THE FULL TAILS GAMMA DISTRIBUTION APPLIED TO MODEL EXTREME VALUES

BY

JOAN DEL CASTILLO, JALILA DAOUDI AND ISABEL SERRA

ABSTRACT

In this paper, we introduce the simplest exponential dispersion model containing the Pareto and exponential distributions. In this way, we obtain distributions with support $(0, \infty)$ that in a long interval are equivalent to the Pareto distribution; however, for very high values, decrease like the exponential. This model is useful for solving relevant problems that arise in the practical use of extreme value theory. The results are applied to two real examples, the first of these on the analysis of aggregate loss distributions associated to the quantitative modelling of operational risk. The second example shows that the new model improves adjustments to the destructive power of hurricanes, which are among the major causes of insurance losses worldwide.

KEYWORDS

Risk models, exponential models, heavy tailed distributions, Pareto distribution, power-law distribution, type III distribution.

1. INTRODUCTION

Extreme value theory is used by many authors for modelling exceedances in several fields, such as hydrology, insurance, finance and environmental science, see Furlan (2010), Coles and Sparks (2006) and Moscadelli (2004). However, the theory shows some surprises in practical applications. For instance, Dutta and Perry (2006) observed, in an empirical analysis of models for estimating operational risk, that even when a Pareto distribution fits the data it may result in unrealistic capital estimates (sometimes more than 100% of the asset size), see also Degen *et al*. (2007), and it may happen that it works in the central region but not for larger values.

The peaks-over-threshold (PoT) method for estimating high quantiles is based on the Pickands–Balkema–DeHaan Theorem, see McNeil *et al*. (2005) and Embrechts *et al*. (1997). This result is the mathematical solution to the

Astin Bulletin 47(3), 895-917. doi: 10.1017/asb.2017.9

question, but there still remains the practical problem of whether the threshold is high enough. These challenges should motivate us to find a new methodology for describing the characteristics of the data rather than limit the data so that it matches the characteristics of a narrow model.

In this paper, a new statistical approach to estimating high quantiles is provided. We analyze how to be a statistical model for non-light tails data sets, see Definition 2. Of course, the Pareto distribution fulfils the definition, and we introduce a new example, here called full tails gamma (FTG) distribution, that also meets the definition. Moreover, it is shown that the Pareto and gamma distributions are nested in this statistical model, see Theorem 6.

The gamma distribution is one of the most studied families of distributions, since Fisher (1922) used it to show that maximum likelihood estimation (MLE) is more efficient that the method of moments. For life theory and reliability usually the two-parameter right-truncated gamma distribution is considered, since Chapman (1956), which also refers the difficulty of considering the unknown origin. Den Broeder (1955) considered the left-truncated gamma distribution but with known scale parameter. Stacy (1962) introduced a three-parameter generalized gamma distribution that includes, as special cases, the two-parameter gamma and the two-parameter Weibull. Harter (1967) extends the model to a four-parameter family, by including a location parameter. Hegde and Dahiya (1989) obtain necessary and sufficient conditions for the existence of the MLE of the parameters of a right-truncated gamma distribution. The right-truncated gamma with unknown origin is a non-regular model, hence MLE is not necessarily the best method of estimation and the uniformly minimum variance unbiased estimator may be better, see Dixit and Phal (2005). For simulation of right- and left-truncated gamma distributions, see Philippe (1997). The FTG distribution is related to very old families of distributions such as the Pareto type III distribution, see Arnold (1983, pp 3), Davis, *et al.* (1979).

The FTG is the simplest exponential dispersion model containing the Pareto and exponential distribution, see Corollary 4. The likelihood theory for FTG distribution is studied in this paper, and it is applied to two illustrative data sets. With the current specialized computer programs for statistical analysis is not difficult to deal with the FTG distribution, since the incomplete gamma function and its derivatives are now easily available, see Abramowitz and Stegun (1972). The work of pioneers like Chapman (1956) must be viewed in this way. This new model allows us to find distributions close to the Pareto distribution in a long range but with exponential decrease for very large values. Another approach for lighter tails is in Akinsete *et al.* (2008).

The FTG distribution, introduced in Section 2, is a scale parameter family closed by taking conditional distributions over a threshold. This fact provides a clear interpretation of its three parameters (α, θ, ρ). The FTG distribution for $\alpha > 0$ is the left-truncated gamma distribution relocated to the origin, hence all distributions have support in $(0,\infty)$. The FTG distribution for $\alpha \leq 0$ appears as the full exponential model generated from a canonical statistic, see Barndorff-Nielsen (1978), Brown (1986) and Letac (1992).

Section 3 describes the most basic statistical properties of the FTG, as the moment-generating function, a simulation method and the standard tools for MLE. In Section 4, using an actual data set, we show how to apply the FTG to the analysis of aggregate loss distributions associated to the quantitative modelling of operational risk; see Degen *et al.* (2007). Unlike the Pareto distribution, the new model provides realistic and highly stable risk-capital estimation. Section 5 demonstrates that FTG is better than the Pareto distribution for modelling the most destructive hurricanes, which are among the major causes of insurance losses worldwide. The most prominent fact is that, unlike other competitors, FTG does not contradict the general theory, since it contains the Pareto distribution.

FTG是gamma分布的右尾条件分布：dgamma(x+u)/[1-pgamma(u)];u=pho/theta

2. The full tails gamma distribution

The *FTG distribution* is the three-parameter family of continuous probability distributions, with support on $(0, \infty)$, defined for instance by $\alpha \in \mathbb{R}$, $\theta > 0$, $\rho > 0$ by

$$f(x; \alpha, \theta, \rho) = \theta (\rho + \theta x)^{\alpha-1} \exp(-(\rho + \theta x))/\Gamma(\alpha, \rho), \tag{1}$$

where $\Gamma(\alpha, \rho)$ is the (upper) incomplete gamma function, see Abramowitz and Stegun (1972),

$$\Gamma(\alpha, \rho) = \int_{\rho}^{\infty} t^{\alpha-1} e^{-t} dt, \tag{2}$$

in particular $\Gamma(\alpha, 0) = \Gamma(\alpha)$ is the gamma function. The FTG distribution extends to some boundary parameters as will be seen in this Section.

If $\alpha > 0, \theta > 0$ and $\rho = 0$, the family (1) clearly extends to the probability density function of the gamma distribution, defined by

$$g(x; \alpha, \theta) = \theta^{\alpha} x^{\alpha-1} \exp(-\theta x) / \Gamma(\alpha). \tag{3}$$

For $\alpha > 0$ the FTG is the left-truncated gamma distribution relocated to the origin. Note that in this paper, for simplicity, *tail* is used in the sense of conditional exceedances over a threshold. Proposition 1 establishes exactly what we say.

Suppose $F(x)$ is an absolutely continuous cumulative distribution function of a non-negative random variable, X. The probability density function of the exceedances of X at $u > 0$ is defined by

$$f_u(x) = f(x+u)/(1 - F(u)), \tag{4}$$

where $f(x) = F'(x)$. The values of $(X - u)$ conditional on $X > u$ are denoted by X_u and it will be said that $f_u(x)$ is the probability density function of the tail of $f(x)$ at u. Often, we write the threshold as u and the threshold exceedances as X_u.

Proposition 1. *If $\alpha > 0$ and $\rho > 0$, then the FTG distribution (1) is the probability density function of the exceedances of a gamma distribution at $\sigma > 0$, with $\sigma = \rho/\theta$.*

$$g_\sigma(x; \alpha, \theta) = \theta^\alpha (x+\sigma)^{\alpha-1} \exp(-\theta(x+\sigma)) / \Gamma(\alpha, \sigma\theta). \tag{5}$$

Proof. Let $G(x)$ be the cumulative distribution function corresponds to $g(x; \alpha, \theta)$, see (3). Consider $\sigma > 0$. Using $dt = \theta dx$, we obtain

$$1 - G(\sigma) = \int_\sigma^\infty \theta^\alpha x^{\alpha-1} \exp(-\theta x)\, dx / \Gamma(\alpha) = \int_{\sigma\theta}^\infty t^{\alpha-1} \exp(-t) dt / \Gamma(\alpha).$$

Then, the probability density function of the tail of $g(x; \alpha, \theta)$ at σ is (5), since (4) and (2). Finally, for $\sigma = \rho/\theta$ we obtain (1). ■

The FTG distribution is related with the Pareto type III model from Arnold (1983, pp3), characterized by the survivor function

$$S(t) = (1 + t/\phi)^{-\lambda} \exp(-\theta t). \tag{6}$$

Proposition 2. *Let $p(\theta, \lambda, \phi)$ be the probability density function of the Pareto type III and f the probability density function of $FTG(\alpha, \theta, \rho)$, see (1) then*

$$p(\theta, \lambda, \phi) = c\, f(-\lambda, \theta, \theta\phi) + (1-c)\, f(-\lambda+1, \theta, \theta\phi), \tag{7}$$

where $c = \lambda(\theta\phi)^\lambda \exp(\theta\phi)\Gamma(-\lambda, \theta\phi)$.

Proof. The probability density function p is minus the derivative of (1). To verify (7) using the property

$$-a\Gamma(a, b) + \Gamma(a+1, b) = \exp(-b) b^a$$

for all a, b. This property is obtained from integration by parts applied to the definition of $\Gamma(a, b)$, see (2). ■

Remark that Pareto type III model is a particular case of the six-parameter mixture model of two FTG models.

2.1. FTG as an exponential model

A full exponential model, generated by *Lebesgue* measure on $[0, \infty)$, with canonical statistic $T(x)$ is the set of all densities

$$\exp(\theta \cdot T(x)) / C(\theta) \tag{8}$$

for $\theta \in D$, where D denotes the largest set of parameters such that the Laplace transform

$$C(\theta) = \int_0^\infty \exp(\theta \cdot T(x))\, dx \tag{9}$$

converges. D is called the natural domain of parameters. If D is an open set, then the likelihood equations have one and only one solution provided the observation is in the domain of the means, Barndorff-Nielsen (1978, pp 238). The *domain of the means* is the image of the interior of D by the gradient map

$$\theta \mapsto \nabla k(\theta), \tag{10}$$

where $k(\theta) = \log C(\theta)$. Given a sample $\mathbf{x} = \{x_1, \ldots, x_n\}$ such that the sample value of the statistic T, $t(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} T(x_i)$, is in the interior of the domain of the means, then the maximum likelihood estimator for the sample is in the interior of the natural domain of parameters.

Example 1. *The gamma family of probability density function defined by* (3) *is the exponential model on* $(0, \infty)$*, generated by Lebesgue measure on* $(0, \infty)$ *with sufficient statistics* $(x, \log(x))$ *and the domain of parameters is exactly:* $\{(\alpha, \theta) : \alpha > 0,\ \theta > 0\}$.

In general, given an exponential model with sufficient statistic $T(x)$ and given $\sigma > 0$ a fixed threshold, then we can consider a new exponential model with statistic $T(1 + x/\sigma)$. Remark, this model contains the densities of the tails of densities in $T(x)$ at σ. Applied to the Example 1, we obtain the FTG model. Exactly, for $\sigma > 0$ fixed, the FTG corresponds to the full exponential model generated by Lebesgue measure on $(0, \infty)$ with sufficient statistic $T(x) = (x/\sigma, \log(1 + x/\sigma))$. The goal is that the natural domain of parameters of FTG is bigger than the extension of gamma model.

Proposition 3. *Let* $\sigma > 0$ *fixed in* (5)*, the Laplace transform of the sufficient statistics* $(x, \log(x + \sigma))$

$$C(\theta, \alpha - 1) = \int_0^\infty \exp(-\theta x + (\alpha - 1)\log(x + \sigma))\, dx$$

converges for the set of parameters given by $\Theta = \{\alpha \in \mathbb{R}, \theta > 0\}$ *and by* $\Theta_0 = \{\alpha < 0, \theta = 0\}$.

Proof. First of all assume $\theta > 0$. Taking $y = x + \sigma$,

$$C(\theta, \alpha - 1) = e^{\theta\,\sigma} \int_\sigma^\infty \exp(-\theta y)\, y^{\alpha-1} dy = \theta^{-\alpha}\, e^{\theta\,\sigma}\, \Gamma(\alpha, \theta\,\sigma),$$

where the integral converges for $\alpha \in \mathbb{R}$, since there is no singularity at zero and the exponential dominates any power function for large values of x.

If $\theta = 0$, then for $\alpha < 0$ the following integral is also convergent:

$$C(0, \alpha - 1) = \int_0^\infty (x + \sigma)^{\alpha-1}\, dx = -\sigma^\alpha/\alpha.$$

■

Proposition 3 shows that given $\sigma > 0$ the natural domain of parameters for (5) is $D = \Theta \cup \Theta_0$. The FTG distribution (1) is the extension of (5) for $\alpha < 0$ and it has been seen that it extends to the boundary parameters Θ_0.

Finally, given a sample $\mathbf{x} = \{x_1, \ldots, x_n\}$, to compute the MLE we can use the procedure: to determine σ and if $t(\mathbf{x})$ is in the interior of the domain of the means, use it to solve the likelihood equations. Therefore, we have to determine the domain of the means.

Example 2. *The probability density function of the Pareto distribution, defined for $x > 0$, is*

$$p(x; \alpha, \sigma) = -\alpha\sigma^{-1}(1 + x/\sigma)^{\alpha-1}, \tag{11}$$

where $\alpha < 0$ and $\sigma > 0$. For fixed σ, the Pareto model is a full exponential model with sufficient statistic $log(1 + x/\sigma)$. This parameterization will be used later to show that the Pareto distribution appears in this boundary of the FTG distribution.

Corollary 4. *Given $\sigma > 0$, the FTG is the smallest full exponential model containing the Pareto distribution with σ fixed and the exponential distribution. That is, the FTG is the smallest full exponential dispersion model containing the Pareto and exponential distribution. The FTG model is a model for tails.*

Remark that the Corollary 4 characterizes the FTG as the simplest exponential dispersion model for modelling non-light tails.

Proposition 5. *The domain of the means of FTG is*

$$\{(x, y)\ ;\ x > 0, \log(1 + x) > y > x/(1 + x)\},$$

for all parameters α, θ and ρ.

Proof. The model FTG is scale invariant, so it suffices to consider $\rho/\theta = \sigma = 1$, and following the results of Castillo and Puig (1999), we write

$$T(x) = (x, \log(1 + x)) = (x, S(x)).$$

Since $-S(x)$ is convex function, $\Theta = \mathbb{R}^+ \times \mathbb{R}$ and $\Theta_0 = \{0\} \times \mathbb{R}^+$, then by the Proposition 4 in Castillo and Puig (1999)

$$S(\mu) \geq E(S(X)) \geq E(S(X_\mu^0)), \tag{12}$$

where X_μ^0 is a random variable with probability density function function of the form $f(x; b) = \exp(-(b)\, S(x))/C(b)$, $b > 0$, such that $E(X_\mu^0) = \mu$. Given a sample $\mathbf{x} = \{x_1, \ldots, x_n\}$ of FTG distribution the likelihood equations can be described in terms of the sufficient statistic $(\bar{x}, \bar{s})$ by

$$E(X) = \frac{\partial k}{\partial a}(\theta, \beta) = \overline{x}, \text{ and } E(S(X)) = \frac{\partial k}{\partial p}(\theta, \beta) = \overline{s},$$

where $\beta = -(\alpha - 1) > 1$. Consider a random variable X^0 with the Pareto distribution, so from (11) the probability density function is

$$p(x; \beta, 1) = \beta\,(1+x)^{-\beta-1} = \exp(-\beta\, S(x))/C(\beta).$$

Then,

$$E(X^0) = \int_0^\infty \beta\, x\,(1+x)^{-\beta-1} dx = \frac{1}{\beta - 1}.$$

For each sample, exist $\beta > 1$ such that

$$E(X^0) = \frac{1}{\beta - 1} = \bar{x} \Leftrightarrow \beta = \frac{\overline{x}+1}{\overline{x}}.$$

Finally, the domain of the means described in (12) corresponds to

$$\log(1+\overline{x}) \geq \overline{s} \geq \frac{\overline{x}}{1+\overline{x}},$$

since $E(S(X^0_{\overline{x}})) = \int_0^\infty \beta\, \log(1+x)\,(1+x)^{-\beta-1} dx = 1/\beta$ for all β. ■

2.2. Models for tails: FTG and Pareto distributions

The Pareto distribution (11) is related to the *power-law distribution*, Sornette (2006), Clauset (2009), with probability density function

$$p_w\,(y; \alpha, c) = -\alpha c^{-1}\,(y/c)^{\alpha-1}\,,\ \ y > c. \tag{13}$$

It seems as if taking $y = (x + c)$ the two models are the same, but this is not true. Note that the probability density functions in (13) have variable support; however, in (11) the support is fixed at $(0, \infty)$. A sample from (11) consists of positive numbers, but a sample from (13) consists of real numbers greater than c. If c is known, the change of variable $y = (x + c)$ gives an one-parameter subfamily of (11).

Likewise, given a statistical model, $\mathcal{M} = \{ f(x; \theta)\}$, of probability density functions on a sample space $S \subset \mathbb{R}$, there is also a clear difference between the left-truncated distributions of $\mathcal{M}$ over a threshold c and the tail distributions of $\mathcal{M}$. For instance, for $[c, \infty) \subset S$, the truncated distributions of $\mathcal{M}$ over the threshold c are a set of probability density functions all of them defined on $[c, \infty)$, where c is fixed. Hence, for a threshold $u > c$, the probability density function of the exceedances at u defined by (4) does not necessarily belongs to the truncated family, even if the corresponding distribution is translated to c. If we denote by $\mathcal{M}_c$ the truncated distributions of $\mathcal{M}$ over the threshold c, translated to 0, then the tails model associated to $\mathcal{M}$ and c is $\cup_{u \geqslant c} \mathcal{M}_u$. In some cases the two concepts agree, for instance in the Pareto distribution or the exponential distribution, which are scale models and, moreover, truncation is a change of scale.

Definition 1. *Let $\mathcal{M}$ be a statistical model on a sample space, $S \subset \mathbb{R}$, and $c \in S$. The tails model associated to $\mathcal{M}$ and c is the set of probability density functions of the exceedances at u defined by* (4) *for any $u > c$ and for each $f \in \mathcal{M}$.*

For $\alpha \leq 0$, the FTG distribution (1) has not been carefully studied yet, to our knowledge, and we shall proceed to do so.

Theorem 6. *Let $\sigma = \rho/\theta > 0$ fixed in* (1) *and $\alpha < 0$. If ρ tends to zero, then the probability density function* (1) *tends to the probability density function of the Pareto distribution (11) in L^1 norm. Moreover, the convergence extends to the moments, provided the corresponding moments for Pareto distribution are finite.*

Proof. Observe that if ρ tends to 0, then θ tends to 0, since $\sigma = \rho/\theta > 0$ is fixed. Using $\theta = \rho/\sigma$,

$$f(x; \alpha, \theta, \rho) = \rho^{\alpha}\sigma^{-1}(1 + x/\sigma)^{\alpha-1} \exp(-\rho(1 + x/\sigma))/\,\Gamma(\alpha, \rho)$$

converges pointwise to the probability density function (11), since the property (5.1.23) of Abramowitz and Stegun (1972), under the assumptions,

$$\lim_{\rho \to 0} \rho^{\alpha}/\,\Gamma(\alpha, \rho) = -\alpha \tag{14}$$

holds. Observe that for ρ small

$$\begin{aligned} f(x; \alpha, \theta, \rho) &= \rho\sigma^{-1}(1 + x/\sigma)^{\alpha-1} \exp(-\rho(1 + x/\sigma))/\,\Gamma(\alpha, \rho) \\ &\leq -2\alpha\sigma^{-1}(1 + x/\sigma)^{\alpha-1} = 2p\,(x; \alpha, \sigma)\,, \end{aligned}$$

since from the limit (14) we can consider the boundedness $\rho^{\alpha}/\,\Gamma(\alpha, \rho) \leq -2\alpha$. Finally, from the dominated convergence theorem we obtain convergence in L^1. Moreover, whenever the moments of Pareto distribution are finite, the convergence extends to these moments. ■

The family (1) has been extended to the boundary parameter sets corresponding to the gamma distribution and the Pareto distribution, $\{\alpha > 0, \theta > 0, \rho = 0\}$ and $\{\alpha < 0, \theta = 0, \sigma > 0\}$, respectively. Hence, it can be assumed that the family (1) includes the gamma distribution, the truncated gamma distribution ($\alpha > 0$), its extension to a full exponential model ($\alpha < 0$) and the Pareto distribution, see Figure 1. We will see that the FTG distribution can be used for modelling tails in the sense of the next definition.

Definition 2. *Let us define $\mathcal{M}$ a model for (non-light) tails if the following conditions hold:*

1. *$\mathcal{M}$ is equal to the tails model associated to some model model $\mathcal{M}'$ and some $c \in \mathbb{R}$.*
2. *$\mathcal{M}$ is a scale model.*
3. *The exponential and Pareto distributions* (11) *belong to $\mathcal{M}$.*

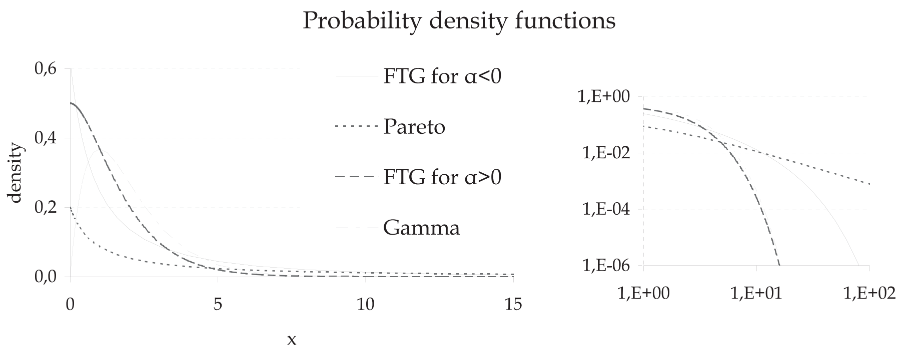

FIGURE 1: The Figure in the left shows some probability density functions in FTG family. The FTG with $\alpha = 2, \sigma = 1$ and $\theta = 1$ corresponds to the tails of a gamma and the FTG for $\alpha = -0.2, \sigma = 1$ and $\theta = 0.1$ is not. For the boundary parameter sets of the family, we consider gamma with $\alpha = 2$ and $\theta = 1$ and Pareto with $\alpha = -0.2$ and $\sigma = 1$. The picture in the right shows the same plot in common logarithm for the tail of the functions. We see exponential decay except for Pareto probability density function.

Proposition 7. *If X is a random variable distributed as $FTG(\alpha, \theta, \rho)$, then*

a. For $\lambda > 0$, the random variable λX is distributed as $FTG\,(\alpha, \theta/\lambda, \rho)$.
b. For any threshold, $u > 0$, the threshold exceedances, X_u is distributed as $FTG(\alpha, \theta, \rho + \theta u)$.

Proof. The first result holds from the probability density function of λX for $\lambda > 0$,

$$\begin{aligned} f\,(x/\lambda; \alpha, \theta, \rho)\,/\lambda &= (\theta/\lambda)\,(\rho + \theta x/\lambda)^{\alpha-1} \exp(-(\rho + \theta x/\lambda))/\,\Gamma(\alpha, \rho) \\ &= f(x; \alpha, \theta/\lambda, \rho); \end{aligned}$$

remark that for $\theta = 0$

$$p\,(x/\lambda; \alpha, \sigma)\,/\lambda = -\alpha(\lambda\sigma)^{-1}\,(1 + x/(\lambda\sigma))^{\alpha-1} = p(x; \alpha, \lambda\sigma).$$

And the second one is a consequence of (4) . For $\theta > 0$,

$$\begin{aligned} \frac{f\,(x + u; \alpha, \theta, \rho)}{1 - F(u)} &= \frac{\theta\,(\rho + \theta u + \theta x)^{\alpha-1}}{\Gamma(\alpha, \rho + \theta u)} \exp(-(\rho + \theta u + \theta x)) \\ &= f(x; \alpha, \theta, \rho + \theta u) \end{aligned}$$

and for $\theta = 0$,

$$\begin{aligned} \frac{p\,(x + u; \alpha, \sigma)}{1 - P(u)} &= -\frac{\alpha\,(1 + u/\sigma + x/\sigma)^{\alpha-1}}{\sigma(1 + u/\sigma)^{\alpha}} = -\frac{\alpha\,(1 + x/(\sigma + u))^{\alpha-1}}{(\sigma + u)^{-1}} \\ &= p(x; \alpha, \sigma + u). \end{aligned}$$

■

From the last result, it is clear that the FTG distribution (1) is a scale parameter family on $(0, \infty)$ closed by truncation, in the sense of (4). Hence, it is appropriate for modelling tails. The parameter $\beta = 1/\theta$ is the scale parameter and the parameter ρ is the truncation parameter. The parameter $(-\alpha)$ shall be interpreted in terms of the Pareto distribution as the weight of the tail, since the model includes the Pareto distribution as a limit case. Then, each one of the three-parameter separately has a clear interpretation.

It can also be seen (1) as a weighted version of the Pareto distribution, with the weight $w(x) = \exp(-\theta x)$, that is also known as a *exponential tilting* of the distribution, see Barndorff-Nielsen and Cox (1994).

Corollary 8. *The FTG model is a model for tails.*

Proof. Let $\mathcal{M}$ denote the FTG model. By Proposition 7, FTG model is an scale model and $f_u \in \mathcal{M}$, for all $f \in \mathcal{M}$ and for all $u > 0$. In the other hand, for arbitrary (α, θ, ρ) parameters of FTG and for any ϵ such that $0 < \epsilon < \rho/\theta$, if $u = \rho/\theta - \epsilon$ and g is the FTG distribution with parameters $(\alpha, \theta, \rho - \epsilon\theta)$, then g_u is the FTG distribution with parameters (α, θ, ρ) , since Proposition 7. Then $\mathcal{M} = \cup_{u\geqslant 0}\mathcal{M}_u$, where $\mathcal{M}_u = \{ f_u \mid f \in \mathcal{M} \}$, i.e. the tails model associated to FTG model and $0 \in \mathbb{R}$ is exactly the FTG model. Finally, by Theorem 6, the exponential and Pareto distributions (11) belong to $\mathcal{M}$. ■

Proposition 9. *The cumulative distribution function corresponding to the family* (1) *is*

$$F(x; \alpha, \theta, \rho) = 1 - \Gamma(\alpha, \rho + \theta x)/\Gamma(\alpha, \rho)$$

and for the Pareto distribution we have to consider the limit case, corresponding to $P(x; \alpha, \sigma) = 1 - (1 + x/\sigma)^{\alpha}$.

Proof. For $\theta > 0$, taking $y = \rho + \theta t$

$$F(t; \alpha, \theta, \rho) = \int_0^{\rho+\theta t} y^{\alpha-1} \exp(-y)dy/\Gamma(\alpha, \rho) = 1 - \Gamma(\alpha, \rho + \theta t)/\Gamma(\alpha, \rho).$$

Remark that for $\rho = 0$ holds since $\Gamma(\alpha, \rho) = \Gamma(\alpha)$ and that is corresponding to the gamma distribution. ■

3. Statistical tools and MLE

With the current specialized computer programs for statistical analysis it is not difficult to deal with the FTG distribution. The incomplete gamma function, $\Gamma(\alpha, \rho)$ and its derivatives are now easily available. Symbolic differentiation allows us to get the moments of a distribution from the moment-generating function. Simulation and optimization algorithms are available in the same way. The work of pioneers like Chapman (1956) must be viewed in this way.

3.1. Moments-generating function

The FTG distribution has moment-generating function in the interior of the domain of parameters. Hence, it is possible to calculate the moments of all orders. In addition, it is also possible to calculate the moments of the conditional distribution over a threshold, by Proposition 7.

Proposition 10. *For $\alpha \in \mathbb{R}, \theta > 0, \rho > 0$, the moment-generating function of the FTG distribution* (1) *exists and is given by*

$$\begin{aligned} M(t) &= M(t; \alpha, \theta, \rho) \\ &= (1 - t/\theta)^{-\alpha} \exp(-\rho t/\theta)\, \Gamma(\alpha, \rho(1 - t/\theta))/\, \Gamma(\alpha, \rho), \quad t < \theta. \end{aligned} \tag{15}$$

For $\alpha > 0$, it extends for $\rho = 0$ and coincides with the moment-generating function of gamma distribution $M_g(t) = (1 - t/\theta)^{-\alpha}$.

Proof. By definition $M(t) = E[\exp(tX)]$ is given by

$$M(t) = \int_0^\infty \frac{\theta\,(\rho + \theta x)^{\alpha-1} \exp[-(\rho + (\theta - t)\,x)]}{\Gamma(\alpha, \rho)} dx,$$

and using $\rho + \theta dx - t\rho/\theta - tdx = dy$, we obtain

$$M(t) = (1 - t/\theta)^{-\alpha} \exp(-\rho t/\theta)\, \Gamma(\alpha, \rho(1 - t/\theta))/\, \Gamma(\alpha, \rho),$$

for $t < \theta$. ∎

The cumulant-generating function is given by

$$\begin{aligned} K(t) = \log(M(t)) &= -t\rho/\theta - \alpha \log(1 - t/\theta) - \log \Gamma(\alpha, \rho) \\ &\quad + \log \Gamma(\alpha, (1 - t/\theta)\,\rho). \end{aligned}$$

Hence, the first moments are

$$\begin{aligned} E[X] &= K'(0) = (\alpha - \rho + \mu)/\theta, \\ Var[X] &= K''(0) = (\alpha + (1 + \rho - \alpha)\mu - \mu^2)/\theta^2, \end{aligned}$$

where $\mu = e^{-\rho}\rho^\alpha / \Gamma(\alpha, \rho)$. Notice that using Proposition 7, to calculate the conditional expectation for any threshold fixed $u > 0$, is the same as to calculate the expectation with modified parameters

$$E[X \mid X > u] = (\alpha - \rho + \mu')/\theta, \tag{16}$$

where $\mu' = e^{-(\rho+\theta u)}(\rho + \theta u)^\alpha / \Gamma(\alpha, \rho + \theta u)$.

3.2. Random variates generation

Simulation methods for Pareto and gamma distributions are well known. The simulation of truncated gamma distribution (5) has also been well studied, see

Philippe (1997). Hence, only the set of parameters $\{\alpha < 0, \theta > 0, \rho > 0\}$ for FTG distribution is considered here.

A simple way to simulate the distribution is the inversion method, since the cumulative distribution function has an easy expression, see Proposition 9; however, it needs to use complex numerical processes using the incomplete gamma function.

A simple and efficient method from a numerical point of view is obtained with an idea from Devroye (1986) on a generalization of the rejection method. We emphasize the simplicity of this algorithm, since it does not require the use of the incomplete gamma function.

First of all, since $1/\theta$ is a scale parameter it is enough consider simulations for $\theta = \rho$. That is, to simulate $FTG(\alpha, \theta, \rho)$, we can first simulate $FTG(\alpha, \rho, \rho)$ and finally we apply the change of scale to the random sample.

For $\theta = \rho$, the probability density function (1) splits in three terms

$$f(x; \alpha, \rho, \rho) = \underbrace{\left(\rho^{\alpha-1} e^{-\rho} / \Gamma(\alpha, \rho)\right)}_{\text{常数}} \underbrace{\left(\rho e^{-\rho x}\right)}_{\text{指数分布}} (1+x)^{\alpha-1} = c g(x) \psi(x), \tag{17}$$

where the function $\psi(x) = (1+x)^{\alpha-1}$ is $[0, 1]$-valued, $g(x) = \rho e^{-\rho x}$ is a probability density function easy to simulate and c is a normalization constant at least equal to 1.

The rejection algorithm for this case can be rewritten as follows. Generate independent random variates (X, U), where X has probability density function $g(x)$ and U is uniformly distributed in $[0, 1]$ until $U \leq \psi(X)$. This method produces a random variable X with probability density function $f(x)$, (Devroye, 1986).

The following code applies the method to our case, see R Development Core Team (2010).

```
#to generate a sample of size n of FTG(a,t,r)
rFTG<-function(n,a,t,r) {
     sample<-c(); m<-0
     while (m<n) {
          x<-rexp(1,rate=r);u<-runif(1)
          if (u<=(1+x)^(a-1)) sample[m+1]<-x
          m<-length(sample) }
     sample*r/t }
```

3.3. Maximum likelihood estimates of the parameters

In (1), FTG distribution has been introduced with parameters (α, θ, ρ), since each one separately has a clear interpretation. For MLE estimation it is better to use (α, σ, ρ), with dispersion parameter $\sigma = \rho/\theta$, since for fixed σ the FTG distribution is an exponential model. Hence, from Barndorff-Nielsen (1978), it is known that the maximum likelihood estimator exists and is unique. Really, the FTG distribution is an exponential dispersion model, but generalizing a *non-*

natural exponential model. The summary of the procedure to compute the MLE is search the dispersion parameter and then to optimize the problem for others.

For $x = \{x_1, \ldots, x_n\}$ of size n, the log-likelihood function for FTG distribution is

$$l(\alpha, \sigma, \rho) = -n \left(\log \Gamma(\alpha, \rho) + \log\left(\sigma \rho^{-\alpha}\right) - \frac{\alpha - 1}{n} \sum_{i=1}^{n} \log\left(1 + \frac{x_i}{\sigma}\right) + \frac{\rho}{n} \sum_{i=1}^{n} \left(1 + \frac{x_i}{\sigma}\right) \right) \tag{18}$$

To simplify, we denote

$$d = d(\alpha, \rho) = \log \Gamma(\alpha, \rho), \tag{19}$$

and we consider (r, s) the sufficient statistics of the exponential model for σ fixed as

$$r(x; \sigma) = (1 + x/\sigma) \quad \text{and} \quad s(x; \sigma) = \log(1 + x/\sigma).$$

Then, we denote the sample means by

$$\bar{r}(\mathrm{x}; \sigma) = \frac{1}{n} \sum_{i=1}^{n} (1 + x_i/\sigma) \quad \text{and} \quad \bar{s}(\mathrm{x}; \sigma) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + x_i/\sigma).$$

To simplify, we use the parameters in subscript to denote the partial derivatives, and we omit the dependence of the parameters in these derivatives. Hence, the scoring is $(l_\alpha, l_\sigma, l_\rho)$ and is given by

$$l_\alpha = -n \{d_\alpha - \log(\rho) - \bar{s}(\mathrm{x}; \sigma)\}, \tag{20}$$

$$l_\sigma = -n \left\{\sigma^{-1} - (\alpha - 1)\bar{s}_\sigma + \rho \bar{r}_\sigma\right\}, \tag{21}$$

$$l_\rho = -n \left\{d_\rho - \alpha \rho^{-1} + \bar{r}(\mathrm{x}; \sigma)\right\}, \tag{22}$$

the observed information matrix is given by

$$I_O(\alpha, \sigma, \rho) = -n \begin{pmatrix} d_{\alpha\alpha} & -\bar{s}_\sigma & d_{\alpha\rho} - \rho^{-1} \\ -\bar{s}_\sigma & -\sigma^{-2} - (\alpha - 1)\bar{s}_{\sigma\sigma} + \rho \bar{r}_{\sigma\sigma} & \bar{r}_\sigma \\ d_{\alpha\rho} - \rho^{-1} & \bar{r}_\sigma & d_{\rho\rho} + \alpha \rho^{-2} \end{pmatrix},$$

and it can be used to compute the confidence interval for the maximum likelihood estimates $\hat{\alpha}$, $\hat{\sigma}$ and $\hat{\rho}$ of the parameters α, σ and ρ, respectively. In fact, the expected information Fisher matrix can be computed in closed form, see the appendix.

To compute the MLE is convenient to solve the Equation (21) for to get $\hat{\sigma}$ using $(\hat{\alpha}(\sigma), \hat{\rho}(\sigma))$ for the parameters (α, ρ) or, more general, to maximize the profile log-likelihood equation

$$l_p(\sigma) = -n \left(\log \Gamma\left(\hat{\alpha}(\sigma), \hat{\rho}(\sigma)\right) + \log\left(\sigma \hat{\rho}(\sigma)^{-\hat{\alpha}(\sigma)}\right) - (\hat{\alpha}(\sigma) - 1)\bar{s}(\mathrm{x}, \sigma) \right.$$
$$\left. + \hat{\rho}(\sigma)\bar{r}(\mathrm{x}, \sigma)\right), \tag{23}$$

where $(\hat{\alpha}(\sigma), \hat{\rho}(\sigma))$ is the only one solution of the system in (α, ρ) consisting of the Equations (20) and (22) for σ fixed, since FTG is an exponential model in this situation. Remark that, from a practical point of view, is convenient to consider this pair of equations to simplify the Equation (23) (or (21)) in an equation as simple as possible. To do that, we will need to consider the relations (A2) found in the appendix. For instance, the Equation (23) can be simplified by

$$l_p(\sigma) = -n \left(\log \Gamma\left(\hat{\alpha}(\sigma), \hat{\rho}(\sigma)\right) - \log\left(\hat{\rho}(\sigma)\sigma^{-1}\right) - (\hat{\alpha}(\sigma) - 1)d_\alpha \right.$$
$$\left. - \hat{\rho}(\sigma)d_\rho + \hat{\alpha}(\sigma)\right). \tag{24}$$

Remark that this expression does not involve the sample explicitly.

A procedure to obtain the MLE in R is computing the MLE of the standardized sample $y = \{x_i/\bar{x}\}_{1\leq i\leq n}$, considering the initial estimates as follows. We have two options: to take the initial estimates as $(\dot{\alpha}, 1, \dot{\theta})$, where $(\dot{\alpha}, \dot{\theta})$ is the MLE of gamma model or take the initial estimates as $(\dot{\alpha}, \dot{\sigma}, \dot{\rho})$, where $(\dot{\alpha}, \dot{\sigma})$ is the MLE of Pareto model and $\dot{\rho}$ is obtained by the relation (from the Equation (22))

$$d_\rho - \dot{\alpha}\dot{\rho}^{-1} + 1 + \dot{\sigma}^{-1} = 0.$$

Finally, $(\hat{\alpha}, \hat{\sigma}, \hat{\rho})$ (the MLE for the sample x) is obtained using the Proposition 7, in fact we obtain $\hat{\alpha} = \hat{\alpha'}$, $\hat{\sigma} = \hat{\sigma'}/\bar{x}$ and $\hat{\rho} = \hat{\rho'}$.

Some additional comments have to be considered. For instance, it might be appropriate to consider log-scale for σ and ρ. R has an optimization package, which greatly simplifies the calculation the MLE. If we use Maple, then we have to be careful with a big sample size.

4. Analysis of aggregate loss distributions

Financial institutions use internal and external loss data in order to compare several approaches for modelling *aggregate loss distributions*, associated to quantitative modelling of operational risk, see Dutta and Perry (2006), Degen *et al*. (2007) and Moscadelli (2004). The data used for the analysis was collected by several banks participating in the survey to provide individual gross operational losses above a threshold, starting in 2002. The data was grouped by eight standardized *business lines* and seven *event types*.

Risk capital is measured as the 99.9% percentile level of the simulated capital estimates for aggregate loss distributions in holding period (1 year). A *loss event*

L_i (also known as the loss severity) is an incident for which an entity suffers damages that can be measured with a monetary value. An aggregate loss over a specified period of time can be expressed as the sum

$$S = \sum_{i=1}^{N} L_i, \tag{25}$$

where N is a random variable representing the frequency of losses that occur over the period. As usual, here it is assumed that the L_i are independent and identically distributed, and each L_i is independent of N, that is Poisson distributed, with parameter λ.

The data set used here corresponds to the 40 largest losses associated with the business line *corporate finance* and the event type *external fraud*, observed over a threshold, $u = \$20,000$. To simplify, the data $\{x_j\}$ has been scaled to threshold zero and mean 100, according to

$$y_j = 100\left(\frac{x_j - u}{\bar{x} - u}\right), \tag{26}$$

where $u = 20,000$ and $\bar{x} = 942211.3$.

The 40 exceedances, rounded to two decimal place, were: 0.07, 0.11, 0.26, 0.40, 0.46, 0.62, 0.70, 0.75, 0.89, 1.08, 1.52, 1.64, 1.69, 2.04, 2.19, 2.52, 2.73, 3.16, 3.74, 4.04, 4.63, 5.44, 5.86, 6.02, 10.32, 19.63, 29.13, 30.36, 30.88, 35.78, 40.07, 46.12, 137.52, 237.05, 311.14, 314.19, 396.29, 552.48, 864.88, 891.62.

Aggregate losses are determined mainly by the extreme values of loss events distribution. In this case, risk capital depends on 40 exceedances, but, to calculate the 99.9% quantile, a model is required. Under the PoT approach extreme values are modelled with Pareto distribution, see Degen *et al*. (2007) and Moscadelli (2004). The Pickands–Balkema–DeHaan theorem justifies the approach, see McNeil *et al*. (2005). However, this approach may result in unrealistic capital estimates, especially when the fitted Pareto distribution has infinite expectation.

Since the data set has only exceedances over a threshold, the PoT method is the appropriate way. When all losses are recorded, Dutta and Perry (2006) use a four-parameter distribution, called g-and-h, to model the data. If we focus on extreme events of financial assets returns, both upside and downside, standard methodologies also include the classical Student t and stable Paretian distributions, see Rachev *et al*. (2010). It cannot be considered a model for tails in the sense of Definition 2 and therefore suitable for combining with the methodology PoT.

Table 1 gives the MLE of parameters for exponential, gamma, Pareto and FTG distributions, as well as its standard deviations and the value of log-likelihood function, for the last data set. First of all, we observe that for Pareto distribution the parameter is in the range $0 < (-\alpha) < 1$, that is, a distribution with infinite expectation. This cannot be rejected with the goodness of fit test for

TABLE 1

MLE AND THE VALUE OF LOG-LIKELIHOOD FUNCTION BY EXPONENTIAL, GAMMA, PARETO AND FTG DISTRIBUTIONS OF LOSSES BY EXTERNAL FRAUD.

		λ			l
Exponential Distribution	MLE	0.010			−224.207
	s.e.	0.002			
		α	θ		l
Gamma Distribution	MLE	0.271	0.003		−181.938
	s.e.	0.048	0.001		
		α	σ		l
Pareto Distribution	MLE	−0.447	1.382		−174.440
	s.e.	0.102	0.732		
		α	σ	ρ	l
FTG Distribution	MLE	−0.197	0.651	4.3e-4	−172.369
	s.e.	0.152	0.586	6.2e-4	

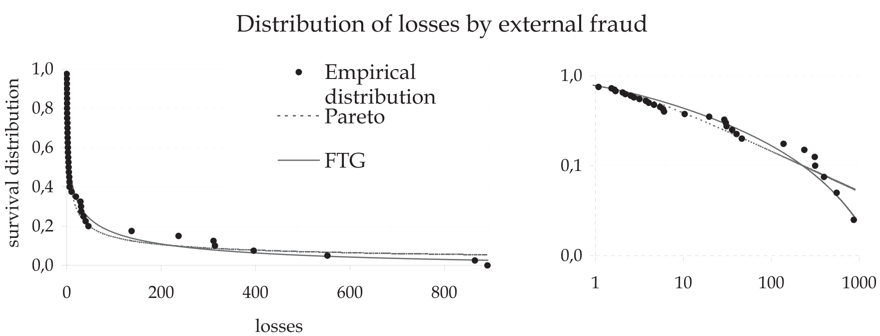

FIGURE 2: The FTG and the Pareto distributions fit the the empirical survival function in a similar way in the range of the observed sample. However, the estimated high quantiles differ greatly.

Pareto distribution given by Choulakian and Stephens (2001), since the parameter is outside the range of parameters provided by their tables. However, the Pareto distribution is nested in FTG distribution (Theorem 6) and likelihood ratio test is 4.142, with p-value 0.042. Hence, the FTG distribution fits the data set better than the Pareto distribution with a 95% confidence.

Figure 2 shows the empirical survival (or reliability) function and its fit given by Pareto and FTG distributions. The probability to exceed the maximum of the sample is estimated at 5.52% for the Pareto distribution and 2.65% for the FTG distribution; this difference does not seem essential. However, the estimation of high quantiles depends heavily on the model. The 0.999 quantile is 6.95×10^6 for the Pareto distribution and 3.93×10^3 for the FTG distribution. Moreover,

TABLE 2

PARAMETER ESTIMATES AND THE RISK CAPITAL FROM THE PARETO DISTRIBUTION AND THE FTG DISTRIBUTION FOR 10 BOOTSTRAP SAMPLES AND THE ORIGINAL DATA SET.

	Pareto Distribution			FTG Distribution			
Sample	α	σ	Risk Capital	α	$\log\theta$	$\log\rho$	Risk Capital
1	−0.310	0.367	2.47E+13	−0.038	−7.093	−9.250	10832.98
2	−0.373	1.122	3.50E+11	0.003	−6.771	−8.251	9292.72
3	−0.410	1.719	5.36E+10	−0.106	−7.341	−7.792	13407.05
4	−0.423	1.351	1.87E+10	−0.057	−6.543	−7.460	6934.26
5	−0.441	2.195	1.23E+10	−0.006	−6.520	−7.217	7603.78
6	−0.460	1.205	2.63E+09	−0.298	−8.039	−8.287	16860.11
7	−0.486	1.097	6.78E+08	−0.276	−7.313	−7.828	8921.12
8	−0.538	1.769	1.78E+08	−0.360	−7.613	−7.444	10997.30
9	−0.612	3.923	3.86E+07	−0.257	−6.723	−6.141	6503.94
10	−0.763	3.916	1.66E+06	−0.371	−6.113	−5.461	3276.98
Original	−0.448	1.382	5.78E+09	−0.197	−7.325	−7.754	10820.37

the difference is even greater to calculate the expected tail loss over this quantile that is the expected value of a loss if a tail event does occur; it is 12970.6 for the FTG distribution, since (16), and infinite for the Pareto distribution. Note that these quantities are measured in a monetary unit (such as dollars) to calculate risk capital, hence a factor of 10^3 is really important.

Risk capital has been calculated as 0.999 quantile of the aggregate losses, computed from (25), by simulating 10^5 times N loss events, where N is Poisson distributed with parameter $\lambda = 20$ and the loss events, L_i, are simulated from the fitted Pareto and FTG distributions. Using the FTG distribution the risk capital is 10820.4, using Pareto distribution is 5.78×10^9, see the last file in Table 2. In the monetary scale, see (26), these values correspond to \$99,806,952 for the FTG and $\$5.33 \times 10^{13}$ for the Pareto distribution. The USA Gross National Product in the first quarter of 2013 was $\$16.2 \times 10^{12}$, hence Pareto risk capital estimation is unrealistic. In the other hand, the FTG estimation is 12 times the maximum of the sample (\$8,242,664 in the monetary scale) that is a reasonable result.

In order to see the sample dependence of the risk capital estimate, we generated several *bootstrap* samples of the same size as the original data set. One observes immediately, with a small number of samples, the instability of the risk capital estimates obtained with the Pareto distribution. However, the estimates obtained with FTG distribution are much more stable.

Table 2 reports the parameter estimates and the risk capital from the Pareto distribution and the FTG distribution for 10 bootstrap samples and for the original data set. In all cases, risk capital has been calculated in the same way. Samples were selected from 100 bootstrap samples, ordered by the parameter α, choosing one out of 10, for more diversity. Note that only sample 2 corresponds to the truncated gamma distribution (5) and their behaviour is not different from the rest. The most prominent fact is that, in addition to the unrealistic risk capital estimation with Pareto distribution, its estimation is highly unstable, with a factor of 10^7. Hence, FTG distribution can be a valuable alternative to Pareto distribution on operational risk. Remark that the bootstrap method is only used in order to construct other scenarios that allow us to assess the stability of the estimated value for the risk capital and the compute this value is not subject to the weaknesses of the bootstrap method for heavy tails.

5. Wind-damage losses from hurricanes

Hurricanes are among the major causes of the most expensive insurance losses worldwide. The average annual hurricane damage in continental United States is about 10 billion (in 2005 dollars), see Pielke *et al.* (2008). The potential of a tropical cyclone to inflict damage is currently described by the maximum sustained surface wind speed (through the Saffir–Simpson intensity scale). However, economic losses not only depend on geophysical quantities. They also depend on the wealth, population density and coastal vulnerability of the impacted area. Official estimates of economic damage are usually obtained by doubling insured loss estimates.

Climate scientists are trying to determine whether hurricanes are becoming more frequent or destructive. Economic data are, in general, not the best way to ascertain changes in underlying geophysical variables, and such changes are best explored using geophysical data directly. There is a scientific debate on how best to describe a hurricane's destructive potential, improving maximum wind-speed measurement. Alternative methods based on measurements of energy are now considered more relevant to damage by wind, storm surge and waves; see Powell and Reinhold (2007).

Corral *et al.* (2010) study the influence of climate variability and global warming through the occurrence of tropical cyclones. Their approach is based on the application of an estimation of released energy to individual tropical cyclones. We will compare our model with its statistical analysis of power-law 热带气旋 distribution for 494 tropical cyclones in the north Atlantic between 1966 and 2009. We indicate (below) that FTG is much more realistic than the Pareto distribution, especially for the most destructive hurricanes, which are those that largely determine the global losses.

To measure the importance of the tropical cyclones, we use an estimation of released energy, the power dissipation index (PDI), defined by

$$\boxed{\mathrm{PDI} = \sum_t v_t^3 \Delta t},$$

where t denotes time and runs over the entire lifetime of the storm and v_t is the maximum sustained surface wind velocity at time t (PDI units are $\mathrm{m}^3/\mathrm{s}^2$). The PDI of the original data is between 5.38×10^8 and $2.54\ 10^{11}$. Deviations from

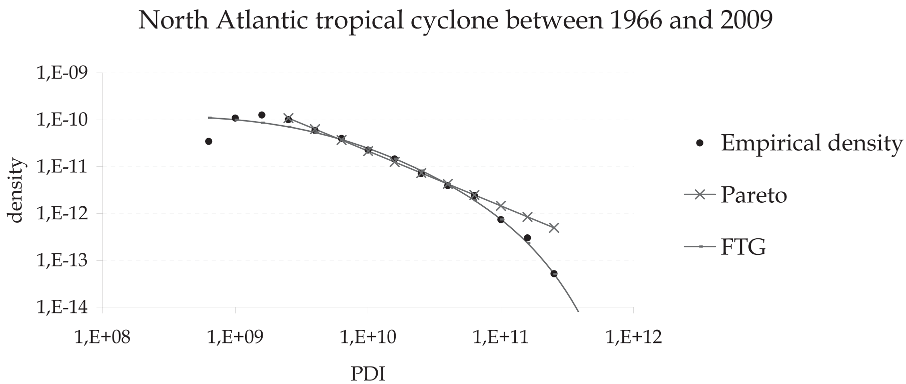

FIGURE 3: FTG distribution fits better than Pareto distribution the tropical cyclones data set, especially in the tail of the observations. The plot is scaled in common logarithm for both axes.

the power law at small PDI values were attributed to the deliberate incompleteness of the records for "no significant" storms. Their estimation only considers tropical cyclones with PDI bigger than 3×10^9, that is a sample of size 372 (75% of the original data).

Figure 3 shows the fit of the power-law distribution with an empirical approximation probability density function of the sample. Given the sample of tropical cyclones $\{x_i\}$ for $1 \le i \le n$ with $n = 494$, Corral *et al.* (2010) approximate the probability density function at points $p_r = 10^{8+(r-1)/5}$ for $1 < r < m$, where $m = 21$, for the histogram values

$$h_r = \frac{\#\{\, x_i \;:\; l_r < x_i \le l_{r+1}\}}{n\,(l_{r+1} - l_r)}$$

for the intervals given by $l_s = 0.5\, 10^{8+s/5}\, 11^{1/5}$ with $1 < s < m+1$. The goal of their method is to plot in common logarithm (base 10) scale for both axes, since the power-law probability density function in this situation corresponds to a straight line. The fit is done by least squares for a set of points $\{(u_r, v_r)\}$, where $u_r = \log_{10} p_r$ and $v_r = \log_{10} h_r$.

Our first contribution consists in fitting the FTG distribution by MLE for the whole sample. The FTG distribution shows a really best fit especially in the tail of the data, see Figure 3. The more rapid decay at large PDI is associated with the finite size of the ocean basin. That is, the storms with the largest PDI do not have enough room to last a longer time. The relevant thing is that FTG distribution fits the date even in this situation.

Theorem 6 shows that Pareto distribution is nested in FTG distribution, hence likelihood inference is now available. MLE of parameters and its standard deviations are shown in Table 3 for the FTG distribution and Pareto distribution (with two parameters). The values of log-likelihood function are −667.58 for the FTG case (truncated gamma distribution) and −680.06 for the Pareto case.

TABLE 3

MLE FOR FTG AND PARETO DISTRIBUTIONS FOR TROPICAL CYCLONE OCCURRED IN THE NORTH ATLANTIC BETWEEN 1966 AND 2009. THE DATA USED CORRESPONDS TO PDI OVER 3×10^9, WITH THE ORIGIN SHIFTED TO ZERO (UNITS ARE 10^{10} m^3/s^2). THIS CHANGE DOES NOT AFFECT THE LIKELIHOOD RATIO TEST, LRT, AND THE α PARAMETER.

	Pareto Distribution			FTG Distribution				
	α	σ	l	α	σ	ρ	l	LRT
MLE	−1.63	2.01	−680.06	0.28	0.09	0.02	−667.58	24.96
s.e.	0.22	0.41		0.15	0.11	0.02		

First of all, the goodness of fit test for the Pareto distribution given by Choulakian and Stephens (2001) rejects with p-value less than 0.001 for both statistics, $W^2 = 0.28$ and $A^2 = 2.4$, of the method, but it offers no alternative to the model. Finally, the likelihood ratio test can be used to find a confidence region around the FTG parameters, concluding that the difference between the FTG and the Pareto distribution is highly significant. The p-value is 5.8×10^{-7}.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I.A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover.

AKINSETE, A., FAMOYE, F. and LEE, C. (2008) The beta-Pareto distribution. *Statistics,* **42**, 547–563.

ARNOLD, B.C. (1983) *Pareto Distributions*. Fairland, Maryland: Interna- tional Cooperative Publishing House.

BARNDORFF-NIELSEN, O. (1978) *Information and Exponential Families in Statistical Theory,* Wiley Series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons.

BARNDORFF-NIELSEN, O. and COX, D. (1994) *Inference and asymptotics.* Monographs on Statistics and Applied Probability, vol. 52. London: Chapman & Hall.

BROWN, L. (1986) *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory,* Lecture Notes Monograph Series, vol. 9. Hayward, CA: Institute of Mathematical Statistics.

CASTILLO, J. DEL and PUIG, P. (1999) Invariant exponential models applied to reliability theory and survival analysis. *Journal of the American Statistical Association,* **94**, 522–528.

CHAPMAN, D.G. (1956) Estimating the parameters of a truncated gamma distribution. *Annals of Mathematical Statistics,* **27**, 498–506.

CHOULAKIAN, V. and STEPHENS, M.A. (2001) Goodness-of-fit for the generalized Pareto distribution. *Technometrics,* **43**, 478–484.

CLAUSET, A., SHALIZI, C.R. and NEWMAN, M.E.J. (2009) Power- law distributions in empirical data. *SIAM Review,* **51**, 661–703.

COLES, S. and SPARKS, R.S.J. (2006) Extreme value methods for modelling historical series of large volcanic magnitudes. In *Statistics in Volcanology* (eds. H.M. Mader. S.G. Coles, C.B. Connor, and L.J. Connor), pp. 47–56. Special publication of IAVCEI, vol. 1. London: Geological Society.

CORRAL, A., OSSO, A. and LLEBOT, J.E. (2010) Scaling of tropical-cyclone dissipation. *Nature Physics,* **6**, 693–696.

DAVIS, H.T. and MICHAEL L.F. (1979) The generalized Pareto law as a model for progressively censored survival data. *Biometrika* **66**, 299–306.

DEGEN, M., EMBRECHTS, P. and LAMBRIGGER, D. (2007) The quantitative modeling of operational risk: between g-and-h and EVT. *Astin Bulletin,* **37**, 265–291.

DEN BROEDER, G.G. (1955) On parameter estimation for truncated Pearson type III distributions. *Annals of Mathematical Statistics* **26**, 659–663.

DEVROYE, L. (1986) *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

DIXIT, U.J. and PHAL, K.D. (2005) Estimating scale parameter of a truncated gamma distribution. *Soochwon J. Math.* **31**, 515–523.

DUTTA, K. and PERRY, J. (2006) *A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital*. Boston, MA: Federal Reserve Bank of Boston. Working Paper 06-13.

EMBRECHTS, P. KLÜPPELBERG, C. and MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.

FISHER, R.A. (1922) On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, **222**, 309–368.

FURLAN, C. (2010) Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistical Modelling*, **10**, 113–132.

HARTER, H.L. (1967) Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples. *Technometrics,* **9**, 159–165.

HEGDE, L.M. and DAHIYA, R.C. (1989) Estimation of the parameters of a truncated gamma distribution. *Communications in Statistics: Theory and Methods,* **18**, 561–577.

LETAC, G. (1992) *Lectures on Natural Exponential Families and their Variance Functions.* Monografias de Matemática, vol. 50. IMPA, Rio de Janeiro.

MCNEIL, A.J., FREY, R. and EMBRECHTS P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.

MOSCADELLI, M. (2004) *The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee.* Economic Working Papers, 517, Bank of Italy, Rome, Economic Research Department.

PHILIPPE, A. (1997) Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing,* **7**, 173–181.

PIELKE, JR., R.A., GRATZ, J., LANDSEA, C.W., COLLINS, D., SAUNDERS, M.A. and MUSULIN, R. (2008) Normalized hurricane damages in the United States: 1900–2005. *Natural Hazards Review,* **9**, 29–42.

POWELL, M.D. and REINHOLD, T.A. (2007) Tropical cyclone destructive potential by integrated kinetic energy. *Bulletin of the American Meteorological Society,* **87**, 513–526.

R Development Core Team (2010) R: *A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

RACHEV, Z., RACHEVA-IOTOVA, B. and STOYANOV, S. (2010) Capturing fat tails. *Risk Magazine*, May 2010, 72–77.

SORNETTE, D. (2006) *Critical Phenomena in Natural Sciences.* Berlin, Heidelberg, New York: Springer.

STACY, E.W. (1962) A generalization of the gamma distribution.*Annals of Mathematical Statistics,* **33**, 1187–1192.

JOAN DEL CASTILLO (Corresponding author)
Department of Mathematics
Universitat Autònoma de Barcelona
08193 Cerdanyola del Vallès, Spain
E-Mail: castillo@mat.uab.cat

JALILA DAOUDI
Universitat Pompeu Fabra

Barcelona, Spain
jalila.daoudi@upf.edu

ISABEL SERRA
Centre de Recerca Matemàtica
Barcelona, Spain
E-Mail: iserra@crm.cat

APPENDIX A

The expected Fisher information matrix

$$I_E(\alpha, \sigma, \rho) = E(I_O(\alpha, \sigma, \rho)) \tag{A1}$$

can be computed in closed form using the expectations of the relations

$$r_\sigma = \sigma^{-1}(1-r), \quad r_{\sigma\sigma} = -\sigma^{-2}(1-r), \quad s_\sigma = -\sigma^{-1}(1-r^{-1}) \text{ and } s_{\sigma\sigma} = \sigma^{-2}(1-r^{-2}), \tag{A2}$$

since for all $k \in \mathbb{R}$,

$$E(r^k) = \rho^{-k}\Gamma(\alpha + k, \rho)/\Gamma(\alpha, \rho).$$

Further expression by I_E can be obtained using others relations, for instance,

$$\rho E(r) + (\alpha - 1) E(r^{-1}) = \rho + \alpha,$$

which is obtained from $\Gamma(\alpha + 1, \rho)/\Gamma(\alpha, \rho) = \alpha - \rho d_\rho$ or equivalently, $\Gamma(\alpha, \rho) = (\alpha - 1)\Gamma(\alpha - 1, \rho) + \rho^{\alpha-1}\exp(-\rho)$. Both are obtained from integration by parts in (2).

Remark that some terms of the I_E have an explicit expression, since are exactly the terms in observed information matrix. We denote the integral (2) by

$$\Gamma = \Gamma(\alpha, \rho) = \rho^\alpha E_{1-\alpha}(\rho), \, E_n(\rho) = \int_1^\infty t^{-n} e^{-\rho\, t} dt,$$

and the first derivatives can be expressed by

$$d_\rho = \Gamma_\rho / \Gamma \quad \text{and} \quad d_\alpha = \Gamma_\alpha / \Gamma$$

and the seconds by

$$d_{\rho\rho} = \Gamma_{\rho\rho} / \Gamma - (\Gamma_\rho / \Gamma)^2, \quad d_{\rho\alpha} = \Gamma_{\rho\alpha} / \Gamma - (\Gamma_\alpha / \Gamma)(\Gamma_\rho / \Gamma), \quad \Gamma_{\alpha\alpha} = \Gamma_{\alpha\alpha} / \Gamma - (\Gamma_\alpha / \Gamma)^2.$$

Finally, by dominated convergence and/or derivation under the integral sign theorems, the first derivatives of Incomplete Gamma function are

$$\Gamma_\rho = \Gamma_\rho\,(\alpha, \rho) = -\rho^{\alpha-1} e^{-\rho} \quad \text{and} \quad \Gamma_\alpha = \Gamma_\alpha\,(\alpha, \rho) = \int_\rho^\infty \log\,(t)\, t^{\alpha-1} e^{-t} dt,$$

and the seconds

$$\begin{aligned}
\Gamma_{\rho\rho} &= \Gamma_{\rho\rho}\,(\alpha, \rho) = \left(\rho^{\alpha-1} - (\alpha - 1)\,\rho^{\alpha-2}\right) e^{-\rho}, \\
\Gamma_{\rho\alpha} &= \Gamma_{\rho\alpha}\,(\alpha, \rho) = -\log\,(\rho)\,\rho^{\alpha-1} e^{-\rho}, \\
\Gamma_{\alpha\alpha} &= \Gamma_{\alpha\alpha}\,(\alpha, \rho) = \int_\rho^\infty \log^2\,(t)\, t^{\alpha-1} e^{-t} dt.
\end{aligned}$$