2459	0.2421	0.963
			$\hat{\alpha}$	-0.0031	0.2100	0.2015	0.964	-0.0092	0.1440	0.1358	0.960	-0.0007	0.1008	0.0962	0.962	-0.0037	0.0775	0.0755	0.954
		2	$\hat{\beta}$	-0.0862	0.4612	0.4674	0.953	-0.0450	0.3183	0.3148	0.949	-0.0302	0.2206	0.2241	0.955	-0.0199	0.1692	0.1693	0.950
			$\hat{\lambda}$	0.2177	0.7721	0.8310	0.962	0.1053	0.4860	0.4969	0.956	0.0586	0.3257	0.3395	0.959	0.0353	0.2462	0.2493	0.949
			$\hat{\alpha}$	-0.0163	0.4175	0.4128	0.960	-0.0090	0.2898	0.2828	0.949	-0.0130	0.2011	0.2005	0.951	-0.0104	0.1550	0.1552	0.955

Table 3. Cont.

β	True Value		Estimator	bias	$n = 150$			bias	$n = 300$			bias	$n = 600$			bias	$n = 1000$		
	λ	α			SE	RMSE	CP		SE	RMSE	CP		SE	RMSE	CP		SE	RMSE	CP
3	2	1	$\hat{\beta}$	0.0808	0.8935	11.177	0.943	0.0266	0.6027	0.6492	0.956	0.0062	0.4146	0.4351	0.949	0.0107	0.3187	0.3427	0.944
			$\hat{\lambda}$	−0.0649	0.5179	11.678	0.992	0.0032	0.2556	0.4053	0.984	0.0101	0.1646	0.1735	0.954	0.0028	0.1255	0.1337	0.945
			$\hat{\alpha}$	0.0292	0.1709	0.2297	0.960	0.0116	0.1128	0.1260	0.966	0.0044	0.0767	0.0817	0.942	0.0038	0.0588	0.0634	0.945
		2	$\hat{\beta}$	−0.0374	0.4571	0.5153	0.968	−0.0152	0.3038	0.2985	0.970	−0.0122	0.2094	0.2142	0.957	−0.0018	0.1592	0.1612	0.957
			$\hat{\lambda}$	−0.0420	0.4607	12.509	0.991	0.0195	0.2431	0.2444	0.985	0.0147	0.1655	0.1694	0.966	0.0038	0.1252	0.1288	0.952
			$\hat{\alpha}$	0.0285	0.3397	0.4235	0.955	0.0065	0.2233	0.2179	0.965	0.0012	0.1535	0.1561	0.954	0.0051	0.1171	0.1184	0.947
	3	1	$\hat{\beta}$	−0.0653	12.566	12.462	0.920	−0.0268	0.9029	0.8671	0.948	−0.0290	0.6415	0.6335	0.945	−0.0172	0.4986	0.5085	0.934
			$\hat{\lambda}$	0.1842	0.7509	0.7566	0.960	0.0814	0.4759	0.4728	0.971	0.0492	0.3230	0.3345	0.962	0.0279	0.2455	0.2551	0.943
			$\hat{\alpha}$	−0.0032	0.2085	0.2064	0.964	−0.0002	0.1440	0.1380	0.959	−0.0029	0.1005	0.0998	0.952	−0.0015	0.0777	0.0787	0.941
		2	$\hat{\beta}$	−0.0838	0.6846	0.6706	0.942	−0.0934	0.4791	0.4828	0.956	−0.0305	0.3282	0.3320	0.951	−0.0131	0.2523	0.2404	0.966
			$\hat{\lambda}$	0.1653	0.7470	0.7741	0.949	0.1290	0.4913	0.5176	0.964	0.0435	0.3220	0.3313	0.955	0.0229	0.2440	0.2362	0.962
			$\hat{\alpha}$	0.0090	0.4190	0.4055	0.973	−0.0257	0.2879	0.2821	0.954	−0.0026	0.2011	0.2018	0.945	0.0012	0.1553	0.1468	0.967

6. Applications

In this section, we present two applications to real datasets to illustrate the performance of the GTG model and RGTG quantile regression model. The first application is related to the beta-carotene levels in a certain sample of persons. The second application involves explaining Mexican American individuals' body mass index (BMI) in terms of their waist circumference and sex. The BMI is calculated by dividing a person's weight in kilograms by the square of their height in meters, Kg/m². Furthermore, it is a tool that health organizations use to monitor and plan public health programs.

6.1. Without Covariates

We considered the retinol plasma database available at http://lib.stat.cmu.edu/datasets/Plasma_Retinol (accessed on 5 April 2024), which focuses on understanding the determinants of the plasma levels of retinol, beta-carotene and other carotenoids. The main variable of interest is BETADIET, which represents the amount of beta-carotene consumed daily by each individual in micrograms (mcg). The importance of analyzing this variable lies in its direct relationship with the plasma levels of beta-carotene (BETAPLASMA), a key nutrient with antioxidant properties and a precursor of vitamin A. Table 4 presents a summary statistics of BETADIET, in addition to the skewness ($\sqrt{b_1}$) and kurtosis (b_2).

Table 4. Descriptive statistics of the amount of beta-carotene consumed daily by each individual in micrograms (BETADIET).

Dataset	<i>n</i>	\bar{X}	<i>S</i> ²	$\sqrt{b_1}$	<i>b</i> ₂
BETADIET	315	2186	2,172,342	1.606	6.399

It can be seen from Figure 5 that the GTG has a better fit compared to the GT, Weibull and Slash truncation positive normal (STPN) (see Gómez et al. [22]) models, in addition to a good behavior of the fitted GTG cdf compared to the empirical cdf. Based on the Akaike information criterion (AIC) (see Akaike [23]) and Bayesian information criterion (BIC) (see Schwarz [24]) given in Table 5, we also see that the GTG model is preferred (among the fitted models) for this dataset.

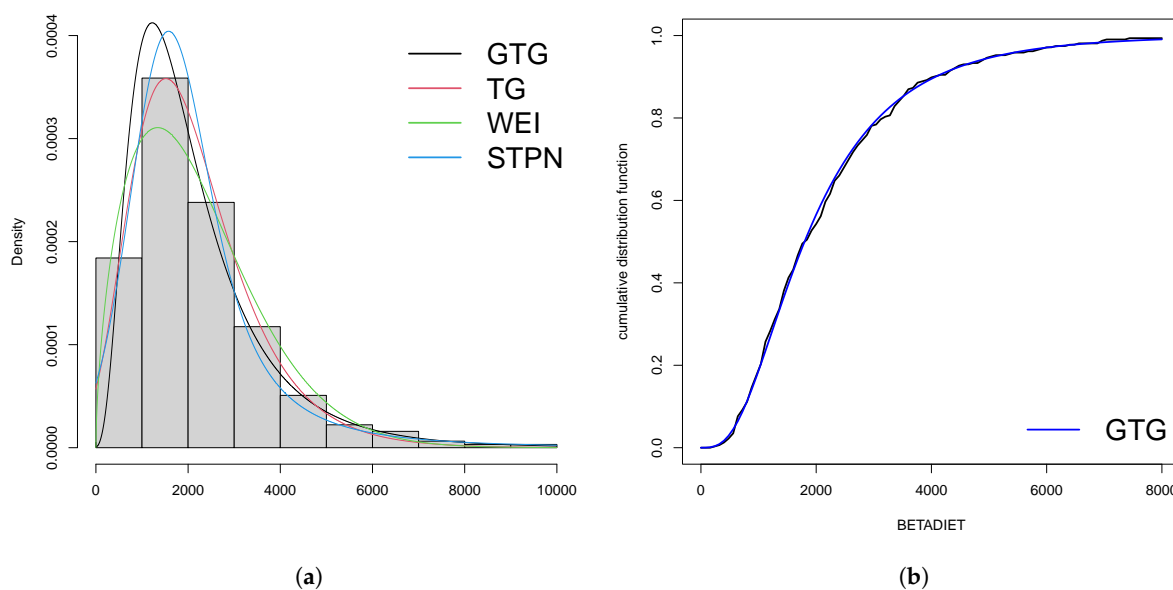


Figure 5. (a) GTG, TG, Weibull and STPN models adjusted using the maximum likelihood method for BETADIET. (b) Empirical (black) and fitted GTG (blue) cdf for the BETADIET dataset.

Table 5. Estimated parameters and their standard errors (in parentheses) for the GTG, TG, Weibull, and STPN models for the BETADIET dataset. The AIC and BIC criteria are also presented.

Parameters	GTG (SE)	TG (SE)	WEI (SE)	STPN (SE)
$\hat{\beta}$	290.83 (126.04)	1041.09 (56.53)	1.61 (0.07)	660.57 (95.37)
$\hat{\lambda}$	2.61 (0.34)	1.46 (0.10)	2454.96 (91.19)	1.97 (0.22)
$\hat{\alpha}$	0.60 (0.08)	-	-	2.46 (0.35)
LLF	-2672.896	-2682.618	-2686.675	-2685.109
AIC	5351.792	5369.236	5377.351	5376.219
BIC	5357.297	5376.742	5384.856	5387.476

6.2. With Covariates

In this application, we fit the quantile regression model to a dataset provided by Cortés et al. [25]. The dataset comprises the body mass index (BMI) measured in Kg/m², waist circumference (Waist) in centimeters, age in years and Sex (1 for female and 0 for male) of 1743 individuals who self-identified as Mexican American in the National Health and Nutrition Examination Survey (NHANES) conducted between 2017 and 2018.

Here, we assumed that the BMI follows an RGTC distribution, denoted as $Z_i \sim \text{RGTC}(p; \rho_{ip}, \alpha_p, \lambda_p)$. Accordingly, we propose the following structure for modeling:

$$\log(\rho_i(p)) = \beta_1(p) + \beta_2(p)\text{Waist}_i + \beta_3(p)\text{Gender}_i \quad \text{and} \quad \log(\alpha(p)) = \nu(p), \quad (17)$$

where $\beta_1(p)$, $\beta_2(p)$, $\beta_3(p)$, $\lambda(p)$ and $\nu(p)$ are the parameters used for the estimation, for $i = 1, \dots, 1743$ and $p \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

The parameter estimates and their corresponding standard errors are presented in Table 6, where it is observed that all estimates are statistically significant at all quantiles p . In relation to the interpretations of the regression coefficients, we can consider the following interpretations. We will present the interpretations in relation to the median, but for this problem, they are also valid for the other quantiles considered:

- The estimated median BMI for females, assuming a waist circumference of 0, is $\exp(2.076) \approx 7.973$.
- As $\exp(\hat{\beta}_2(p)) = \exp(0.013) \approx 1.013$, then for a given gender the estimated median, the BMI is increased by 1.013% for each unit that increases the waist circumference.
- As $\exp(\hat{\beta}_3(p)) = \exp(0.035) \approx 1.036$, then for a given waist circumference, the estimated median BMI for females is increased by approximately 1.036 times in comparison with that for males.
- $\hat{\lambda}(p) = 3.504$ and $\exp(\hat{\nu}(p)) = \exp(1.309) \approx 3.702$ are the estimates of the shape parameters associates with the median BMI.

Additionally, to assess the model’s adequacy, we calculated the normalized quantile residuals (NQRs) along with their respective envelopes; see Dunn and Smyth [26]. These can be observed graphically in Figure A6 of Appendix A.2, where observation #1267 is highlighted as atypical. Also, we obtained the measures of a generalized Cook’s distance and likelihood displacement for $p = 0.9$; see Figure 6. Here, we highlight that the potentially influential observations are #264, #486, #516, #1267 and #1299.

Table 7 presents a classification (https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmi_dis.htm (accessed on 5 April 2024)) of overweight and obesity by sex, BMI, waist circumference, and risk of diseases (type 2 diabetes, hypertension, and cardiovascular disease) for the highlighted observations in the residuals, generalized Cook’s distance and likelihood displacement, considering $p = 0.9$ (obese individuals). We observe from the table that observations #264 and #1267 indicate a very poor health status, emphasizing the importance of their detection.

Table 6. ML estimates of the parameters and their corresponding standard errors (in parentheses) for the RGTG quantile regression model.

p	$\hat{\beta}_1(p)$	$\hat{\beta}_2(p)$	$\hat{\beta}_3(p)$	$\hat{\lambda}(p)$	$\hat{v}(p)$
0.1	1.975 (0.0082)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.2	2.009 (0.0081)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.3	2.034 (0.0081)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.4	2.055 (0.0081)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.5	2.076 (0.0081)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.6	2.096 (0.0081)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.7	2.118 (0.0081)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.8	2.145 (0.0082)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)
0.9	2.183 (0.0084)	0.013 (0.0001)	0.035 (0.0039)	3.504 (0.2444)	1.309 (0.0675)

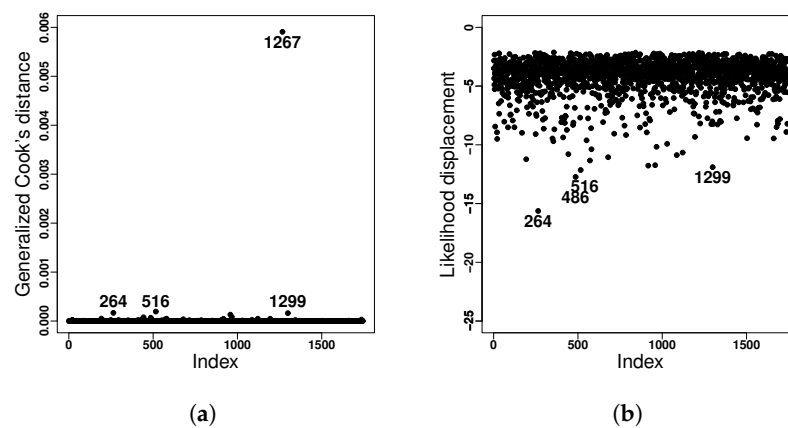


Figure 6. (a) Generalized Cook’s distance and (b) likelihood displacement for the NHANES dataset.

Table 7. Classification of overweight and obesity by BMI, waist circumference and associated disease risks.

Case	Sex	BMI	Classification of Overweight	Obesity Class	Waist	Disease Risk
264	Female	48.45	Extreme obesity	III	115.8	Extremely high
486	Female	37.21	Obesity	II	97.8	Very high
516	Female	18.04	Underweight	-	79.8	-
1267	Female	71.72	Extreme obesity	III	132.3	Extremely high
1299	Male	23.33	Normal	-	101.8	-

7. Final Discussion

In this work, we proposed a new distribution called generalized truncated Gumbel. The model has quite a few interesting properties, mainly associated with having a cumulative distribution function and a quantile function in closed form. For this reason, an extension of the model was proposed to be used in the context of quantile regression. Future extensions of the model are directed in the context of random effects.

Author Contributions: Conceptualization, H.J.G., K.I.S., D.A. and I.E.C.; formal analysis, K.I.S., D.A., D.I.G. and T.M.M.; funding acquisition, H.J.G. and D.A.; 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.A., D.I.G. and T.M.M.; writing—review and editing, D.A., 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 Applications 1 and 2. (http://lib.stat.cmu.edu/datasets/Plasma_Retinol, https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmi_dis.htm) (accessed on 5 April 2024).

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

Appendix A

Appendix A.1. Behaviors of SD, SE, RMSE, RB and CP

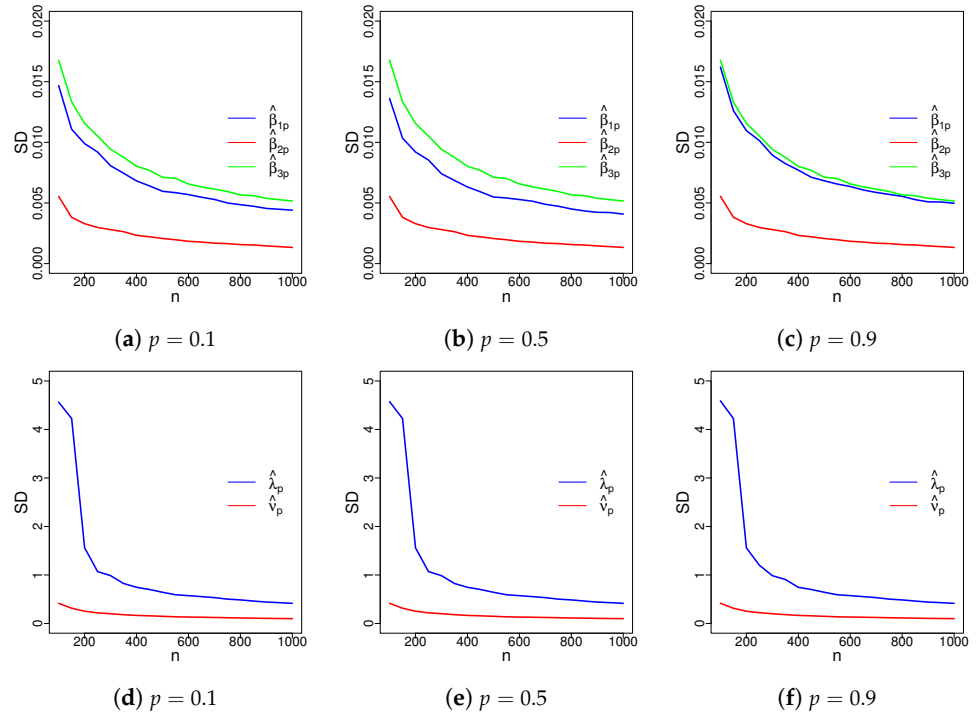


Figure A1. SDs of the 3000 estimates of ρ_q , λ_q and α_q obtained in the RGTG quantile regression model under different sample sizes.

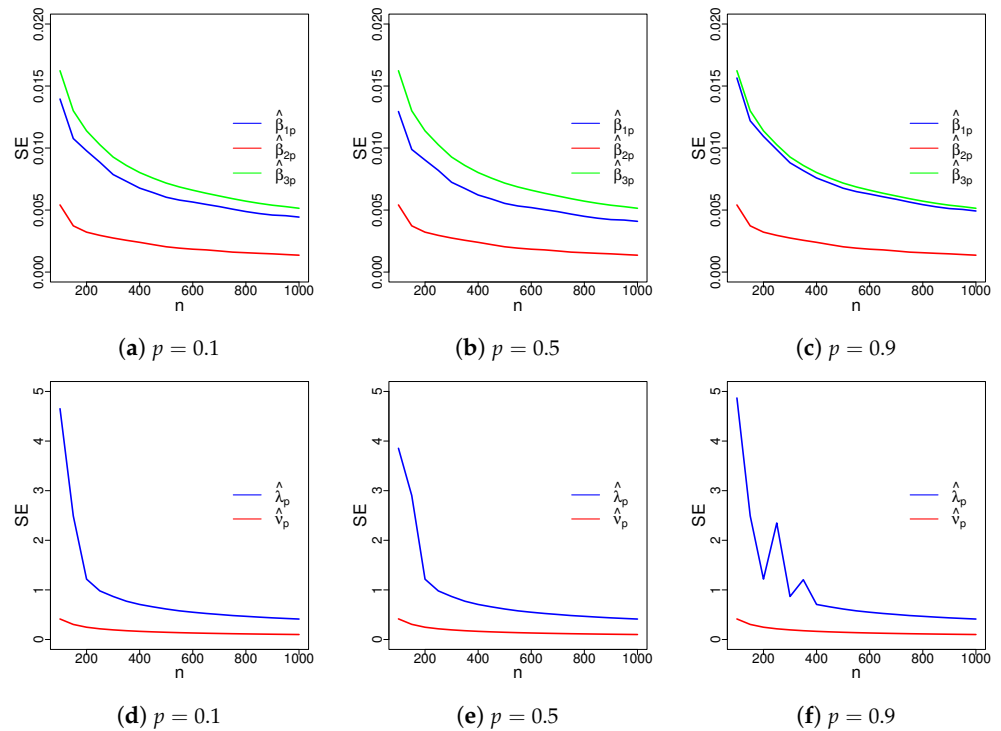


Figure A2. Means of the asymptotic SE on the 3000 estimates of ρ_q , λ_q and α_q obtained in the RGTG quantile regression model under different sample sizes.

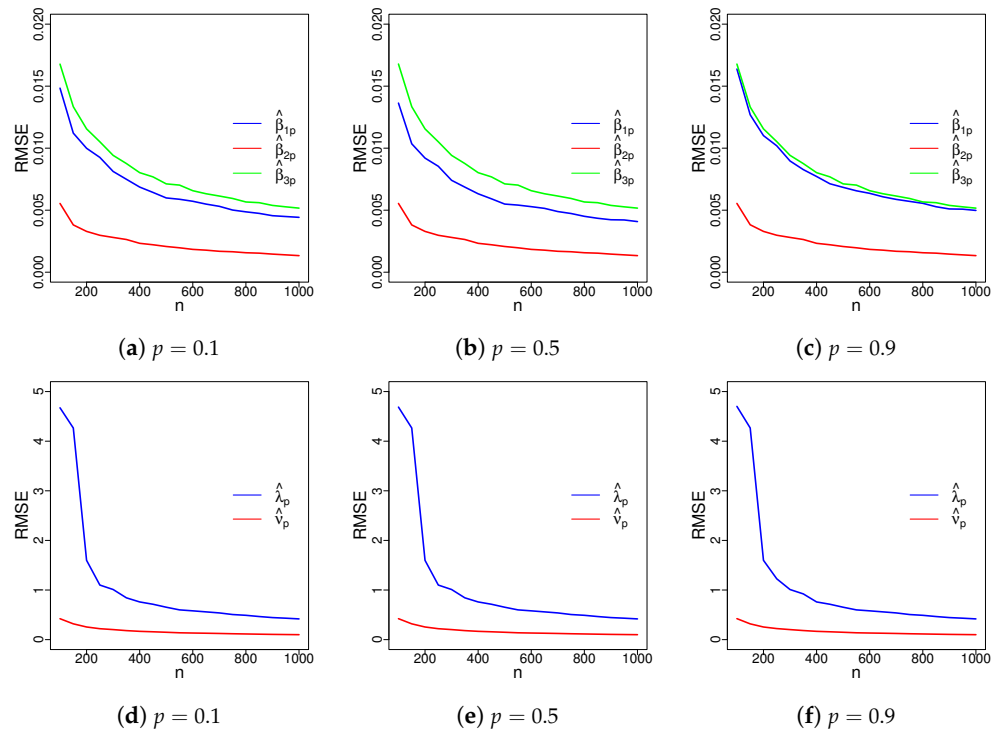


Figure A3. RMSEs of the 3000 estimates of ρ_q , λ_q and α_q obtained in the RGTG quantile regression model under different sample sizes.

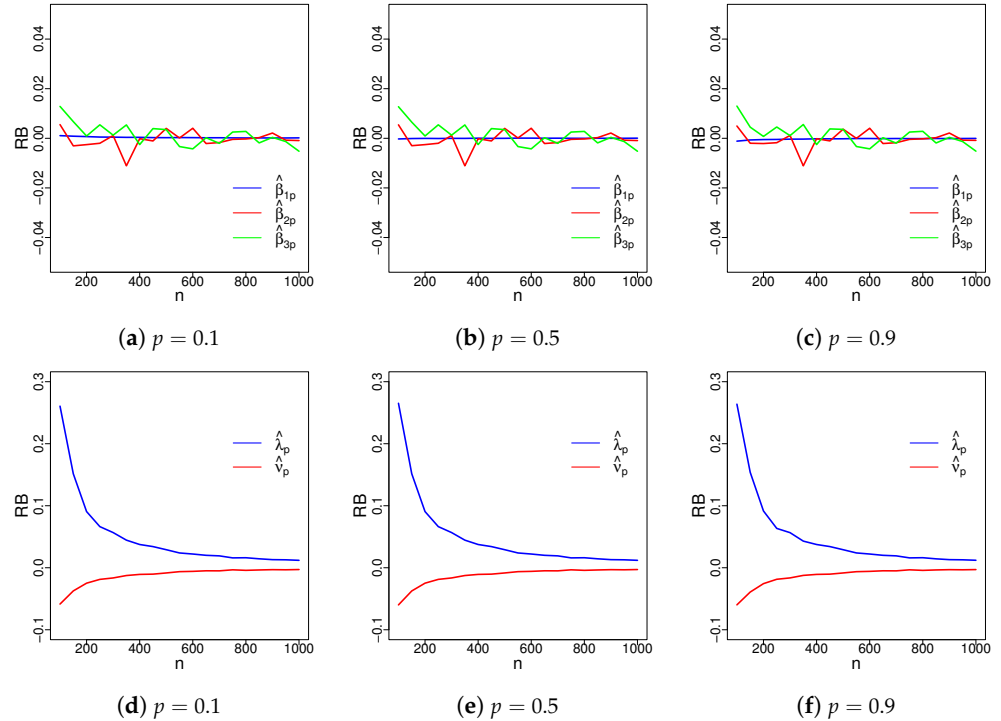


Figure A4. Means of the RB on the 3000 estimates of ρ_q , λ_q and α_q obtained in the RGTG quantile regression model under different sample sizes.

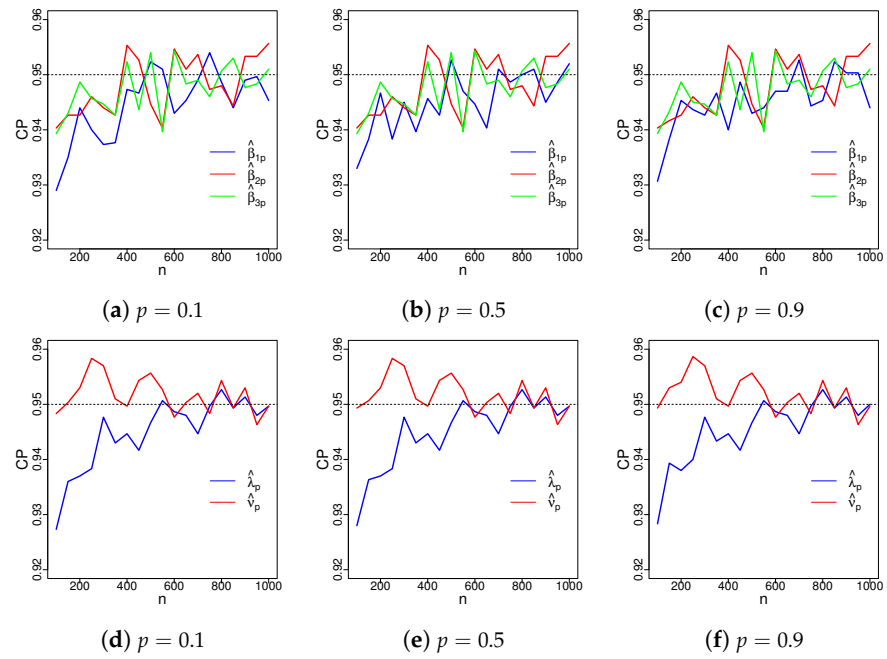


Figure A5. 95% CPs of the ρ_q , λ_q and α_q obtained in the RGTG quantile regression model under different sample sizes.

Appendix A.2. QQ-Plots

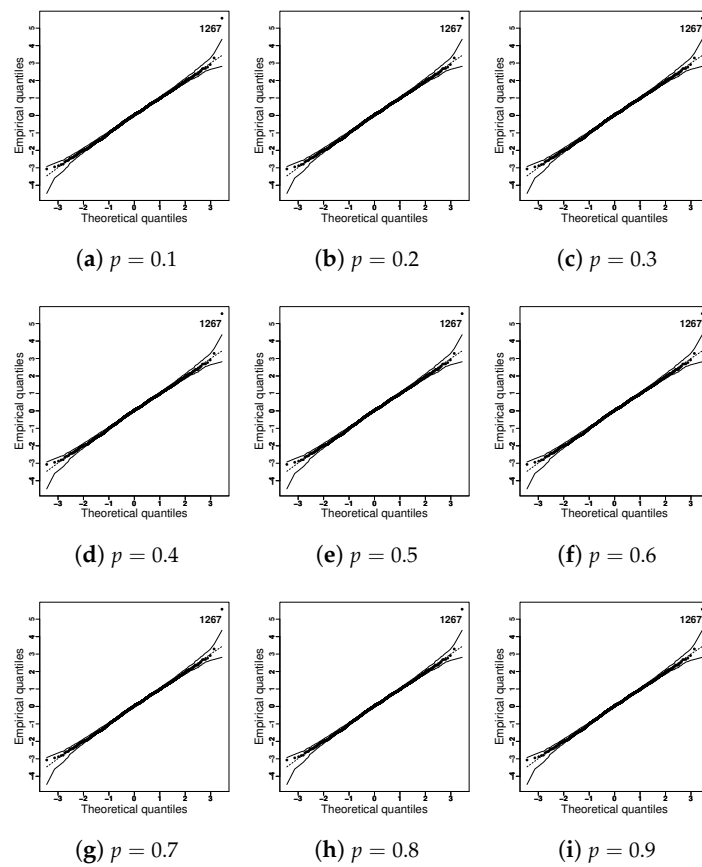


Figure A6. QQ-plots with envelopes for the NQRs under the model from Equation (16) at different probabilities p .

References

1. Bhaskaran, S.; Verma, A.S.; Goupee, A.J.; Bhattacharya, S.; Nejad, A.R.; Shi, W. Comparison of Extreme Wind and Waves Using Different Statistical Methods in 40 Offshore Wind Energy Lease Areas Worldwide. *Energies* **2023**, *16*, 6935. [CrossRef]
2. Gurung, B.; Sarkar, K.P.; Singh, K.N.; Lama, A. Modelling annual maximum temperature of India: A distributional approach. *Theor. Appl. Climatol.* **2021**, *145*, 979–988. [CrossRef]
3. Purohit, S.U.; Lalit, P.N. European option pricing using Gumbel distribution. *Int. J. Financ. Eng. (IJFE)* **2022**, *9*, 2141002. [CrossRef]
4. Li, Y.; Qiao, H.; Yang, A. Reliability Study of Magnesium Oxychloride-Coated Reinforced Concrete Based on Gumbel Distribution. *Materials* **2023**, *16*, 2521. [CrossRef] [PubMed]
5. Kang, D.; Ko, K.; Huh, J. Determination of extreme wind values using the Gumbel distribution. *Energy* **2015**, *86*, 51–58. [CrossRef]
6. Hossain, E.; Abdulrahman, A.T.; Gemeay, A.M.; Alshammari, N.; Alshawarbeh, E.; Mashaqbah, N.K. A novel extension of Gumbel distribution: Statistical inference with COVID-19 application. *Alex. Eng. J.* **2022**, *61*, 8823–8842. [CrossRef]
7. Watthanawisut, A.; Bodhisuwan, B. The Beta Topp-Leone Gumbel Distribution for Modeling the Minimum Flow Data. *Thai J. Math.* **2020**, 165–174.
8. Fayomi, A. A new extended Gumbel distribution: Properties and application. *PLoS ONE* **2022**, *17*, e0267142. [CrossRef] [PubMed]
9. Nagode, M. Gumbel mixture modelling for multiple failure data. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108946. [CrossRef]
10. Oseni, B.M.; Okasha, H.M. Gumbel-geometric Distribution: Properties and Applications. *Gazi Univ. J. Sci.* **2020**, *33*, 935–941. [CrossRef]
11. Koenker, R.; Bassett, G. Regression Quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
12. Cade, B.S.; Terrell, J.W.; Schroeder, R.L. Estimating Effects of Limiting Factors with Regression Quantiles. *Ecology* **1999**, *80*, 311–323. [CrossRef]
13. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.
14. Wei, Y.; Pere, A.; Koenker, R.; He, X. Quantile regression methods for reference growth charts. *Stat. Med.* **2006**, *25*, 1369–1382. [CrossRef] [PubMed]
15. Gómez-Déniz, E.; Leiva, V.; Calderín-Ojeda, E.; Chesneau, C. A novel claim size distribution based on a Birnbaum-Saunders and gamma mixture capturing extreme values in insurance: Estimation, regression, and applications. *Comput. Appl. Math.* **2022**, *41*, 171. [CrossRef]
16. Neamah, M.W.; Qasim, B.A. A New Left Truncated Gumbel Distribution: Properties and Estimation. *J. Phys. Conf. Ser.* **2021**, 1897, 012015. [CrossRef]
17. 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]
18. Bonferroni, C.E. *Elementi di Statistica Generale*; Libreria Seber: Firenze, Italy, 1930.
19. Arcagni, A.; Porro, F. The Graphical Representation of Inequality. *Rev. Colomb. Estadística* **2014**, *37*, 419–436. [CrossRef]
20. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023. Available online: <https://www.R-project.org/> (accessed on 12 October 2023).
21. Wang, H.; Chen, W.; Li, B. Large sample properties of maximum likelihood estimator using moving extremes ranked set sampling. *J. Korean Stat. Soc.* **2024**, *53*, 398–415. [CrossRef]
22. Gómez, H.J.; Gallardo, D.I.; Santoro, K.I. Slash Truncation Positive Normal Distribution and its Estimation Based on the EM Algorithm. *Symmetry* **2021**, *13*, 2164. [CrossRef]
23. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.
24. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
25. Cortés, I.E.; De Castro, M.; Gallardo, D.I. A new family of quantile regression models applied to nutritional data. *J. Appl. Stat.* **2023**, 1–21. [CrossRef] [PubMed]
26. 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.