mathematics

Article

The Heavy-Tailed Gleser Model: Properties, Estimation, and Applications

Neveka M. Olmos [1,*], Emilio Gómez-Déniz [2] and Osvaldo Venegas [3]

1 Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile
2 Department of Quantitative Methods in Economics and TIDES Institute, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain
3 Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile
* Correspondence: neveka.olmos@uantof.cl

Abstract: In actuarial statistics, distributions with heavy tails are of great interest to actuaries, as they represent a better description of risk exposure through a type of indicator with a certain probability. These risk indicators are used to determine companies' exposure to a particular risk. In this paper, we present a distribution with heavy right tail, studying its properties and the behaviour of the tail. We estimate the parameters using the maximum likelihood method and evaluate the performance of these estimators using Monte Carlo. We analyse one set of simulated data and another set of real data, showing that the distribution studied can be used to model income data.

Keywords: gleser distribution; heavy-tailed distribution; maximum likelihood; VaR

MSC: 62E15; 62E20

check for updates

Citation: Olmos, N.M.; Gómez-Déniz, E.; Venegas, O. The Heavy-Tailed Gleser Model: Properties, Estimation, and Applications. *Mathematics* **2022**, *10*, 4577. https://doi.org/10.3390/math10234577

Academic Editor: Manuel Franco

Received: 29 September 2022
Accepted: 21 November 2022
Published: 2 December 2022

1. Introduction

Heavy-tailed distributions have been used to model data in various applied sciences, such as environmental sciences, earth sciences, and economic and actuarial sciences. Insurance datasets tend to be positive and asymmetric to the right, with heavy tails (see Ibragimov and Prokhorov [1]); distributions with these characteristics are therefore used to model insurance data. Some authors have extended certain asymmetric distributions using the slash methodology, for instance, to increase the weight of the right tail, e.g., Olmos et al. [2,3] in the half-normal (HN) and generalized half-normal (GHN) models, Astorga et al. [4] in the generalized exponential model (see Gupta and Kundu [5]; Mudholkar et al. [6]), and Gómez et al. [7] in the Gumbel model. Two other recent works are, for example, Bhati and Ravi [8] in the generalized log-Moyal model and Afify et al. [9] in the heavy-tailed exponential model. It is known that the Pareto distribution and its corresponding generalizations are heavy-tailed distributions; they have been used in several areas of knowledge, e.g., by Choulakian and Stephens [10], Zhang [11], Akinsete et al. [12], Nassar and Nada [13], Mahmoudi [14] and Boumaraf et al. [15]. It is important to study distributions with these characteristics in order to model, for example, insurance datasets and financial yields. In the present paper, we study a distribution with a heavy right tail that provides a good fit to family income data. A very necessary function in this paper is the beta function, denoted by $B(a,b)$, which can be expressed as:

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \int_0^\infty \frac{t^{a-1}}{(1+t)^{a+b}}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (1)$$

where $a > 0$, $b > 0$ and $\Gamma(\cdot)$ is the gamma function. The beta function is the normalisation constant of the beta distribution, i.e., we say that the random variable Y has a beta distribution with parameters a and b if its probability density function (pdf) is given by

$$f_Y(y; a, b) = \frac{1}{B(a, b)} y^{a-1}(1-y)^{b-1}, \quad 0 < y < 1, \tag{2}$$

where $a > 0$ and $b > 0$.

The incomplete beta function is denoted by $B(y; a, b)$ and can be expressed as:

$$B(y; a, b) = \int_0^y t^{a-1}(1-t)^{b-1} dt, \quad 0 < y < 1, \tag{3}$$

where $a > 0$ and $b > 0$. Another related function is the regularised incomplete beta function, denoted by $I_y(a, b)$, and expressed as $I_y(a, b) = B(y; a, b)/B(a, b)$.

Gleser [16] introduced a representation of the gamma distribution, the product of a mixed scale between an unknown distribution and the exponential distribution. The object of the present paper is to study this unknown distribution, which we call the Gleser (G) distribution. A random variable X has a G distribution with parameter α if its pdf is given by

$$f_X(x; \alpha) = \frac{x^{-\alpha}}{B(\overline{\alpha}, \alpha)(1+x)}, \quad x > 0, \tag{4}$$

with $0 < \alpha < 1$ shape parameter and $\overline{\alpha} = 1 - \alpha$. The G distribution is a particular case of the beta prime distribution (see Keeping [17]), also called the inverted beta distribution (see Tiao and Cuttman [18]), but as it is a distribution with only one parameter, we were interested in studying it and applying some of its properties.Taking the density given in (4), and considering a new scale parameter, we obtain a more flexible distribution for modelling positive data with heavy right tail.

The article is organised as follows. In Section 2 we give some properties of the G distribution. In Section 3 we study the behaviour of the tail of the G distribution. In Section 4 we carry out parameter estimation using the maximum likelihood (ML) method and do a simulation study and asymptotic convergence of the ML estimators. Section 5 shows an application with data from the economic field. In Section 6 we offer some conclusions.

2. The G Distribution

In this section, we study the basic properties of the G distribution given in (4) incorporating a scale parameter. A random variable X has a G distribution with positive support and parameters σ and α if its pdf is:

$$f_X(x; \sigma, \alpha) = \frac{\sigma^{\alpha} x^{-\alpha}}{B(\overline{\alpha}, \alpha)(\sigma + x)}, \quad x > 0, \tag{5}$$

where $\sigma > 0$ scale parameter and $0 < \alpha < 1$ shape parameter; we denote this by $X \sim G(\sigma, \alpha)$. The G distribution is an alternative model to distributions with two parameters, to be used for modelling actuarial statistics data such as the Pareto and GHN distributions, among others. Figure 1 shows the graph of the G density for different values of parameter α, fixing $\sigma = 1$.

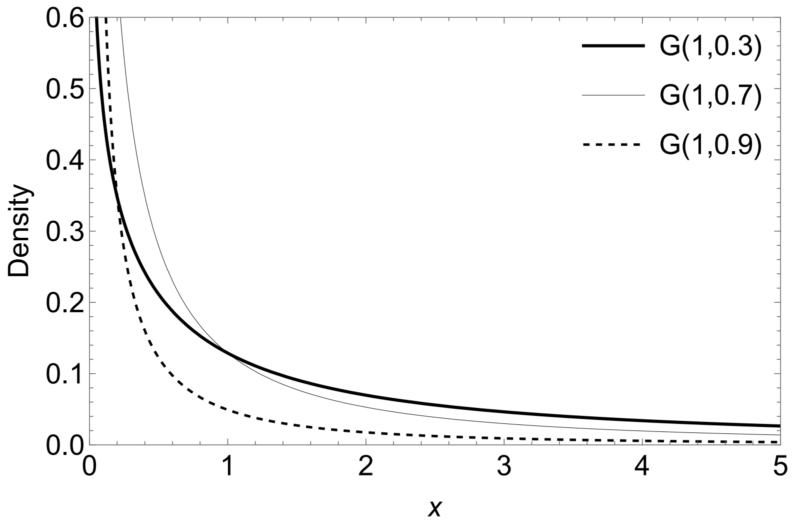

Figure 1. Examples of G(1, 0.3), G(1, 0.7) and G(1, 0.9).

We perform a brief comparison illustrating that the tails of the G distribution are heavier when the α parameter decreases. Table 1 shows $P(X > x)$ for different values of x in this distribution.

Table 1. Tails comparison.

Distribution	$P(X > 2)$	$P(X > 3)$	$P(X > 4)$
G(1, 0.9)	0.04803	0.03537	0.02818
G(1, 0.7)	0.19065	0.15106	0.12697
G(1, 0.3)	0.63376	0.57717	0.53760

2.1. Properties

The following proposition shows some properties of the G distribution.

Proposition 1. *Let $X \sim G(\sigma, \alpha)$ and $Y \sim Beta(1-\alpha, \alpha)$. Then,*

(a) *the G distribution has unimodality (at 0).*

(b) $\frac{\sigma Y}{1-Y} \sim G(\sigma, \alpha)$.

(c) *the cumulative distribution function (cdf) of X is given by*

$$F_X(x; \sigma, \alpha) = I_y(\overline{\alpha}, \alpha), \quad x > 0, \tag{6}$$

where $y = \frac{x}{\sigma + x}$ and $I_y(\cdot, \cdot)$ is the regularised incomplete beta function.

(d) *the hazard function of X is decreasing for all $x > 0$.*

(e) *the r-th moment of the random variable X does not exist for $r \geq \alpha$.*

(f) $\frac{1}{1+X} \sim Beta(\alpha, \overline{\alpha})$.

(g) *the quantile function (Q) of the G distribution is given by*

$$Q(p) = \frac{\sigma I_p^{-1}(\overline{\alpha}, \alpha)}{1 - I_p^{-1}(\overline{\alpha}, \alpha)}, \qquad 0 < p < 1, \tag{7}$$

where I_p^{-1} is the inverse function of the regularised incomplete beta function.

Proof. (d) Using the theorem, item (b), given in Glaser [19] we have that

$$\eta(x) = -\frac{f_X'(x; \sigma, \alpha)}{f_X(x; \sigma, \alpha)} = \frac{\alpha}{x} + \frac{1}{\sigma + x},$$

where $f_X(x;\sigma,\alpha)$ is the pdf given in (5), then the derivative of $\eta(x)$ with respect to x

$$\eta'(x) = -\left(\frac{\alpha}{x^2} + \frac{1}{(\sigma+x)^2}\right) < 0,\ \forall x > 0$$

gives the result.

(e) Considering $\sigma = 1$, we claim that the integral $\int_1^\infty \frac{x^{-\alpha+r}}{1+x}dx$ is divergent. In fact, taking $g_1(x) = \frac{x^{-\alpha+r}}{1+x}$ and $g_2(x) = \frac{1}{x^{1+\alpha-r}}$, we have that

$$\lim_{x\to\infty} \frac{g_1(x)}{g_2(x)} = \lim_{x\to\infty} \frac{x}{1+x} = 1 \neq 0;$$

and as the integral $\int_1^\infty \frac{1}{x^{1+\alpha-r}}dx$ is divergent for $r \geq \alpha$ this proves the claim. On the other hand, since $\mathbb{E}(X^r) = \frac{1}{B(\overline{\alpha},\alpha)} \int_0^\infty \frac{x^{-\alpha+r}}{1+x}dx$, by using comparison

$$0 \leq \int_1^\infty \frac{x^{-\alpha+r}}{1+x}dx \leq \int_0^\infty \frac{x^{-\alpha+r}}{1+x}dx,$$

and the above claim, the result is reached.

□

The survival function $\overline{F}(t)$, which is the probability that an item will not fail before time t, is defined by $\overline{F}(t) = 1 - F(t)$. The survival function for a G random variable is given by $\overline{F}(t) = 1 - I_{t/(\sigma+t)}(\overline{\alpha},\alpha)$.

The hazard function $h(t)$, defined by $h(t) = \frac{f(t)}{\overline{F}(t)}$, for a G random variable is given by

$$h(t) = \frac{\sigma^\alpha t^{-\alpha}}{(\sigma+t)\left(B(\overline{\alpha},\alpha) - B\left(\frac{t}{\sigma+t},\overline{\alpha},\alpha\right)\right)}, \quad t > 0.$$

Figure 2 shows the form of the hazard function for different values of α, considering $\sigma = 1$. As we show in Proposition 1, part (d), the hazard function is always a decreasing function. In the context of reliability, it indicates that failures are more likely to occur earlier in a product's useful life.

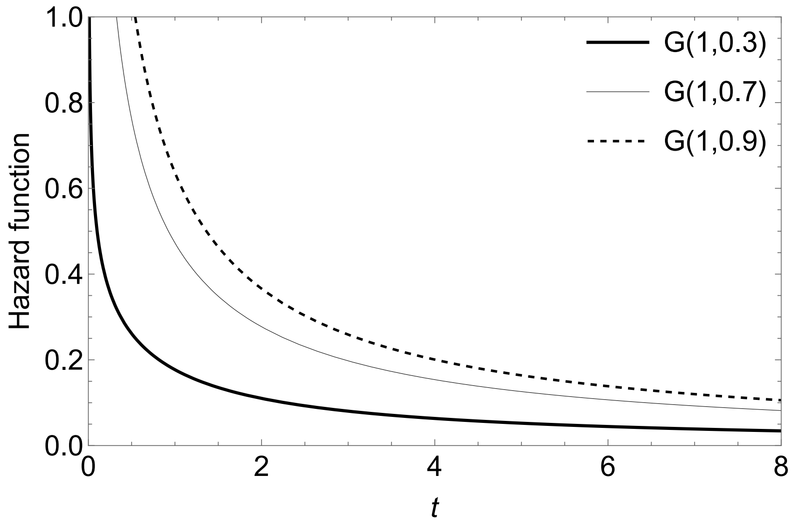

Figure 2. Examples of $h(t)$, for $\alpha = 0.3$, $\alpha = 0.7$, $\alpha = 0.9$.

Using the $Q(p)$ function given in Proposition 1 (g), we can compute the coefficient of skewness ($\sqrt{\beta_1}$) and the coefficient of kurtosis (β_2) for the random variable $X \sim G(\sigma, \alpha)$

$$\sqrt{\beta_1} = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{2}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}, \qquad \beta_2 = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}.$$

Figure 3 depicts plots for the skewness and kurtosis coefficients in the G distribution.

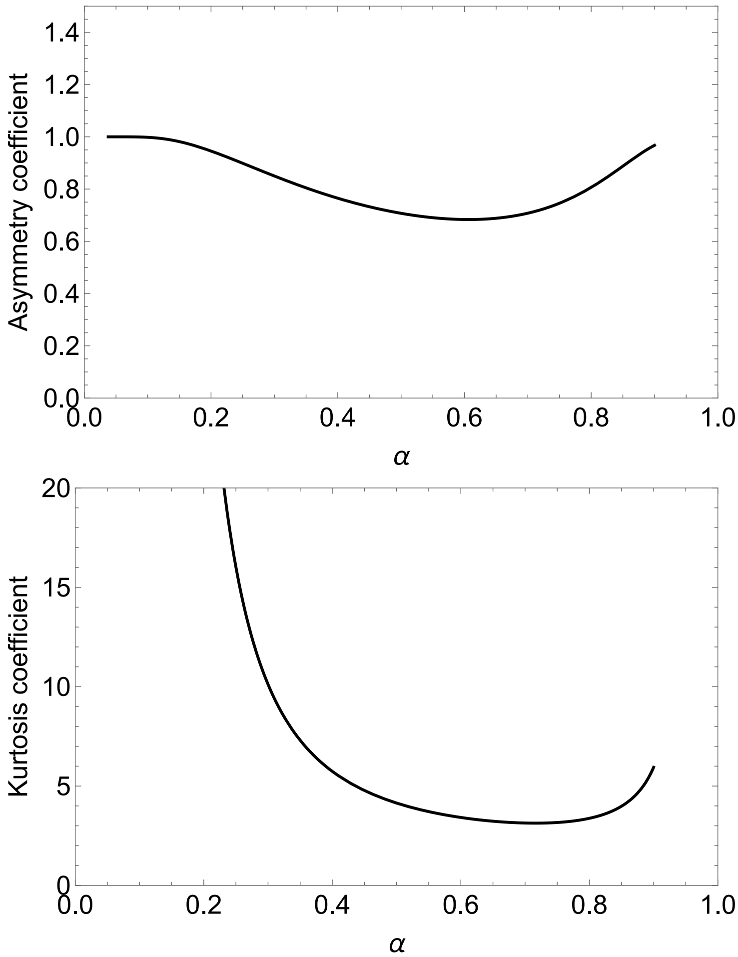

Figure 3. Plots of the skewness and kurtosis coefficients for the G model.

2.2. Actuarial Measure

The value at risk (VaR) is used to assess the risk exposure, i.e., it can be used to determine the amount of capital necessary to liquidate adverse results. The VaR of the random variable $X \sim G(\sigma, \alpha)$ is defined as (see Artzner [20] and Artzner et al. [21])

$$VaR_p = Q(p) = \frac{\sigma I_p^{-1}(\overline{\alpha}, \alpha)}{1 - I_p^{-1}(\overline{\alpha}, \alpha)}, \ \ 0 < p < 1,$$

where I_p^{-1} is the inverse function of the regularised incomplete beta function.

Figure 4 shows graphs of the VaR_p measurement of distribution $\mathrm{G}(10, \alpha)$ for different values of parameter α. We may observe that the smaller the value of parameter α, the larger the value of the VaR_p measurement.

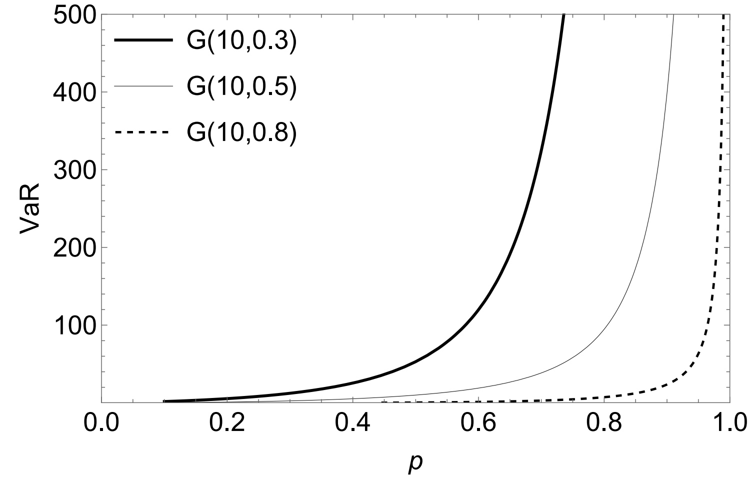

Figure 4. Plots of the VaR, for G(10, 0.3), G(10, 0.5) and G(10, 0.8).

2.3. Order Statistics

Let $X_1, \ldots, X_n$ be a random sample of the random variable $X \sim G(\sigma, \alpha)$, let us denote by $X_{(j)}$ the $jth-$order statistics, $j \in \{1, \ldots, n\}$.

Proposition 2. *The pdf of $X_{(j)}$ is*

$$f_{X_{(j)}}(x;\sigma,\alpha) = \frac{n!}{(j-1)!(n-j)!}\frac{\sigma^{\alpha}x^{-\alpha}}{B(\overline{\alpha},\alpha)(\sigma+x)}\left[I_{x/(\sigma+x)}(\overline{\alpha},\alpha)\right]^{j-1}\left[1-I_{x/(\sigma+x)}(\overline{\alpha},\alpha)\right]^{n-j},$$

where $x > 0$. In particular, the pdf of the minimum, $X_{(1)}$, is

$$f_{X_{(1)}}(x;\sigma,\alpha) = \frac{n\sigma^{\alpha}x^{-\alpha}}{B(\overline{\alpha},\alpha)(\sigma+x)}\left[1-I_{x/(\sigma+x)}(\overline{\alpha},\alpha)\right]^{n-1}, \quad x > 0 \tag{8}$$

and the pdf of the maximum, $X_{(n)}$, is

$$f_{X_{(n)}}(x;\sigma,\alpha) = \frac{n\sigma^{\alpha}x^{-\alpha}}{B(\overline{\alpha},\alpha)(\sigma+x)}\left[I_{x/(\sigma+x)}(\overline{\alpha},\alpha)\right]^{n-1}, \quad x > 0.$$

Proof. Since we are dealing with an absolutely continuous model, the pdf of the $jth-$order statistics is obtained by applying

$$f_{X_{(j)}}(x;\sigma,\alpha) = \frac{n!}{(j-1)!(n-j)!}f_X(x;\sigma,\alpha)\,[F_X(x;\sigma,\alpha)]^{j-1}\,[1-F_X(x;\sigma,\alpha)]^{n-j}, \quad j \in \{1,\ldots,n\}$$

where F and f denote the cdf and pdf of the parent distribution, $X \sim G(\sigma, \alpha)$ in this case. □

2.4. Entropy

In this subsection we will discuss the Shannon and Rényi entropies of the G model.

2.4.1. Shannon Entropy

The following lemma will be very useful for calculating the Shannon entropy.

Lemma 1. *Let $X \sim G(\sigma, \alpha)$, then we have the following results.*

1. $\mathbb{E}(\log X) = \log\sigma + \psi(\overline{\alpha}) - \psi(\alpha)$,
2. $\mathbb{E}(\log(\sigma + X)) = \log\sigma - \gamma - \psi(\alpha)$,

where $\psi(\cdot)$ is the digamma function and $\gamma = -\psi(1)$ is Euler's constant.

Proof. Both results are obtained directly using the pdf given in (5). □

The Shannon (S) entropy (see Shannon [22]) for a random variable X is defined as

$$S(X) = -\mathbb{E}(\log f_X(X)). \tag{9}$$

Therefore, it can be verified that the S entropy for the G model is

Proposition 3. *Let $X \sim G(\sigma, \alpha)$. Then the Shannon entropy of X is*

$$S(X) = \log \sigma + \log B(\overline{\alpha}, \alpha) - \gamma + \alpha\psi(\overline{\alpha}) - (1+\alpha)\psi(\alpha). \tag{10}$$

Figure 5 shows the Shannon entropy for the G model fixing $\sigma = 1$.

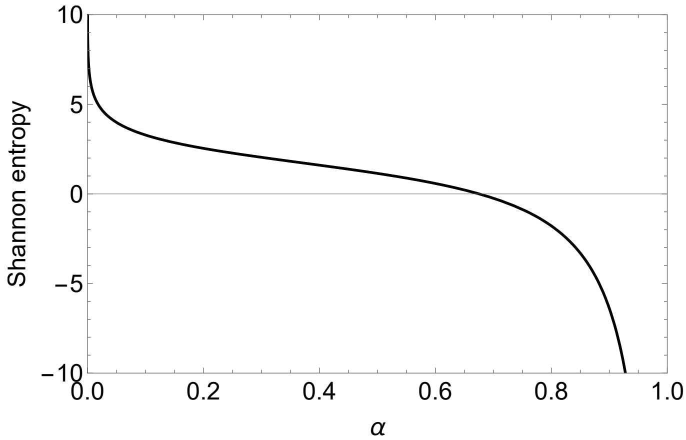

Figure 5. Shannon entropy of $G(\alpha, \sigma = 1)$ for different values of α.

2.4.2. Rényi Entropy

A generalization of the Shannon entropy is the Rényi (R_p) entropy, which is defined as

$$R_p(X) = \frac{1}{1-p} \log\left(\int_0^\infty [f_X(x)]^p dx\right). \tag{11}$$

Therefore, it can be verified that the R_p entropy for the G model is

Proposition 4. *Let $X \sim G(\sigma, \alpha)$. Then the Rényi entropy of X is*

$$R_p(X) = \frac{1}{1-p}(\log \Gamma(\xi) + \log \Gamma(p - \xi) - \log \Gamma(p) - (p-1)\log \sigma - p \log B(\overline{\alpha}, \alpha)), \tag{12}$$

where $\xi = 1 - p\alpha > 0$.

Corollary 1. *Let $X \sim G(\sigma, \alpha)$ with $S(X)$ and $R_p(X)$ the Shannon and Rényi entropies. Then we have*

$$\lim_{p \to 1} R_p(X) = S(X). \tag{13}$$

3. Tail of the Distribution

A random variable with a non-negative support, like the classic Pareto distribution, is commonly used in insurance contexts to model the amount of losses. The size of the distribution tail is fundamental if we want the chosen model to capture quantities sufficiently removed from the start of the distribution support, i.e., atypical (extreme) values. The concept of heavy tail is fundamental for this and other financial scenarios.

The use of heavy right-tailed distributions is of vital importance in general insurance. Pareto, Log-normal and Weibull distributions, among others, have been used to model third party liability insurance losses for motor vehicles, re-insurance and catastrophe insurance.

Let $\mathcal{S}$ be the class of subexponential distributions. That is, $F \in \mathcal{S}$ defined in $\mathbb{R}_+$ satisfies that:

$$\lim_{x\to\infty} \frac{1-F^{*2}(x)}{1-F(x)} = 2,$$

where F^{*j} is the j-fold convolution of F.

The following Lemma, appearing, among others, in Chapet 2, p. 55 in Rolski et al. [23], is required in the next Theorem.

Lemma 2. *Let $F(x)$ and $H(x)$ be two distributions on $\mathbb{R}_+$ and assume that there exists a constant $c > 0$ such that*

$$\lim_{x\to\infty} \frac{\overline{H}(x)}{\overline{F}(x)} = c. \tag{14}$$

Then, $F \in \mathcal{S}$ if and only if $H \in \mathcal{S}$.

Theorem 1. *Let the cdf of $G(\sigma, \alpha)$ given in (6). Then, $F_X(x;\sigma,\alpha) \in \mathcal{S}$.*

Proof. Let $\overline{H}_X(x;\sigma,\alpha) = (\sigma/(\sigma+x))^{\alpha}$ the survival function of the Pareto Type II distribution (Lomax distribution). By using this together with the complementary of the cdf (6) and by computing (14) we get, after applying L'Hopital's rule, that

$$\lim_{x\to\infty} \frac{\overline{H}(x)}{\overline{F}(x)} = \alpha B(\overline{\alpha}, \alpha) > 0.$$

Now, taking into account that the Pareto type II distribution belongs to the class of subexponential distributions, we have the result. □

As a consequence of the previous Theorem (see Rolski et al. [23] p. 50), we have the following Corollary.

Corollary 2. *If X_1 and X_2 are independent and identically distributed random variables with distributions given in (6), then as $x \to \infty$*

$$\Pr(X_1 + X_2 > x) \sim \Pr(\max\{X_1, X_2\} > x).$$

Proof. The result is a consequence of the fact that $F_X(x;\sigma,\alpha) \in \mathcal{S}$. □

Proposition 5. *It is verified that $F_X(x;\sigma,\alpha)$ is heavy right-tailed distribution.*

Proof. Since $F_X(x;\sigma,\alpha) \in \mathcal{S}$, the result is a consequence of applying Theorem 2.5.2 in Rolski et al. [23]. □

Another way to see that $F_X(x;\sigma,\alpha)$ is a heavy right-tailed distribution is by computing the limit $\lim_{x\to\infty}(-\log \overline{F}_X(x;\sigma,\alpha)/x)$, which results 0. Observe that $-\log \overline{F}_X(x;\sigma,\alpha)/x)$ is the hazard function of X.

As a consequence of the last result, we have the following Corollary:

Corollary 3. *It is verified that* $\limsup_{x\to\infty} e^{sx}\overline{F}_X(x;\sigma,\alpha) = \infty$, $s > 0$.

In this case the distribution fails to possess any positive exponential moment, i.e., $\int \exp(sx)\, dF_X(x;\sigma,\alpha) = \infty$ for all $s > 0$ see Chapter 1, p. 2 [24]. Distributions of this type have moment generating function $M_{F_X(x;\sigma,\alpha)}(s) = \infty$, for all $s > 0$. This occurs, for example, with the log-normal distribution.

An important issue in extreme value theory is regular variation (see Bingham [25] and Konstantinides [26]). This is a flexible description of the variation of some functions according to the polynomial form of the type $x^{-\delta} + o(x^{-\delta})$, $\delta > 0$. This concept is formalized in the following definition.

Definition 1. *A CDF (measurable function) is called regular varying at infinity with index* $-\delta$ *if it holds:*

$$\lim_{x\to\infty} \frac{\overline{F}(\tau x)}{\overline{F}(x)} = \tau^{-\delta},$$

where $\tau > 0$ *and the parameter* $\delta \geq 0$ *is called the tail index.*

The following proposition establishes that the survival function of the G distribution is a distribution with regular variation.

Proposition 6. *The survival function of* $G(\sigma, \alpha)$ *is regularly varying with tail index* α*.*

Proof. Applying the above definition and using L'Hopital's Rule, we have that

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha+1} \lim_{x\to\infty} \left(\frac{\sigma+x}{\sigma+tx}\right)^{-\alpha} \left(\frac{\sigma+x}{\sigma+tx}\right)^{\alpha-1} \left(\frac{\sigma+x}{\sigma+tx}\right)^{2} = t^{-\alpha+1} \lim_{x\to\infty} \frac{\sigma+x}{\sigma+tx},$$

Calculating the limit to the right, we obtain the result. □

In actuarial settings and individual and collective risk models, the practitioner is usually interested in the random variable $S_n = \sum_{i=1}^n X_i$ for $n \geq 1$. Although its pdf is difficult or impossible to calculate in practice, we can approximate its probabilities using the following Corollary (see Jessen and Mikosch [27]).

Corollary 4. *If* $X_1, \ldots, X_n$ *are iid random variables* $G(\sigma, \alpha)$ *and* $S_n = \sum_{i=1}^n X_i$*,* $n \geq 1$*, then*

$$\Pr(S_n > x) \sim n \Pr(X_1 > x), \quad \text{as} \quad x \to \infty. \tag{15}$$

Therefore, if $P_n = \max_{i=1,\ldots,n} X_i$, $n \geq 1$, we have that

$$\Pr(S_n > x) \sim n \Pr(X > x) \sim \Pr(P_n > x).$$

This means that for large x the event $\{S_n > x\}$ is due to the event $\{P_n > x\}$. Therefore, high thresholds being exceeded by the sum S_n are due to this threshold being exceeded by the largest value in the sample.

As Jessen and Mikosch [27] point out, expression (15) can be taken as the definition of a subexponential distribution. The class of those distributions is larger than the class of regularly varying distributions. The result, given in Corollary 4, remains valid for subexponential distributions because the subexponentiality of S_n implies subexponentiality of X_1. Usually, this property is referred to as convolution root closure of subexponential distributions. More details can be viewed in [28,29].

On the other hand, let the random variable X, whose support is $0 < x < \infty$, represents either a policy limit or reinsurance deductible (from an insurer's perspective); then the limited expected value function L of X with cdf $F(x)$, is defined by (see Boland [30] and Hogg and Klugman [31]):

$$L(x) = \mathbb{E}[\min(X, x)] = \int_0^x y\, dF(y) + x\overline{F}(x),$$

which is the expectation of the cdf $F(x)$ truncated at this point. In other words, it represents the expected amount per claim retained by the insured on a policy with a fixed amount deductible of x. Observe that we integrate according to the interval $(0, x)$.

Proposition 7. *Let X be a random variable following the pdf (5). Then the limited expected value of X is given by*

$$L(z) = \frac{z^2}{\sigma B(\overline{\alpha}, \alpha)(2-\alpha)} \left(\frac{\sigma}{z}\right)^{\alpha} {}_2F_1(2-\alpha, 1, 3-\alpha, -z/\sigma) + z\overline{F}(z), \quad \alpha < 1,$$

where ${}_2F_1$ represents the hypergeometric function.

Proof. By making the change of variable $t = x/(z-x)$ in the integral

$$I = \frac{\sigma^{\alpha}}{B(\overline{\alpha}, \alpha)} \int_0^z \frac{dx}{x^{\alpha-1}(\sigma + x)},$$

gives the result. □

4. Inference

In this section we estimate the parameters of the G model by the ML method, and discuss a simulation study and asymptotic estimation of the ML estimators.

4.1. ML Estimation

For a random sample $x_1, \ldots, x_n$ derived from the $G(\sigma, \alpha)$ distribution, the log-likelihood function can be written as

$$\ell(\sigma, \alpha) = n\alpha \log \sigma - \alpha \sum_{i=1}^{n} \log x_i - n \log B(\overline{\alpha}, \alpha) - \sum_{i=1}^{n} \log(\sigma - x_i). \tag{16}$$

The score equations are given by:

$$\frac{n\alpha}{\sigma} - \sum_{i=1}^{n} \frac{1}{\sigma + x_i} = 0, \tag{17}$$

$$\log \sigma - \overline{\log x} + \psi(1-\alpha) - \psi(\alpha) = 0, \tag{18}$$

where $\psi(\cdot)$ is the digamma function and $\overline{\log x} = \frac{1}{n}\sum_{i=1}^{n} \log x_i$. From (17) we obtain

$$\hat{\alpha} = \frac{\hat{\sigma}}{n} \sum_{i=1}^{n} \frac{1}{\hat{\sigma} + x_i}, \tag{19}$$

and the ML estimator for σ ($\hat{\sigma}$) is obtained by solving numerically the following equation

$$\log \hat{\sigma} = \overline{\log x} + \psi\left(\frac{\hat{\sigma}}{n} \sum_{i=1}^{n} \frac{1}{\hat{\sigma} + x_i}\right) - \psi\left(1 - \frac{\hat{\sigma}}{n} \sum_{i=1}^{n} \frac{1}{\hat{\sigma} + x_i}\right). \tag{20}$$

Equation (20) can be solved by using numerical procedures such as the Newton-Raphson algorithm. Alternatively, these estimates can be found by directly maximizing the log-likelihood surface given by (16) and using the "optim" subroutine in the R software package [32] version 4.2.1.

4.2. Simulation Study

To examine the behaviour of the ML estimation, we present a simulation study to evaluate its performance, evaluating the σ and α parameters in the G model. The simulation was analysed by generating 1000 samples of size $n = 100$, 200 and 500 from the G model. The object of this simulation was to study the behaviour of the ML of the G model parameters. Algorithms 1 and 2 can be used to generate random numbers from the G model, as shown below.

Algorithm 1 For simulating from the $X \sim G(\sigma, \alpha)$ can proceed as follows

Step 1: Generate $Y \sim Uniform(0,1)$.
Step 2: Compute $X = \frac{\sigma I_Y^{-1}(\bar{\alpha},\alpha)}{1-I_Y^{-1}(\bar{\alpha},\alpha)}$.

Algorithm 2 For simulating from the $X \sim G(\sigma, \alpha)$ can proceed as follows

Step 1: Generate $Y \sim Beta(\bar{\alpha}, \alpha)$.
Step 2: Compute $X = \frac{\sigma Y}{1-Y}$.

For each sample generated from the G distribution, the ML estimates are obtained by applying the Newton-Raphson algorithm. In Table 2 the empirical bias, the mean of the standard errors (SE), the root of the empirical mean squared error (RMSE), and the 95% coverage probability (CP) based on the asymptotic distribution for ML estimators are given for the estimators of the parameters. As Table 2 shows, the performance of the estimates improves when n increases.

Table 2. Empirical bias, SE, RMSE and 95% CP for the ML estimators of σ and α in the G distribution with different combinations of parameters.

True Value			$n=100$				$n=200$				$n=500$			
σ	α	Par.	Bias	SE	RMSE	CP	Bias	SE	RMSE	CP	Bias	SE	RMSE	CP
1	0.2	σ	0.1934	0.5709	0.6543	0.9402	0.1113	0.3727	0.4083	0.9474	0.0494	0.2219	0.2292	0.9540
		α	0.0054	0.0260	0.0270	0.9618	0.0039	0.0181	0.0186	0.9562	0.0022	0.0113	0.0116	0.9542
	0.5	σ	0.2191	0.7544	0.8668	0.9004	0.1006	0.4920	0.5425	0.9156	0.0468	0.3010	0.3131	0.9370
		α	0.0004	0.0679	0.0667	0.9258	−0.0012	0.0496	0.0491	0.9308	0.0005	0.0322	0.0320	0.9424
	0.8	σ	0.0727	0.5090	0.5498	0.8988	0.0348	0.3463	0.3602	0.9216	0.0118	0.2137	0.2143	0.9432
		α	−0.0044	0.0258	0.0271	0.9550	−0.0022	0.0179	0.0183	0.9496	−0.0010	0.0112	0.0114	0.9510
10	0.2	σ	1.9401	5.7013	6.4320	0.9394	1.1103	3.7292	4.0676	0.9486	0.4651	2.2140	2.2556	0.9528
		α	0.0052	0.0260	0.0268	0.9562	0.0037	0.0181	0.0189	0.9492	0.0020	0.0113	0.0112	0.9560
	0.5	σ	1.9611	7.4418	8.4714	0.8872	1.0610	4.9654	5.2373	0.9176	0.4148	2.9938	3.0764	0.9360
		α	−0.0014	0.0681	0.0668	0.9216	0.0002	0.0497	0.0490	0.9344	0.0000	0.0322	0.0315	0.9484
	0.8	σ	0.6832	5.0843	5.3895	0.8952	0.3724	3.4679	3.5625	0.9236	0.1692	2.1465	2.1721	0.9406
		α	−0.0047	0.0259	0.0274	0.9510	−0.0020	0.0179	0.0181	0.9544	−0.0008	0.0112	0.0113	0.9534

4.3. Fisher's Information Matrix

Let us now consider $X \sim G(\sigma, \alpha)$. For a single observation x of X, the log-likelihood function for $\boldsymbol{\theta} = (\sigma, \alpha)$ is given by

$$\ell(\boldsymbol{\theta}) = \alpha \log \sigma - \alpha \log x - \log B(\bar{\alpha}, \alpha) - \log(\sigma + x).$$

The corresponding first and second partial derivatives of the log-likelihood function are derived in Appendix A. It can be shown that the Fisher's information matrix, denoted by $I_F(\cdot)$, for the G distribution is provided by

$$I_F(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\bar{\alpha}\alpha}{2\sigma^2} & -\frac{1}{\sigma} \\ -\frac{1}{\sigma} & \psi'(\alpha) + \psi'(\bar{\alpha}) \end{pmatrix},$$

where $\psi'(\cdot)$ is the trigamma function.

Proposition 8. *The ML estimate of $\boldsymbol{\theta}$ is consistent and asymptotically normal, verifying*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} N_2(\mathbf{0},\ I_F(\boldsymbol{\theta})^{-1}), \quad as \quad n \to \infty.$$

Proof. The distribution satisfies the regularity conditions see Lehmann and Casella [33] p. 449 under which the ML estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is consistent and asymptotically normal. □

Thus the asymptotic variance of the ML estimator $\hat{\boldsymbol{\theta}}$ is the inverse of Fisher's information matrix $I_F(\boldsymbol{\theta})$. Since the parameters are unknown, the observed information matrix is usually considered, where the unknown parameters are estimated by ML.

5. Applications

In this section we analyse two applications, the first with simulated data and the second with real data. To compare the models, we use the Akaike information criterion AIC (see Akaike [34]) and the Bayesian information criterion BIC (see Schwarz [35]). In these applications we use the GHN model (see Cooray and Ananda [36]) and the Pareto model (see Arnold [37]). These models present a certain similarity in form with the G model and have greater flexibility in their coefficient of kurtosis. Hence the Pareto and GHN models are considered in the comparison of the fits to the two data sets. The two models are given respectively by:

1. $f(x;\alpha,\beta) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \quad x > \alpha,$
2. $f(x;\sigma,\alpha) = \frac{2\alpha x^{\alpha-1}}{\sigma^\alpha}\phi\left(\left(\frac{x}{\sigma}\right)^\alpha\right), \quad x > 0,$

where ϕ denotes the density function of the standard normal distribution and $\sigma, \alpha, \beta > 0$.

5.1. Numerical Application

In this numerical application we analyse 200 simulated data generated from the G$(1, 0.3)$ model using Algorithm 1 given in Section 4.2. The object of this numerical example is to use the ML method given in Section 4.1 to see whether, when the G model is fitted, it has its own form.

Table 3 shows the ML estimates for the parameters of the G model, Pareto model and GHN model, as well as the AIC and BIC values for each model.

Table 3. 200 simulated data: Model, ML estimates, AIC and BIC values.

Model	ML Estimates	AIC	BIC
G(σ,α)	$\widehat{\sigma} = 0.890$, $\widehat{\alpha} = 0.295$	1967.556	1974.152
Pareto(α,β)	$\widehat{\alpha} = 0.002$, $\widehat{\beta} = 0.121$	2144.958	2151.555
GHN(σ,α)	$\widehat{\sigma} = 204.871$, $\widehat{\alpha} = 0.144$	2143.358	2151.787

We observe that the smallest values of the AIC and BIC criteria correspond to the G model, meaning that the G model fits the data better than the other models. This is to be expected, since the data were simulated from the G model. The above values for the measures indicate that the G model has its own form and that it may be difficult to replace it by any other known two-parameter model.

5.2. Application with Real Data

This dataset comes from the Survey of Consumer Finances (SCF), a nationally representative sample that contains extensive information on assets, liabilities, income, and demographic characteristics of those sampled (potential U.S. customers). It contains a random sample of 500 households in USA with positive incomes that were interviewed in the 2004 survey. The variable of interest is the annual income of the family in thousands of US dollars divided by the number of members in the household. The data can be recovered from the web page https://www.federalreserve.gov/econres/scfindex.htm (accessed on 5 January 2022). The descriptive statistics of these data are shown in Table 4, where CS is the coefficient of skewness of the sample and CK is the coefficient of kurtosis of the sample.

Table 4. Descriptive statistics for income data.

n	Median	Mean	Variance	CS	CK
500	21.125	216.709	11,270,001	0.435	1.655

Figure 6 presents two boxplot graphs; the left boxplot shows a very extreme datum and the right boxplot shows the dataset after elimination of the extreme datum to show the other extreme data that cannot be seen in the left boxplot. These atypical data make the right tail heavier. It may be noted that the majority of the observations are around 21,125 dollars per capita per family, and there is a very atypical value which is an income of 75 million dollars.

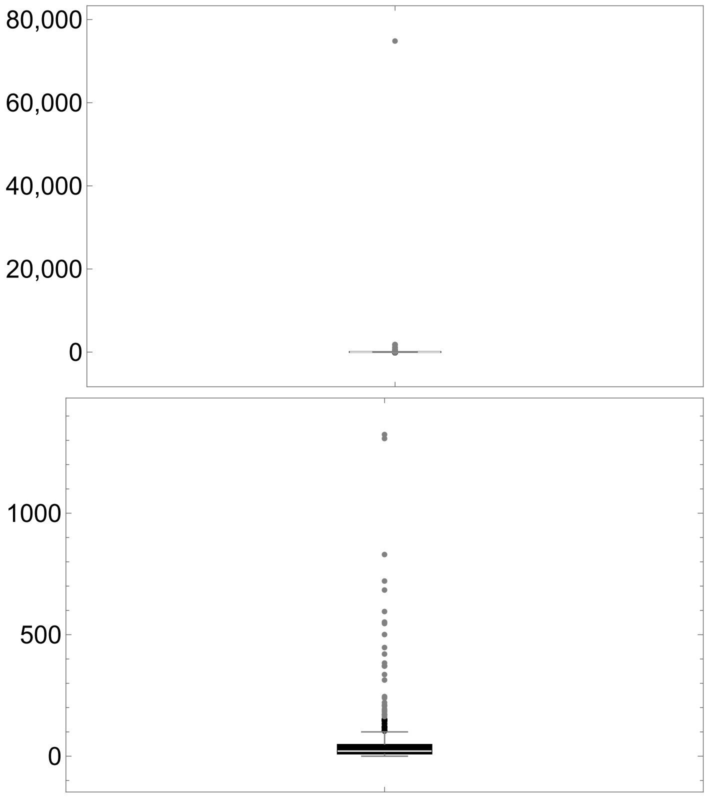

Figure 6. Boxplot for income data (**top**) and boxplot for income data without extreme data (**bottom**).

Table 5 shows the ML estimates for the parameters of the G, Pareto and GHN models, as well as the values of the AIC and BIC criteria for each model.

Table 5. ML estimates for the income data with corresponding standard errors (in parentheses), and AIC and BIC values.

Model	ML Estimates	AIC	BIC
G(σ, α)	$\widehat{\sigma} = 21.555\,(6.264)$, $\widehat{\alpha} = 0.497\,(0.033)$	5139.176	5147.605
Pareto(α, β)	$\widehat{\alpha} = 0.065\,(0.001)$, $\widehat{\beta} = 0.171\,(0.008)$	5867.910	5876.339
GHN(σ, α)	$\widehat{\sigma} = 64.313\,(6.815)$, $\widehat{\alpha} = 0.335\,(0.008)$	5280.360	5288.789

We observe that the smallest values of the AIC and BIC criteria correspond to the G model, meaning that the G model fits the data better than the Pareto and GHN models. The SE of the ML of the G model were calculated using Section 4.1.

Table 6 presents estimates of the VaR for the G, Pareto, and GHN models at the levels of 0.50, 0.60, 0.70, 0.80, 0.90, and 0.95 and empirical quantiles. It is well known that the models with the highest VaR values have the heaviest tails. Then, based on this characteristic, we can see that the G model has higher VaR values than the GHN model. On the other hand, the G model provides higher VaR values up to the 0.70 level than the Pareto model since the Pareto VaRs explode at high (but not too high) quantiles. Due to the selection criteria of the AIC and BIC models, the G model best fits the income data. Figure 7 shows the behavior of the VaR values of the three models. Figure 8 shows the empirical cdf with estimated G and Pareto cdf's, which also shows the excellent agreement between the G model and the income data.

Table 6. Comparison of VaR of different models for income data and empirical quantiles in parentheses.

Model\Significance	**0.50 (21.125)**	**0.60 (28.600)**	**0.70 (40.000)**	**0.80 (60.000)**	**0.90 (100.000)**	**0.95 (210.250)**
G $(\widehat{\sigma}, \widehat{\alpha})$	22.034	41.764	85.060	209.891	889.770	3632.538
Pareto $(\widehat{\alpha}, \widehat{\beta})$	3.744	13.805	74.248	795.165	45800.120	2638007
GHN $(\widehat{\sigma}, \widehat{\alpha})$	19.836	38.426	71.568	134.928	284.355	479.988

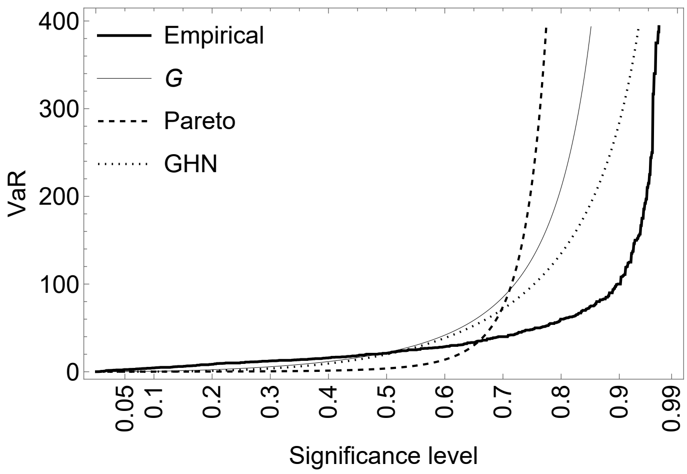

Figure 7. Plots of the VaR using the values in Table 6, for G, Pareto and GHN models.

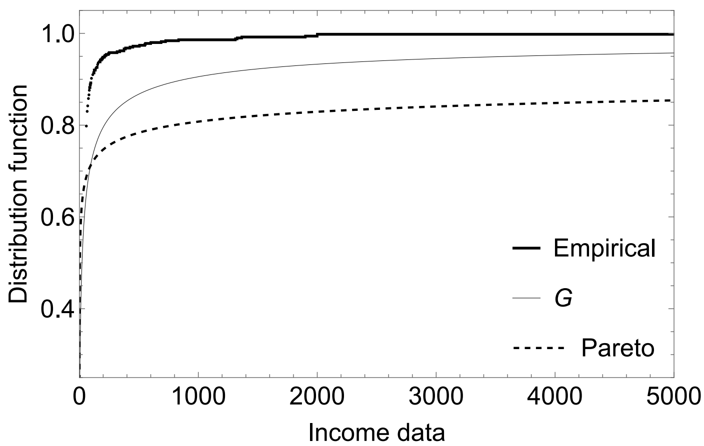

Figure 8. Plots of the empirical cdf. with estimated G cdf and estimated Pareto cdf models.

6. Discussion

This paper presents a study of the G model, used in a characterization of the gamma distribution. The G model is a special case of the prime beta model with one scale parameter. The G model has two parameters and this makes it an attractive competitor against various two-parameter models used in actuarial statistics. The G model appears to be a viable alternative for fitting data with extreme observations. Some other characteristics of the G model are:

- The smaller parameter α, the heavier the right tail of the G model.
- The G model has an explicit representation given in Proposition 1 (b).
- Cdf, risk function and quantile function are explicit and are represented by known functions.
- The VaR measurement is explicit and is used to show that the right tail of the G model is heavy.
- The applications show that the G model has its own characteristic compared with other two-parameter models, and that the G model can be a good candidate for modelling income data with a heavy tail.

Author Contributions: Conceptualization, N.M.O. and E.G.-D.; methodology, N.M.O. and E.G.-D.; software, N.M.O. and E.G.-D.; validation, N.M.O. and E.G.-D. and O.V.; formal analysis, N.M.O. and O.V.; investigation, E.G.-D. and N.M.O.; writing—original draft preparation, N.M.O. and E.G.-D.; writing—review and editing, O.V. and E.G.-D.; funding acquisition, O.V. All authors have read and agreed to the published version of the manuscript.

Funding: The research of N.M. Olmos was supported by Semillero UA-2022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is available on the web site https://www.federalreserve.gov/econres/scfindex.htm (accessed on 5 January 2022).

Acknowledgments: We thank the reviewers and the editor for their valuable comments, which undoubtedly helped to improve the presentation of our results.

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

Appendix A

The first derivatives of $\ell(\boldsymbol{\theta})$ are given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = \frac{\alpha}{\sigma} - \frac{1}{\sigma + x}, \qquad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \log(\sigma) - \log(x) + \psi(\overline{\alpha}) - \psi(\alpha).$$

The second derivatives of $l(\boldsymbol{\theta})$ are:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{\alpha}{\sigma^2} + \frac{1}{(\sigma + x)^2}, \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \alpha} = \frac{1}{\sigma}, \quad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha^2} = -\psi'(\overline{\alpha}) - \psi'(\alpha),$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions respectively.

References

1. Ibragimov, R.; Prokhorov, A. *Heavy Tails and Copulas: Topics in Dependence Modelling in Economics and Finance*; World Scientific: Singapore, 2017.
2. Olmos, N.M.; Varela, H.; Gómez, H.W.; Bolfarine, H. An extension of the half-normal distribution. *Stat. Pap.* **2012**, *53*, 875–886. [CrossRef]
3. Olmos, N.M.; Varela, H.; Bolfarine, H.; Gómez, H.W. An extension of the generalized half-normal distribution. *Stat. Pap.* **2014**, *55*, 967–981. [CrossRef]
4. Astorga, J.M.; Gómez, H.W.; Bolfarine, H. Slashed generalized exponential distribution. *Commun. Stat. Theory Methods* **2017**, *46*, 2091–2102. [CrossRef]
5. Gupta, R.D.; Kundu, D. Generalized exponential distributions. *Aust. N. Z. J. Stat.* **1999**, *41*, 173–188. [CrossRef]
6. Mudholkar, G.S.; Srivastava, D.K.; Freimer, M. The exponentiated Weibull family: A reanalysis of the Bus-Motor-Failure data. *Technometrics* **1995**, *37*, 436–445. [CrossRef]
7. Gómez, Y.M.; Bolfarine, H.; Gómez, H.W. Gumbel distribution with heavy tails and applications to environmental data. *Math. Comput. Simul.* **2019**, *157*, 115–129. [CrossRef]
8. Bhati, D.; Ravi, S. On generalized log-Moyal distribution: A new heavy tailed size distribution. *Insur. Math. Econ.* **2018**, *79*, 247–259. [CrossRef]

9. Afify, A.Z.; Gemeay, A.M.; Ibrahim, N.A. The Heavy-Tailed Exponential Distribution: Risk Measures, Estimation, and Application to Actuarial Data. *Mathematics* **2020**, *8*, 1276. [CrossRef]
10. Choulakian, V.; Stephens, M.A. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* **2001**, *43*, 478–484. [CrossRef]
11. Zhang, J. Likelihood moment estimation for the generalized Pareto distribution. *Aust. N. Z. J. Stat.* **2007**, *49*, 69–77. [CrossRef]
12. Akinsete, A.; Famoye, F.; Lee, C. The Beta-Pareto distribution. *Statistics* **2008**, *42*, 547–563. [CrossRef]
13. Nassar, M.M.; Nada, N.K. The beta generalized Pareto distribution. *J. Stat. Adv. Theory Appl.* **2011**, *6*, 1–17.
14. Mahmoudi, E. The beta generalized Pareto distribution with application to lifetime data. *Math. Comp. Simul.* **2011**, *81*, 2414-2430. [CrossRef]
15. Boumaraf, B.; Seddik-Ameur, N.; Barbu, V.S. Estimation of Beta-Pareto Distribution Based on Several Optimization Methods. *Mathematics* **2020**, *8*, 1055. [CrossRef]
16. Gleser, L.J. The Gamma Distribution as a Mixture of Exponential Distributions. *Am. Stat.* **1989**, *43*, 115–117.
17. Keeping, E.S. *Introduction to Statistical Inference*; Van Nostrand: New York, NY, USA, 1962.
18. Tiao, G.G.; Cuttman, I. The Inverted Dirichlet Distribution with Applications. *J. Am. Stat. Assoc.* **1965**, *60*, 793–805. [CrossRef]
19. Glaser, R.E. Bathtub and Related Failure Rate Characterizations. *J. Am. Stat. Assoc.* **1980**, *75*, 667–672. [CrossRef]
20. Artzner, P. Application of coherent risk measures to capital requirements in insurance. *N. Am. Actuar. J.* **1999**, *3*, 11–25. [CrossRef]
21. Artzner, P.; Delbaen, F.; Eber, J.M.; Heath, D. Coherent measures of risk. *Math. Financ.* **1999**, *9*, 203–228. [CrossRef]
22. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1978**, *27*, 379–423. [CrossRef]
23. Rolski, T.; Schmidli, H.; Schmidt, V.; Teugel, J. *Stochastic Processes for Insurance and Finance*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
24. Foss, S.; Korshunov, D; Zachary, S. *An Introduction to Heavy-Tailed and Subexponential Distributions*; Springer Series in Operations Research and Financial Engineering; Springer: New York, NY, USA, 2011.
25. Bingham, N. *Regular Variation*; Cambridge University Press: Cambridge, UK, 1987.
26. Konstantinides, D. *Risk Theory. A Heavy Tail Approach*; World Scientific Publishing: Singapore, 2018.
27. Jessen, A.H.; Mikosch, T. Regularly varying functions. *Publ. Inst. Math. Nouvelle Série* **2006**, *80*, 171–192. [CrossRef]
28. Embrechts, P.; Goldie, C.M. On closure and factorization properties of subexponential and related distributions. *J. Aust. Math. Soc.* **1980**, *29*, 243–256. [CrossRef]
29. Embrechts, P.; Goldie, C.M. On convolution tails. *Stoch. Process. Their Appl.* **1982**, *13*, 263–278. [CrossRef]
30. Boland, P. *Statistical and Probabilistic Methods in Actuarial Science*; Chapman & Hall: Boca Raton, FL, USA, 2007.
31. Hogg, R.V. and Klugman, S.A. *Loss Distributions*; John Wiley & Sons: Hoboken, NJ, USA, 1984.
32. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing; Vienna, Austria, 2021. Available online: https://www.R-project.org/ (accessed on 15 January 2022).
33. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*; Springer: New York, NY, USA, 1998.
34. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716–723. [CrossRef]
35. Schwarz, G. Estimating the dimension of a model. *Ann. Statist.* **1978** *6*, 461–464. [CrossRef]
36. Cooray, K.; Ananda, M.M.A. A generalization of the half-normal distribution with applications to lifetime data. *Commun. Stat. Theory Methods* **2008**, *37*, 1323–1337. [CrossRef]
37. Arnold, B.C. Pareto Distributions. In *Monographs on Statistics & Applied Probability*, 2nd ed.; Chapman & Hall: Boca Raton, FL, USA, 2015.