Jurnal EurekaMatika

Journal homepage: https://ejournal-science.upi.edu/jem/index
doi: https://doi.org/10.17509/jem.v14i2.1495

The Nested Beta-Pareto Distribution: Properties and Applications

Alfred Ayo Ayenigba[1], Ifeyinwa Vivian Omakem[2], Alabi Nurudeen Olawale[3] and Olajide Oyewole Agboola[1]*

[1]Ajayi Crowther University, Faculty of Natural Sciences, Department of Mathematical Sciences, Oyo, Nigeria
[2]University of Ilorin, Faculty of Physical Sciences, Department of Statistics, Ilorin, Nigeria
[3]Federal Polytechnic Ilaro, Department of Mathematics and Statistics, Ilaro, Nigeria
*Correspondence: aa.ayenigba@acu.edu.ng

ABSTRACT

Heavy-tailed data with a lower threshold occur widely in hydrology, insurance and reliability, and are usually modelled with the Pareto law or with beta-generated extensions of it. These extensions are flexible, but their survival functions involve incomplete beta or gamma functions, which complicates reliability analysis. This paper proposes a three-parameter subfamily of the beta-Pareto distribution, obtained by fixing the first generator parameter at two, whose survival function is exactly elementary: a linear combination of two Pareto survival functions. The distribution is the product of two independent Pareto variables, its logarithm is hypo exponential, and its tail is an exact member of Hall's second-order Pareto class. These facts give closed forms for the reliability functions, moments, entropies, order-statistic moments and stress-strength reliability, and they fix the hazard rate as upside-down bathtub shaped, with an explicit mode, for every parameter value. Maximum likelihood estimation reduces to an explicit two-dimensional profile search, and a simulation study confirms accurate estimation of the tail index and threshold. Across two flood- and insurance-loss applications, the model gives the best fit among eight competitors for River Nidd exceedances and is statistically indistinguishable from its own Pareto sub-model for Danish fire losses. A built-in lower bound on the log-scale dispersion of the data gives a necessary, though not sufficient, condition for the model to improve on the Pareto law, a partial diagnostic rather than a guarantee.

© 2026 Kantor Jurnal dan Publikasi UPI

ARTICLE INFORMATION

Article History:

Received October 1, 2026
Revised November 1, 2026
Approved November 9, 2026
Available online November 18, 2026
Published November 18, 2026

Keywords:
Beta-Pareto Distribution,
Flood Exceedances,
Hazard Rate,
Maximum Likelihood.

1. INTRODUCTION

Power-law tails are a defining feature of flood peaks, insurance losses, city sizes and other quantities observed only above a threshold (Clauset et al., 2009). The Pareto distribution, with survival function $\bar{F}(x) = \left(\frac{x}{\theta}\right)^{-k}$ for $x \geq \theta$, is the classical model for such data, but its single shape parameter controls both body and tail, which empirical studies repeatedly find too restrictive (Cirillo, 2013). Much recent work therefore embeds the Pareto law in larger families through distribution generators, continuing a broader tradition of deriving exact, closed-form distributional properties wherever a model admits them, as in recent work on the uniform-exponential distribution (Ayenigba & Ogunkoya, 2025).

The beta-generated class of Eugene et al. (2002) transforms a baseline cumulative distribution function (CDF) F into

$$G(x) = \frac{1}{B(a,b)} \int_0^{F(x)} t^{a-1} (1-t)^{b-1} \, \mathrm{d}t = I_{F(x)}(a,b), \qquad a,b > 0, \tag{1}$$

where $I_z(a,b)$ is the regularized incomplete beta function. Jones (2004) linked this class to order statistics. Later generators include the Kumaraswamy-G family (Cordeiro & de Castro, 2011), the T-X framework (Alzaatreh, Lee, et al., 2013) and the alpha-power transformation (Mahdavi & Kundu, 2017); see Lee et al. (2013) and Tahir and Nadarajah (2015) for reviews. With a Pareto baseline, (1) gives the beta-Pareto (BP) distribution of Akinsete et al. (2008), with density

$$g_{\mathrm{BP}}(x) = \frac{k}{\theta B(a,b)} \left\{1 - \left(\frac{x}{\theta}\right)^{-k}\right\}^{a-1} \left(\frac{x}{\theta}\right)^{-kb-1}, \qquad x \geq \theta. \tag{2}$$

Other Pareto extensions include the exponentiated Pareto (Gupta et al., 1998), beta generalised Pareto (Mahmoudi, 2011), gamma-Pareto (Alzaatreh et al., 2012), Weibull-Pareto (Alzaatreh, Famoye, et al., 2013; Tahir et al., 2016), Kumaraswamy-Pareto (Bourguignon et al., 2013), Kumaraswamy-transmuted Pareto (Chhetri et al., 2017), alpha-power Pareto (Ihtisham et al., 2019), Marshall-Olkin generalised Pareto (Haj Ahmad & Almetwally, 2020), odds exponential-Pareto IV (Baharith et al., 2020), several further generalised Pareto models (Ghitany et al., 2018; Salahuddin et al., 2021; Karakaya et al., 2024), the extended Pareto (Aga, 2026) and a gamma-generated Pareto-type model (Hmingthansanga et al., 2026); see Hamed et al. (2018) for a survey.

These models share one drawback. Their survival functions are incomplete beta or gamma functions, or compositions of them. The hazard rate, cumulative hazard and mean residual life then require special-function evaluation, shape results for the hazard are hard to establish, and censored likelihoods become cumbersome. For the BP in particular, $\bar{G}(x) = I_z(b,a)$ with $z = \left(\frac{x}{\theta}\right)^{-k}$, which is elementary only in special cases.

Within the beta-Pareto class of Akinsete et al. (2008), no member combines an elementary survival function with an interpretable split between body and tail except the one obtained by fixing $a = 2$ and writing $b = \frac{\lambda}{k}$, giving a three-parameter subfamily we denote NBP (Nested Beta-Pareto), whose survival function is a linear (non-mixture) combination of two Pareto survival functions; isolating this subfamily, and developing its reliability, estimation and applied properties, is the contribution of this paper. Four claims organise the rest of the paper: (i) the model is the product of two independent Pareto variables, its logarithm hypoexponential, and its tail an exact member of the second-order Pareto class of Hall (1982); (ii) these facts make the standard reliability measures, moments and entropies elementary; (iii) the hazard rate is upside-down bathtub shaped for every parameter value, with an explicit mode, and the likelihood equation for the tail index is explicitly solvable; (iv)

the same facts impose a testable restriction, a lower bound on the log-scale dispersion of the data, that gives a necessary condition, checkable in advance of fitting, for the model to be competitive.

Because Theorem 2 shows that the NBP arises as the scaled product of two independent Pareto variables, with logarithm hypo exponential, it also sits within an existing literature on products of Pareto-type variables (Nadarajah, 2008; Obeid & Kadry, 2022), which derives the distribution of such a product for its own sake, typically treating the two Pareto rates as unrelated nuisance parameters. What distinguishes the present construction is not the product representation itself, which is a consequence of the model rather than its starting point, but the specific coupling of the two rates through a single additional parameter: this is exactly the restriction that turns a generic product of two Pareto variables into the three-parameter subfamily studied here, with one elementary survival function, one interpretable body-tail split, and a closed-form inferential apparatus throughout.

The trade-off is scope: fixing $a = 2$ buys tractability at the cost of a fixed hazard shape and a restricted range of log-scale dispersion, so the model is not a universal replacement for the wider beta-Pareto class; Section 3.5 states this limitation precisely. Section 2 constructs the model, relates it to existing Pareto-type distributions, and develops its reliability measures, moments, quantile function, entropies, order statistics, stress-strength reliability and estimation method. Section 3 reports a simulation study and two real-data applications, together with a formal limitations statement and section 4 concludes.

2. METHODOLOGY

2.1 Model Construction

Construction. Write $y = \frac{x}{\theta} \geq 1$ and $z = y^{-k}$, so that $0 < z \leq 1$. Since $1 - F(x) = z$ for the Pareto baseline, the beta-Pareto (BP) survival function is

$$\bar{G}(x) = 1 - I_{1-z}(a, b) = I_z(b, a). \tag{3}$$

Lemma 1. For $b > 0$ and $0 \leq z \leq 1$, $I_z(b, 2) = z^b\{1 + b(1 - z)\}$.

Proof. Since $B(b, 2) = \frac{1}{b(b+1)}$, and $\int_0^z t^{b-1}\,(1 - t)\,\mathrm{d}t = \frac{z^b}{b} - \frac{z^{b+1}}{b+1}$, we obtain $I_z(b, 2) = (b + 1)z^b - bz^{b+1} = z^b\{1 + b(1 - z)\}$. ▫

More generally, $I_z(b, a) = z^b \sum_{j=0}^{a-1} \frac{(b)_j (1-z)^j}{j!}$ for integer a, a standard combinatorial identity for the incomplete beta function. The case $a = 1$ returns the Pareto law, so $a = 2$ is the simplest case that adds a parameter while keeping the survival function elementary. Setting $a = 2$ in (2) and writing $b = \frac{\lambda}{k}$ couples the second generator parameter to the Pareto shape through the tail index $\lambda = kb$. Define

$$\tau_k(y) = \frac{1 - y^{-k}}{k}, \qquad y \geq 1,\ k > 0, \tag{4}$$

which increases in y, decreases in k, tends to $\log y$ as $k \to 0$, and satisfies $k\,\tau_k(y) \to 1$ as $k \to \infty$.

Definition 1. A random variable X has the NBP distribution, written $X \sim \mathrm{NBP}(\lambda, k, \theta)$, if its probability density function (PDF) is

$$g(x) = \frac{\lambda(\lambda + k)}{\theta}\, y^{-(\lambda+1)}\, \tau_k(y), \qquad x > \theta, \quad y = \frac{x}{\theta}, \tag{5}$$

where $\lambda > 0$ is the tail index, $k > 0$ the body parameter and $\theta > 0$ the threshold.

Remark 1 (identifiability). The density (5) is driven by two hypo exponential rates, λ and $\lambda + k$ (see Stochastic representation, below), which are not two free, exchangeable rates: λ

is fixed by construction as the smaller of the two, and $k > 0$ is the gap between them. No ordering constraint on a pair of raw rates is therefore needed, and $k = 0$ is not a boundary singularity: it is the ordinary interior limit $\lambda + k \to \lambda$, at which (5) reduces smoothly to the gamma-Pareto sub-model described below. This is a deliberate feature of the (λ, k) parameterisation, and it is why no case distinction on the relative size of the two rates appears anywhere below. This is a structural property of the parameterisation itself, holding for every sample size; it does not imply that the body parameter is well determined by any particular finite sample, a separate, statistical question addressed quantitatively in Section 3.1 and Section 3.5.

Theorem 1. The survival function and CDF of $X \sim \text{NBP}(\lambda, k, \theta)$ are

$$\bar{G}(x) = y^{-\lambda}\{1 + \lambda\,\tau_k(y)\} = \frac{\lambda + k}{k}\,y^{-\lambda} - \frac{\lambda}{k}\,y^{-(\lambda+k)}, \qquad G(x) = 1 - \bar{G}(x). \tag{6}$$

Proof. By Lemma 1 with $b = \frac{\lambda}{k}$, equation (3) gives $\bar{G}(x) = z^{\frac{\lambda}{k}}\{1 + \frac{\lambda}{k}(1 - z)\}$. Since $z^{\frac{\lambda}{k}} = y^{-\lambda}$ and $\frac{1-z}{k} = \tau_k(y)$, this is the first form in (6). Expanding $\tau_k(y) = \frac{1-y^{-k}}{k}$ inside the braces gives the second form. Differentiating the second form with respect to x returns the density (5). ▫

The survival function is thus a linear combination (not a probability mixture, since the second weight is negative) of Pareto survival functions with indices λ and $\lambda + k$, with coefficients $\frac{\lambda+k}{k}$ and $-\frac{\lambda}{k}$.

Theorem 2. Let $P_1 \sim \text{Pareto}(\lambda, 1)$ and $P_2 \sim \text{Pareto}(\lambda + k, 1)$ be independent, and let E_1, E_2 be independent standard exponential variables. Then

$$X \overset{d}{=} \theta P_1 P_2, \qquad S = \log\frac{X}{\theta} \overset{d}{=} \frac{E_1}{\lambda} + \frac{E_2}{\lambda + k}. \tag{7}$$

Proof. If $P \sim \text{Pareto}(\nu, 1)$ then $\log P$ is standard exponential with rate ν, so $\log P_1$ and $\log P_2$ are independent exponentials with rates λ and $\lambda + k$. Their sum $S = \log P_1 + \log P_2$ is hypoexponential, with density $f_S(s) = \left\{\frac{\lambda(\lambda+k)}{k}\right\}\left(e^{-\lambda s} - e^{-(\lambda+k)s}\right)$ for $s > 0$, a standard result for the hypoexponential distribution obtained by direct convolution of the two exponential densities (see also Ayenigba et al., 2025, for convolution methods for sums of independent random variables). Writing $X = \theta P_1 P_2$, so that $\log\left(\frac{X}{\theta}\right) = S$, the transformation $x = \theta e^s$ gives the density $f_X(x) = \frac{f_S(\log y)}{x}$. This expression coincides with (5), which proves the stochastic representation. ▫

Theorem 2 yields the exact sampler $X = \theta U_1^{-\frac{1}{\lambda}} U_2^{-\frac{1}{\lambda+k}}$, where U_1, U_2 are independent uniform variables on $(0,1)$. It also shows that the rates λ and $\lambda + k$ enter symmetrically, so the constraint $k > 0$ ensures identifiability.

Sub-models, ordering and shape. As $k \to \infty$, $\bar{G}(x) \to y^{-\lambda}$, the Pareto law, with $g(x) \to \lambda\theta^{\lambda}x^{-\lambda-1}$. As $k \to 0$ (the equal-rate limit of Remark 1), $\tau_k(y) \to \log y$ by l'Hôpital's rule, so $\bar{G}(x) \to y^{-\lambda}(1 + \lambda\log y)$ and $g(x) \to \lambda^2\theta^{\lambda}x^{-\lambda-1}\log y$, the gamma-Pareto law of Alzaatreh et al. (2012) with shape two. Both limits are attained smoothly, with no discontinuity in the density or its derivatives. Because $\tau_k(y) \le \log y$ and τ_k decreases in k, the survival function decreases in both λ and k. Hence X is stochastically decreasing in each parameter, and $\text{Pareto}(\lambda, \theta) \le_{\text{st}} \text{NBP}(\lambda, k, \theta)$.

The density vanishes at θ and at infinity. Solving $\frac{\mathrm{d}\{y^{-\lambda-1} - y^{-\lambda-k-1}\}}{\mathrm{d}y} = 0$ gives the unique mode

$$x_{\text{mode}} = \theta\left(\frac{\lambda + k + 1}{\lambda + 1}\right)^{\frac{1}{k}}. \tag{8}$$

Near the threshold, $g(x) \sim \lambda(\lambda + k)\theta^{-2}(x - \theta)$. Figure 1(a) shows the range of shapes.

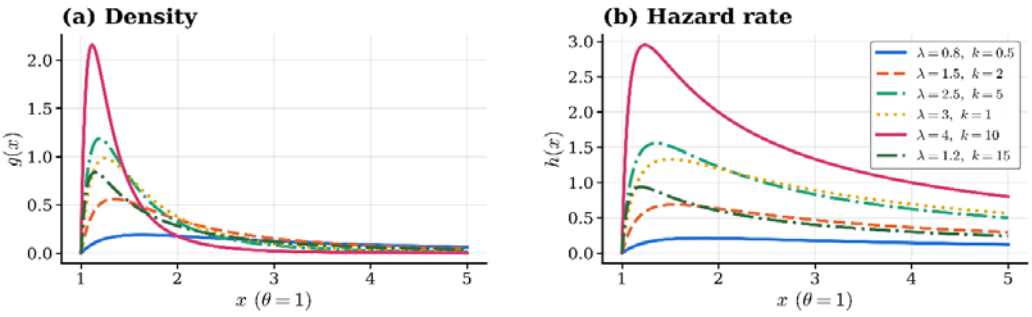

Figure 1. *(a) NBP densities and (b) hazard rates for selected* (λ, k) *with* $\theta = 1$.

Relationship with existing Pareto-type models. The NBP is the $a = 2$ member of the beta-Pareto (BP) class of Akinsete et al. (2008), under the coupling $b = \frac{\lambda}{k}$ introduced above; it is a genuine subfamily, not a reparameterisation of an already-elementary model, since it is not equivalent under any parameter mapping to the five other Pareto-type extensions in common use. It nests the *Pareto* law only as $k \to \infty$; for every finite k the two differ near the threshold, where the NBP survival function is concave and the Pareto survival function is convex. *Exponentiated Pareto* (Gupta et al., 1998) and *Kumaraswamy-Pareto* (Bourguignon et al., 2013) both raise $(1 - y^{-k})$ to one or two outer powers, so a single exponent k controls the entire body, with a survival function that admits a finite expansion in powers of y^{-k} only for integer outer parameters, and only in multiples of k; (6), by contrast, has exactly two additively independent exponents, λ and $\lambda + k$, which no choice of their parameters reproduces except at the shared Pareto boundary. Within the *BP* family itself, the identity above (for $I_z(b, m)$) needs m terms: $a = 1$ returns the Pareto law and $a \geq 3$ needs three or more terms, so NBP ($a = 2$) is the unique member with a two-term elementary survival function. The *gamma-Pareto* law (Alzaatreh et al., 2012) is not a competing model but the exact $k \to 0$ boundary sub-model of NBP (Remark 1, above), reached smoothly rather than approximated. *Weibull-Pareto* (Alzaatreh, Famoye, et al., 2013) and the *generalised Pareto* (Hosking & Wallis, 1987) instead come from generators outside the beta-Pareto class (a Weibull-integral transform and a threshold-exceedance limit law) and do not embed it under any parameter choice. Table 8 (Section 3.4) extends this comparison analytically and empirically for BP, GaP and the Pareto sub-model, the three families NBP nests most directly.

2.2 Reliability Measures and Tail Behaviour

Reliability functions. From (5) and (6), the hazard, cumulative hazard, reversed hazard and odds functions are

$$h(x) = \frac{\lambda(\lambda + k)\,\tau_k(y)}{\theta\, y\,\{1 + \lambda\tau_k(y)\}}, \qquad H(x) = \lambda \log y - \log\{1 + \lambda\tau_k(y)\}, \tag{9}$$

$$r(x) = \frac{\lambda(\lambda + k)\, y^{-(\lambda+1)}\tau_k(y)}{\theta\,[1 - y^{-\lambda}\{1 + \lambda\tau_k(y)\}]}, \qquad O(x) = \frac{y^{\lambda}}{1 + \lambda\tau_k(y)} - 1. \tag{10}$$

For $\lambda > 1$ and $t > \theta$, with $y = \frac{t}{\theta}$, the mean residual life is

$$m(t) = \frac{\theta}{\bar{G}(t)}\left\{\frac{(\lambda + k)\, y^{1-\lambda}}{k(\lambda - 1)} - \frac{\lambda\, y^{1-\lambda-k}}{k(\lambda + k - 1)}\right\}, \tag{11}$$

and $m(t) \sim \frac{t}{\lambda-1}$ as $t \to \infty$. None of these quantities involves a special function.

Theorem 3. For all $\lambda, k > 0$ the hazard rate is upside-down bathtub shaped, with $h(\theta) = 0$ and $x\,h(x) \to \lambda$ as $x \to \infty$. Its unique maximum is at $x_0 = \theta\,(u_0)^{-\frac{1}{k}}$, where

$$u_0 = \frac{(2\lambda + k + k^2) - \sqrt{(2\lambda + k + k^2)^2 - 4\lambda(\lambda + k)}}{2\lambda}. \tag{12}$$

Proof. Put $x = \theta e^s$ and $u = e^{-ks}$. Because $f_X(x) = \frac{f_S(s)}{x}$ and $\bar{G}(x) = \bar{F}_S(s)$, we have $h(x) = \frac{h_S(s)}{x}$, where $h_S(s) = \frac{\lambda(\lambda+k)(1-u)}{\lambda+k-\lambda u}$ is the hazard of S in (7). Hence $\frac{\mathrm{d}\log h(x)}{\mathrm{d}x} = x^{-1}\{D(s) - 1\}$, with $D(s) = \frac{\mathrm{d}\log h_S(s)}{\mathrm{d}s} = \frac{k^2 u}{(1-u)(\lambda+k-\lambda u)}$. The numerator of D increases in u and its positive denominator decreases in u, so D decreases strictly in s. As $s \to 0$, $u \to 1$ and $D(s) \to +\infty$; as $s \to \infty$, $u \to 0$ and $D(s) \to 0$. The equation $D(s) = 1$ therefore has exactly one root, and h rises before it and falls after it. The equation $D = 1$ is equivalent to $\lambda u^2 - (2\lambda + k + k^2)u + (\lambda + k) = 0$. Its left side is positive at $u = 0$ and equals $-k^2$ at $u = 1$, so its root in $(0,1)$ is (12). Finally, $h_S(0) = 0$ and $h_S(s) \to \lambda$ as $s \to \infty$, which give the stated limits. ▫

The BP hazard can be increasing, decreasing or unimodal, depending on its parameters. The NBP has a single interpretable shape: risk rises after the threshold, peaks at x_0 and then decays at the Pareto rate $\frac{\lambda}{x}$ (Glaser, 1980).

Tail behaviour. The survival function (6) can be written as

$$\bar{G}(x) = C\,x^{-\lambda}(1 + D\,x^{-k}), \qquad C = \frac{\theta^\lambda(\lambda + k)}{k}, \quad D = -\frac{\lambda\,\theta^k}{\lambda + k}. \tag{13}$$

This is exactly the second-order Pareto model of Hall (1982), with no remainder term: in Hall's own notation $\bar{F}(x) = Cx^{-\frac{1}{\gamma}}\{1 + Dx^\rho + o(x^\rho)\}$, (13) gives $\gamma = \frac{1}{\lambda}$, $\rho = -k$ and the coefficients C, D above exactly, with the $o(x^\rho)$ term identically zero. The tail is regularly varying with index $-\lambda$, X lies in the Fréchet domain of attraction with extreme-value index $\frac{1}{\lambda}$, and the second-order exponent $\rho = -k$ depends on the parameterisation only through k, not through λ (Hall, 1982; Hall & Welsh, 1985). Because Hall's class is the standard setting for studying the bias of the Hill estimator (Hall & Welsh, 1985), the NBP supplies a proper distribution in which this bias can be computed exactly. Figure 2(a) shows that the NBP tail runs parallel to the Pareto tail with the same λ.

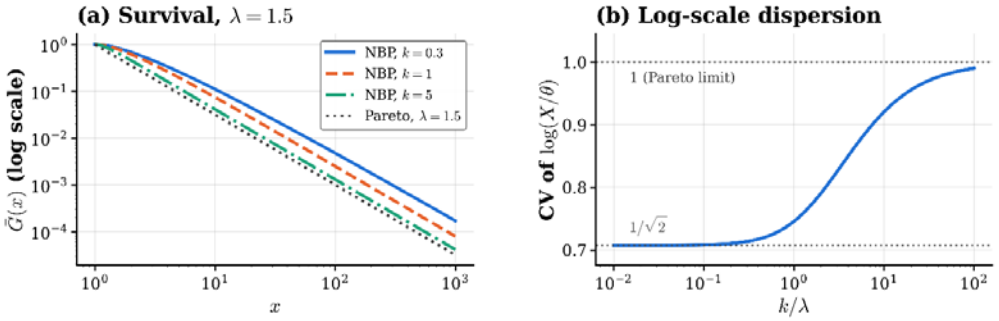

Figure 2. *(a) NBP survival functions for $\lambda = 1.5$ against the Pareto law on log-log axes. (b) Coefficient of variation of $log\left(\frac{X}{\theta}\right)$ as a function of $\frac{k}{\lambda}$.*

2.3 Moments and Integral Transforms

Theorem 4. For real $r < \lambda$,

$$\mu_r{}' = E(X^r) = \theta^r \, \frac{\lambda(\lambda + k)}{(\lambda - r)(\lambda + k - r)}, \tag{14}$$

and $E(X^r) = \infty$ for $r \geq \lambda$.

Proof. By Theorem 2, $E(X^r) = \theta^r E(P_1^r) E(P_2^r)$. For $P \sim$ Pareto$(\nu, 1)$, $E(P^r) = \frac{\nu}{\nu - r}$ when $r < \nu$, and $E(P^r) = \infty$ otherwise. Applying this with $\nu = \lambda$ and $\nu = \lambda + k$ gives (14). ▫

The mean exists for $\lambda > 1$ and the variance for $\lambda > 2$:

$$\begin{aligned} \mu &= \frac{\theta\lambda(\lambda + k)}{(\lambda - 1)(\lambda + k - 1)}, \\ \sigma^2 &= \theta^2 \left\{ \frac{\lambda(\lambda + k)}{(\lambda - 2)(\lambda + k - 2)} - \frac{\lambda^2(\lambda + k)^2}{(\lambda - 1)^2(\lambda + k - 1)^2} \right\}. \end{aligned} \tag{15}$$

The skewness ($\lambda > 3$) and kurtosis ($\lambda > 4$) follow from (14) and do not depend on θ; Table 1 shows that both decrease in λ, uniformly over the tabulated range of k, reported in Table 1 and Table 5 as raw (Pearson) kurtosis -- the normal value is 3, and the corresponding excess kurtosis is this value minus 3.

Table 1. First four raw moments, variance, skewness, kurtosis and median of the NBP distribution ($\theta = 1$).

λ	k	$\mu_1{}'$	$\mu_2{}'$	$\mu_3{}'$	$\mu_4{}'$	σ^2	γ_1	γ_2	**Median**
5	1	1.5000	2.5000	5.0000	15.0000	0.2500	4.0000	57.0000	1.3595
5	5	1.3889	2.0833	3.5714	8.3333	0.1543	4.1113	60.5271	1.2784
5	20	1.3021	1.8116	2.8409	5.9524	0.1162	4.5339	71.2134	1.1999
6	2	1.3714	2.0000	3.2000	6.0000	0.1192	3.1653	28.3778	1.2759
8	2	1.2698	1.6667	2.2857	3.3333	0.0542	2.5174	16.3191	1.2072
8	10	1.2101	1.5000	1.9200	2.5714	0.0357	2.7404	18.9605	1.1579
10	5	1.1905	1.4423	1.7857	2.2727	0.0251	2.2620	13.0333	1.1487

By (7), the cumulants of $S = \log\left(\frac{X}{\theta}\right)$ are:

$$\kappa_r(S) = (r - 1)! \{\lambda^{-r} + (\lambda + k)^{-r}\}, \qquad r \geq 1. \tag{16}$$

With $\eta = \frac{\lambda}{\lambda + k}$, the coefficient of variation of S is:

$$\mathrm{CV}(S) = \frac{\sqrt{1 + \eta^2}}{1 + \eta}, \qquad \frac{1}{\sqrt{2}} \leq \mathrm{CV}(S) < 1. \tag{17}$$

This bound (Figure 2b) characterises the data the model is designed for: power-law data that, on the log scale above the threshold, are at least as dispersed as a gamma variable with shape two.

For $t > \theta$, $y = \frac{t}{\theta}$ and $r \notin \{\lambda, \lambda + k\}$, the rth incomplete moment is

$$m_r(t) = \int_\theta^t x^r \, g(x) \, \mathrm{d}x = \theta^r \, \frac{\lambda(\lambda + k)}{k} \left\{ \frac{1 - y^{r - \lambda}}{\lambda - r} - \frac{1 - y^{r - \lambda - k}}{\lambda + k - r} \right\}. \tag{18}$$

The mean deviations about the mean and the median, and the Lorenz and Bonferroni curves, are all elementary functions of m_1 and are omitted for brevity.

Since the tail is regularly varying, $E(e^{tX}) = \infty$ for every $t > 0$, and the moment generating function does not exist. Let $E_p(z) = \int_1^\infty e^{-zt} \, t^{-p} \, \mathrm{d}t$ be the generalised exponential integral. Integrating (5) termwise gives the characteristic function:

$$\varphi(t) = E\left(e^{\mathrm{i}tX}\right) = \frac{\lambda(\lambda + k)}{k} \{E_{\lambda + 1}(-\mathrm{i}t\theta) - E_{\lambda + k + 1}(-\mathrm{i}t\theta)\}, \qquad t \in \mathbb{R}, \tag{19}$$

and replacing $-\mathrm{i}t$ by $s \geq 0$ gives the Laplace transform. Such transform-based characterisations have a discrete counterpart in probability generating functions, used analogously for distribution generation, branching processes and queueing models (Ayenigba et al., 2026).

2.4 Quantile Function

The quantile $Q(p) = \theta y_p$ is the unique root of

$$y^{-\lambda}\{1 + \lambda\tau_k(y)\} = 1 - p, \qquad 0 < p < 1, \tag{20}$$

and it lies in the bracket $(1-p)^{-\frac{1}{\lambda}} \leq y_p \leq \left\{\frac{k(1-p)}{\lambda+k}\right\}^{-\frac{1}{\lambda}}$, so bisection always converges. When $k = \lambda$, (20) is a quadratic in $y^{-\lambda}$ and

$$Q(p) = \theta\left(1 - \sqrt{p}\right)^{-\frac{1}{\lambda}}. \tag{21}$$

The median, Bowley skewness and Moors kurtosis follow from Q. Random numbers are generated exactly by the product representation of Theorem 2.

2.5 Entropies

Theorem 5. For $\xi > 0$, $\xi \neq 1$ and $\xi(\lambda+1) > 1$, the Rényi entropy is

$$I_R(\xi) = \frac{1}{1-\xi}\log\left[\left\{\frac{\lambda(\lambda+k)}{k\theta}\right\}^{\xi}\frac{\theta}{k}\,B\left(\frac{\xi(\lambda+1)-1}{k},\,\xi+1\right)\right]. \tag{22}$$

Proof. Rényi entropy requires $\int_\theta^\infty g\,(x)^\xi\,\mathrm{d}x$, and by (5) this is proportional to $\int_\theta^\infty y^{-\xi(\lambda+1)}\,(1-y^{-k})^\xi\,\mathrm{d}x$. The substitution $w = y^{-k}$, so that $\mathrm{d}x = -\left(\frac{\theta}{k}\right)w^{-\frac{1}{k}-1}\,\mathrm{d}w$, transforms this integral into $\left(\frac{\theta}{k}\right)\int_0^1 w^{\frac{\xi(\lambda+1)-1}{k}-1}\,(1-w)^\xi\,\mathrm{d}w$. This is a complete beta integral, equal to $\left(\frac{\theta}{k}\right)B\left(\frac{\xi(\lambda+1)-1}{k},\,\xi+1\right)$, which is finite if and only if $\xi(\lambda+1) > 1$. Collecting the constant $\left\{\frac{\lambda(\lambda+k)}{k\theta}\right\}^\xi$ from g^ξ and applying $(1-\xi)^{-1}\log(\cdot)$ gives (22). ▫

Theorem 6. The Shannon (1948) entropy is

$$H_S = \log\frac{k\theta}{\lambda(\lambda+k)} + (\lambda+1)\left(\frac{1}{\lambda}+\frac{1}{\lambda+k}\right) + \frac{\lambda(\lambda+k)}{k}\left\{\frac{\Psi\left(1+\frac{\lambda}{k}\right)+\gamma}{\lambda} - \frac{\Psi\left(2+\frac{\lambda}{k}\right)+\gamma}{\lambda+k}\right\}, \tag{23}$$

where Ψ is the digamma function and γ is Euler's constant. Unlike the Rényi entropy, H_S is finite for every $\lambda, k > 0$: $E(S)$ is finite because S is hypoexponential, and the series for $E\{-\log(1-e^{-kS})\}$ in the proof converges absolutely by comparison with $\sum_j j^{-2}$.

Proof. Writing $S = \log\left(\frac{X}{\theta}\right)$, the density (5) gives $-\log g(X) = \log\left\{\frac{k\theta}{\lambda(\lambda+k)}\right\} + (\lambda+1)S - \log(1-e^{-kS})$. By (16), $E(S) = \lambda^{-1} + (\lambda+k)^{-1}$, which supplies the second term of $H_S = E\{-\log g(X)\}$. Expand $-\log(1-e^{-kS}) = \sum_{j\geq 1} j^{-1}\,e^{-jkS}$ and take expectations term by term. Since S is hypoexponential with rates λ and $\lambda+k$, $E\left(e^{-jkS}\right) = \frac{\lambda(\lambda+k)}{(\lambda+jk)(\lambda+(j+1)k)}$.

A partial-fraction decomposition in j reduces $\sum_{j\geq 1} j^{-1}\,E\left(e^{-jkS}\right)$ to

$$\left\{\frac{\lambda(\lambda+k)}{k}\right\}\left[\frac{\Psi\left(1+\frac{\lambda}{k}\right)+\gamma}{\lambda} - \frac{\Psi\left(2+\frac{\lambda}{k}\right)+\gamma}{\lambda+k}\right].$$

Adding this to the first two terms gives (23).

2.6 Order Statistics and Stress-Strength Reliability

Let $X_{1:n} \le \cdots \le X_{n:n}$ be the order statistics of a random sample from NBP(λ, k, θ). The density of $X_{i:n}$ is $g_{i:n}(x) = \left\{\frac{n!}{(i-1)!(n-i)!}\right\} g(x)G(x)^{i-1}\bar{G}(x)^{n-i}$, which is elementary. Using $g\bar{G}^N = -(N+1)^{-1}\frac{\mathrm{d}\bar{G}^{N+1}}{\mathrm{d}x}$, integrating by parts and expanding $\bar{G}^{N+1}$ binomially, we obtain for $r < (N+1)\lambda$

$$J_N(r) = \int_\theta^\infty x^r\, g\, \bar{G}^N \mathrm{d}x = \frac{\theta^r}{N+1}\left\{1 + r\sum_{j=0}^{N+1}\binom{N+1}{j}\frac{\left(\frac{\lambda+k}{k}\right)^{N+1-j}\left(-\frac{\lambda}{k}\right)^j}{(N+1)\lambda + jk - r}\right\}. \tag{24}$$

Expanding $G^{i-1} = (1-\bar{G})^{i-1}$ then gives the exact finite-sum moments, with $\bar{G}_{1:n} = \bar{G}^{\,n}$ in particular, so the sample minimum has tail index $n\lambda$:

$$E(X_{i:n}^r) = \frac{n!}{(i-1)!\,(n-i)!}\sum_{m=0}^{i-1}(-1)^m\binom{i-1}{m}J_{n-i+m}(r). \tag{25}$$

Let the stress $X_1 \sim$ NBP(λ_1, k_1, θ) and the strength $X_2 \sim$ NBP(λ_2, k_2, θ) be independent and share a common threshold θ. Since every term of $g_1\bar{G}_2$ is a power of y, the reliability $R = P(X_1 < X_2)$ is

$$R = \lambda_1(\lambda_1 + k_1)\left\{\frac{\lambda_2 + k_2}{k_2(\lambda_1+\lambda_2)(\lambda_1+\lambda_2+k_1)} - \frac{\lambda_2}{k_2(\lambda_1+\lambda_2+k_2)(\lambda_1+\lambda_2+k_1+k_2)}\right\}. \tag{26}$$

This expression equals $\frac{1}{2}$ when the two distributions coincide, and it tends to the Pareto-strength value as $k_2 \to \infty$. A common threshold is essential to the closed form; with $\theta_1 \neq \theta_2$ the two supports no longer align and R must be evaluated by numerical integration.

2.7 Parameter Estimation

Maximum likelihood. For a random sample $x_1, \ldots, x_n$ with $\theta < x_{1:n}$ and $y_i = \frac{x_i}{\theta}$, the log-likelihood is

$$\ell(\lambda, k, \theta) = n\log\lambda + n\log(\lambda+k) - n\log\theta - (\lambda+1)\sum_{i=1}^n \log y_i + \sum_{i=1}^n \log \tau_k(y_i), \tag{27}$$

with score functions

$$\frac{\partial\ell}{\partial\lambda} = \frac{n}{\lambda} + \frac{n}{\lambda+k} - \sum_{i=1}^n \log y_i, \qquad \frac{\partial\ell}{\partial k} = \frac{n}{\lambda+k} - \frac{n}{k} + \sum_{i=1}^n \frac{\log y_i}{y_i^k - 1},$$
$$\frac{\partial\ell}{\partial\theta} = \frac{n\lambda}{\theta} - \frac{k}{\theta}\sum_{i=1}^n \frac{1}{y_i^k - 1}. \tag{28}$$

Theorem 7. For fixed (k, θ), let $T = n^{-1}\sum_i \log\left(\frac{x_i}{\theta}\right)$. The log-likelihood is maximised in λ at

$$\hat{\lambda}(k, \theta) = \frac{2 - kT + \sqrt{k^2T^2 + 4}}{2T}. \tag{29}$$

Proof. Setting $\frac{\partial\ell}{\partial\lambda} = 0$ in (28) gives $\lambda^{-1} + (\lambda+k)^{-1} = T$. Clearing denominators gives the quadratic $T\lambda^2 + (kT-2)\lambda - k = 0$.

The product of its roots is $-\frac{k}{T} < 0$, so exactly one root is positive, and its discriminant is $k^2T^2 + 4$. Solving the quadratic for the positive root gives (29). Since $\frac{\partial^2 \ell}{\partial \lambda^2} = -n\lambda^{-2} - n(\lambda + k)^{-2} < 0$, this root is the maximiser. ▫

As $k \to 0$, (29) tends to the gamma-Pareto estimator $\frac{2}{T}$, and as $k \to \infty$ it tends to the Pareto estimator $\frac{1}{T}$. Substituting (29) into (27) leaves a two-dimensional profile likelihood in (k, θ), which we maximise by the Nelder-Mead method from a grid of starting values.

Information and inference. Summing the series $E\{s^2 e^{-ks}(1 - e^{-ks})^{-2}\}$ under (7) gives the per-observation expected information for (λ, k):

$$\mathcal{I} = \begin{pmatrix} \lambda^{-2} + (\lambda + k)^{-2} & (\lambda + k)^{-2} \\ (\lambda + k)^{-2} & (\lambda + k)^{-2} - k^{-2} + \dfrac{2\lambda(\lambda + k)}{k^4}\, \zeta\left(3, 1 + \dfrac{\lambda}{k}\right) \end{pmatrix}, \tag{30}$$

where $\zeta(3, a) = \sum_{j \geq 0} (j + a)^{-3}$ is the Hurwitz zeta function. Standard errors follow from $(n\mathcal{I})^{-1}$. The threshold makes the problem non-regular. Near θ, $g(x) \sim c\,(x - \theta)^{\gamma - 1}$ with $\gamma = 2$, and for this case Smith (1985) shows that the maximum likelihood estimator of θ exists and is asymptotically normal at rate $(n \log n)^{\frac{1}{2}}$, while the other parameters keep their regular theory. For the BP, exponentiated Pareto, Kumaraswamy-Pareto and gamma-Pareto models, the density is unbounded at the threshold when the boundary shape parameter is below one, and the likelihood then has no maximum. The NBP likelihood is always bounded. Because the Pareto law is the boundary case $k = \infty$, the likelihood ratio statistic for this sub-model is referred to the mixture $1/2\,\chi_0^2 + 1/2\,\chi_1^2$ (Self & Liang, 1987).

3. RESULTS AND DISCUSSION

3.1 Simulation Study

Samples were generated by the exact sampler of Theorem 2 for $(\lambda, k, \theta) = (1.5,3,1)$, $(2,6,1)$ and $(3,12,1)$, with $n \in \{50,100,250,500\}$ and 1000 replications per setting. Each sample was fitted by the profile method with fixed starting values. Table 2 reports the bias and root mean squared error (RMSE) of $\hat{\lambda}$, $\hat{\theta}$ and $\hat{\eta} = \frac{\hat{\lambda}}{\hat{\lambda} + \hat{k}}$, and the coverage probability (CP) of the 95% Wald interval for λ. For k, whose estimator can diverge in small samples, it reports the median estimate and the percentage of boundary fits ($\hat{k} < 10^{-4}$ or $\hat{k} > 10^3$), a threshold chosen to be at least two orders of magnitude beyond the simulated parameter range (k = 3 to 18 in Table 2), so as to flag only clear optimiser excursions rather than to calibrate a precise scientific criterion.

Table 2. Monte Carlo results (1000 replications): bias and RMSE of $\hat{\lambda}$, $\hat{\theta}$ and $\hat{\eta} = \frac{\hat{\lambda}}{\hat{\lambda} + \hat{k}}$, coverage probability (CP) of the nominal 95% Wald interval for λ, median of $\hat{k}$ and percentage of boundary fits for k.

(λ, k, θ)	n	Bias $\hat{\lambda}$	RMSE $\hat{\lambda}$	CP $\hat{\lambda}$	Bias $\hat{\theta}$	RMSE $\hat{\theta}$	Bias $\hat{\eta}$	RMSE $\hat{\eta}$	Median $\hat{k}$	Boundary (%)
$(1.5,3,1)$	50	0.1400	0.4846	0.650	0.0463	0.0718	0.0347	0.3746	4.690	47.0
	100	0.0887	0.3600	0.773	0.0214	0.0418	0.0452	0.3028	3.593	21.3
	250	0.0521	0.2269	0.920	0.0080	0.0182	0.0278	0.1886	3.217	3.8
	500	0.0201	0.1413	0.945	0.0046	0.0119	0.0110	0.1164	3.086	0.8
$(2,6,1)$	50	0.2051	0.6398	0.743	0.0296	0.0475	0.0457	0.3420	9.008	43.7
	100	0.1077	0.4477	0.819	0.0139	0.0278	0.0363	0.2610	7.090	17.7
	250	0.0417	0.2442	0.941	0.0056	0.0126	0.0088	0.1376	6.451	1.8
	500	0.0147	0.1396	0.963	0.0030	0.0079	0.0023	0.0738	6.187	0.0

(λ, k, θ)	n	Bias $\hat{\lambda}$	RMSE $\hat{\lambda}$	CP $\hat{\lambda}$	Bias $\hat{\theta}$	RMSE $\hat{\theta}$	Bias $\hat{\eta}$	RMSE $\hat{\eta}$	Median $\hat{k}$	Boundary (%)
$(3,12,1)$	50	0.3038	0.9093	0.759	0.0180	0.0287	0.0378	0.3013	17.506	43.2
	100	0.1390	0.6206	0.855	0.0084	0.0170	0.0275	0.2315	14.153	18.9
	250	0.0471	0.3150	0.961	0.0035	0.0078	0.0028	0.1084	13.086	1.4
	500	0.0171	0.1865	0.951	0.0019	0.0048	-0.0001	0.0572	12.435	0.0

The tail index is estimated well. Its bias is about 10% of the true value at $n = 50$ and about 1% at $n = 500$. A log-log regression of RMSE on n over the four sample sizes gives a slope of -0.53 for $(\lambda, k) = (1.5,3)$, steepening to -0.66 and -0.70 for $(2,6)$ and $(3,12)$; the $n^{-\frac{1}{2}}$ rate is thus a conservative lower bound on the observed convergence, not a uniform description. The threshold estimator is faster still, with slopes of -0.79 to -0.80 across the three configurations -- steeper than first-order asymptotics predict. Smith's (1985) non-regular rate for the threshold is the square root of n log n, not of n alone; its log-log slope over n = 50 to 500 is close to only −0.60, so the observed −0.79 to −0.80 reflects faster finite-sample convergence than the theory by itself implies. It remains, in any event, a non-regular, better-than-$\sqrt{n}$ theory. The body parameter k is harder to estimate. At $n = 50$ between 43% and 47% of fits lie on a boundary, and the coverage for λ is only 0.65 to 0.76. From $n = 250$ onwards, boundary fits are rare, the median of $\hat{k}$ is within 10% of the truth, and coverage lies between 0.92 and 0.96. This behaviour is typical of second-order parameters in extreme-value theory (Hall & Welsh, 1985). Figure 3 plots the RMSEs.

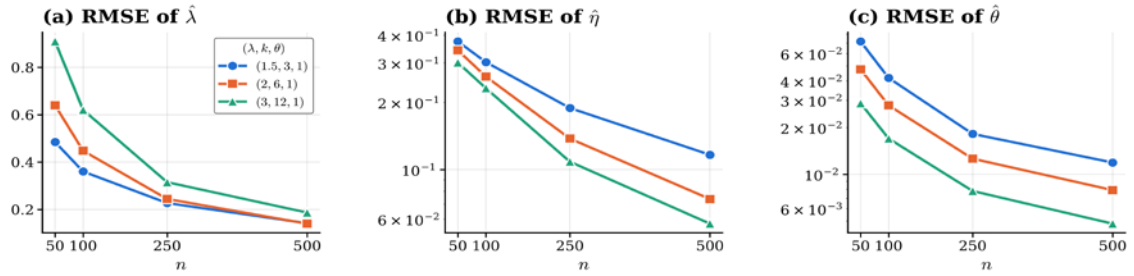

Figure 3. *Monte Carlo RMSE of $\hat{\lambda}$, $\hat{\eta}$ and $\hat{\theta}$ against sample size (1000 replications per point).*

3.2 The Data

Two exceedance samples are analysed, chosen to test the NBP honestly against data of differing body shape. The first is the 154 exceedances of 65 $\frac{\mathrm{m}^3}{\mathrm{s}}$ by the River Nidd at Hunsingore, Yorkshire, originally reported in the UK Flood Studies Report of 1975 and analysed by Davison and Smith (1990); it is distributed as nidd.thresh in the R package **evir** and listed in ascending order in Table 3. The second is the 109 losses exceeding 10 million Danish krone among the large fire-insurance claims recorded in Denmark from 1980 to 1990, supplied by Mette Rytgaard of Copenhagen Re and analysed by McNeil (1997); it is distributed as danish in the same package and listed in ascending order in Table 4. Table 5 gives descriptive statistics for both samples. The Nidd sample rises steeply just above its threshold before decaying slowly, the pattern the NBP is designed for; the Danish sample decays closer to a pure power law from the threshold outward.

Table 3. River Nidd exceedances of 65 m^3/s (n = 154), listed in ascending order.

65.08	**65.60**	**66.53**	**67.02**	**67.02**	**67.15**	**67.15**	**67.28**	**67.67**	**68.45**	**68.45**
68.84	69.49	69.49	69.49	69.62	70.01	70.14	70.49	70.52	70.78	71.91
72.21	72.34	72.34	72.34	72.34	73.61	73.61	74.02	74.15	74.54	74.54
74.80	74.93	74.93	74.93	75.02	75.06	75.87	75.97	76.22	76.22	76.22
76.22	76.61	76.61	76.72	76.74	77.00	77.13	77.26	77.26	77.52	77.52

65.08	**65.60**	**66.53**	**67.02**	**67.02**	**67.15**	**67.15**	**67.28**	**67.67**	**68.45**	**68.45**
77.78	78.04	78.17	78.29	78.29	78.55	78.81	78.81	78.98	78.98	79.33
79.33	79.46	80.11	80.12	80.75	81.27	81.40	81.40	81.40	81.40	81.40
81.65	82.54	83.56	83.69	83.94	85.85	86.48	86.48	86.73	87.19	87.75
87.76	87.76	87.76	88.04	88.76	88.89	89.02	89.46	89.65	90.28	90.88
91.55	91.80	91.80	92.82	92.82	94.08	95.47	96.25	96.86	97.24	97.87
97.87	99.08	99.14	99.14	99.93	100.40	102.92	104.19	104.19	104.19	107.58
107.97	108.14	110.48	110.98	111.54	111.74	115.52	116.77	119.28	121.73	123.71
131.82	131.92	133.06	138.72	143.06	148.63	149.30	151.79	153.04	157.12	158.01
162.99	172.92	179.12	181.59	189.02	213.70	226.48	251.96	257.62	261.82	305.75

Table 4. Danish fire-insurance losses exceeding 10 million Danish krone, 1980-1990 (n = 109), listed in ascending order.

10.01	**10.07**	**10.14**	**10.18**	**10.18**	**10.20**	**10.22**	**10.27**	**10.50**	**10.58**	**10.70**
10.82	10.83	11.00	11.12	11.13	11.37	11.43	11.60	11.62	11.69	11.71
11.80	11.89	12.05	12.06	12.22	12.38	12.47	12.52	12.54	12.63	12.70
12.80	12.90	13.20	13.35	13.50	13.62	13.62	14.01	14.12	14.24	14.29
14.30	14.39	14.39	14.68	14.85	15.28	15.81	15.88	15.93	16.09	16.30
16.42	16.44	16.88	17.07	17.57	17.74	17.74	18.32	18.42	18.55	18.63
18.65	18.85	19.07	19.16	19.27	19.40	19.47	20.05	20.45	20.83	20.86
20.97	21.96	22.14	22.26	23.28	24.56	24.58	24.97	25.29	25.95	26.21
27.26	27.34	27.83	28.63	29.03	29.04	31.06	32.39	32.47	34.14	38.15
42.09	46.50	47.02	50.07	56.23	57.41	65.71	144.66	152.41	263.25	

Table 5. Descriptive statistics for the two exceedance samples.

Data	n	Min	Q_1	Median	Mean	Q_3	Max	SD	Skewness	Kurtosis (raw)
River Nidd	154	65.08	75.26	81.53	97.87	100.28	305.75	40.94	2.667	10.993
Danish fire losses	109	10.01	12.38	16.30	24.08	23.28	263.25	30.87	5.658	40.464

3.3 Model Comparison

On each sample the NBP is compared with the Pareto, Exponentiated Pareto (EP) (Gupta et al., 1998), Gamma-Pareto (GaP) (Alzaatreh et al., 2012), Beta-Pareto (BP) (Akinsete et al., 2008), Kumaraswamy-Pareto (KwP) (Bourguignon et al., 2013), generalised Pareto (Hosking & Wallis, 1987) and three-parameter Weibull (W3) distributions. All parameters, including thresholds and locations, were estimated by maximum likelihood with the threshold below the sample minimum. As discussed in Section 2.7, the boundary shape parameter of EP, GaP, BP, KwP and W3 was restricted to be at least one so that every likelihood is bounded. Each model was fitted from a grid of starting values. We report the maximised log-likelihood $\hat{\ell}$ and, with q the number of estimated parameters (threshold or location included) and n the sample size, $\mathrm{AIC} = -2\hat{\ell} + 2q$ (Akaike, 1974), $\mathrm{AICc} = \mathrm{AIC} + \frac{2q(q+1)}{n-q-1}$ (Hurvich & Tsai, 1989), $\mathrm{BIC} = -2\hat{\ell} + q\log n$ (Schwarz, 1978) and $\mathrm{HQIC} = -2\hat{\ell} + 2q\log\log n$, together with the Kolmogorov-Smirnov (D), Cramér-von Mises (W) and Anderson-Darling (A) statistics (Anderson & Darling, 1952; Stephens, 1974); a boundary-restricted shape parameter is still counted in q.

Table 6. Fitted parameters, log-likelihood, information criteria and goodness-of-fit statistics for the River Nidd data.

Model (q)	Estimates	$\hat{\ell}$	AIC	AICc	BIC	HQIC	D (p)	W	A
NBP (3)	λ = 3.342, k = 13.967, θ = 64.593	−683.61	**1373.21**	**1373.37**	**1382.32**	**1376.91**	0.069 (0.443)	0.114	0.622
EP (3)	a = 1.484, k = 3.622, θ = 64.906	−684.52	1375.04	1375.20	1384.15	1378.74	0.078 (0.284)	0.161	0.875
BP (4)	a = 2.348, b = 0.205, k = 16.327, θ = 64.324	−683.57	1375.14	1375.40	1387.28	1380.07	0.068 (0.446)	0.109	0.598
KwP (4)	a = 3.062, b = 0.216, k = 15.505, θ = 64.020	−683.66	1375.31	1375.58	1387.46	1380.25	0.068 (0.451)	0.109	0.599
GaP (3)	a = 1.409, c = 0.249, θ = 64.948	−684.82	1375.64	1375.80	1384.75	1379.34	0.079 (0.273)	0.176	0.957
GPD (3)	ξ = 0.205, σ = 26.091, μ = 65.080	−687.89	1381.78	1381.94	1390.89	1385.48	0.096 (0.110)	0.300	1.963

Model (q)	Estimates	$\hat{\ell}$	AIC	AICc	BIC	HQIC	D (p)	W	A
Pareto (2)	λ = 2.861, θ = 65.080	−688.96	1381.92	1382.00	1388.00	1384.39	0.122 (0.019)	0.414	2.631
W3 (3)	c = 1.000†, σ = 32.788, μ = 65.080	−691.47	1388.94	1389.10	1398.05	1392.64	0.110 (0.045)	0.521	2.823

The NBP ranks first on all four information criteria for the Nidd data (Table 6). The four-parameter BP and KwP reach essentially the same log-likelihood, so their extra parameter brings no gain. The Pareto sub-model is decisively rejected: $\Lambda = 10.71$, with $p \approx 0.0005$ under the boundary mixture. The estimates are $\hat{\lambda} = 3.342$ (standard error 0.352), $\hat{k} = 13.97$ (Wald standard error 6.45; profile-likelihood 95% CI 5.34 to 37.68 -- a wide interval reflecting weak identification of the body parameter at this sample size) and $\hat{\theta} = 64.593$ (profile-likelihood 95% CI 63.37 to 65.05). The fitted mean is 97.82 (sample 97.87) and the fitted median is 84.64 (sample 81.53). The density mode (8) is at 71.60, and by (12) the hazard peaks at $x_0 = 77.37\ \frac{\mathrm{m}^3}{\mathrm{s}}$. Beyond that flow the conditional risk of a still larger flood decline like $\frac{\lambda}{x}$. A parametric bootstrap with $B = 500$ resamples confirms the fit, lowering the Kolmogorov-Smirnov p-value from the asymptotic 0.443 to a still-comfortable 0.100 once the cost of estimating $(\hat{\lambda}, \hat{k}, \hat{\theta})$ is accounted for. A likelihood-ratio test of the restriction that fixes the first beta-Pareto generator parameter at two gives a statistic of 0.08 (p ≈ 0.78 on one degree of freedom), so the extra parameter of the four-parameter beta-Pareto model is not supported by these data, and the three-parameter NBP is preferred on grounds of parsimony as well as fit. The two tail indices also matter for extrapolation: using the record's 35-year span, that is form 1934 to 1969 (Davison & Smith, 1990) and its 154 exceedances, the 100-year return level is 425.6 cubic metres per second under the NBP against 546.3 under the plain Pareto fit, even though the two models agree closely over the observed range. Figure 4 confirms the close agreement between the fitted model and the data across the whole range.

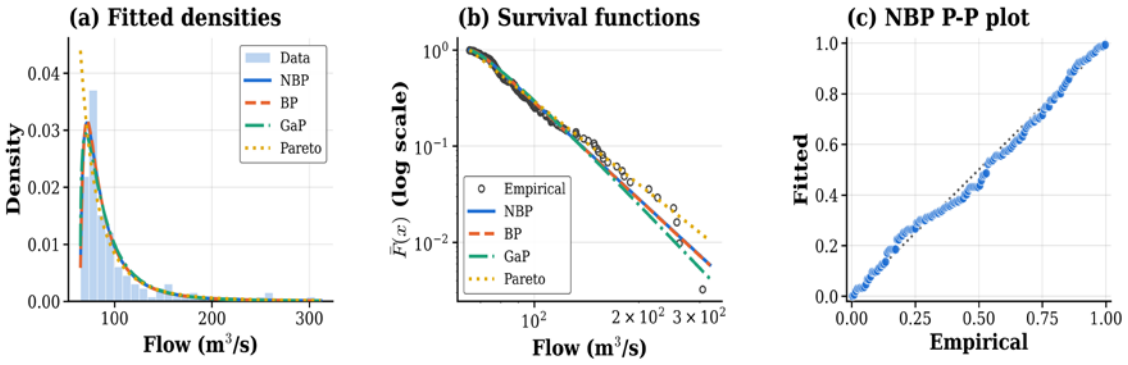

Figure 4. *River Nidd exceedances: (a) histogram with fitted densities, (b) empirical and fitted survival functions on log-log axes, (c) probability-probability plot for the NBP.*

Table 7. Fitted parameters, log-likelihood, information criteria and goodness-of-fit statistics for the Danish fire-loss data.

Model (q)	Estimates	$\hat{\ell}$	AIC	AICc	BIC	HQIC	D (p)	W	A
Pareto (2)	λ = 1.617, θ = 10.011	−375.10	**754.20**	**754.31**	**759.58**	**756.38**	0.064 (0.743)	0.086	0.643
GPD (3)	ξ = 0.499, σ = 6.950, μ = 10.011	−374.72	755.44	755.67	763.51	758.71	0.044 (0.980)	0.033	0.427
NBP (3)	λ = 1.617, k = 2.89e16†, θ = 10.011	−375.10	756.20	756.43	764.27	759.47	0.064 (0.743)	0.086	0.643
GaP (3)	a = 1.000†, c = 0.618, θ = 10.011	−375.10	756.20	756.43	764.27	759.47	0.064 (0.743)	0.086	0.643
EP (3)	a = 1.000†, k = 1.617, θ = 10.011	−375.10	756.20	756.43	764.27	759.47	0.064 (0.743)	0.086	0.643

Model (q)	Estimates	$\hat{\ell}$	AIC	AICc	BIC	HQIC	D (p)	W	A
KwP (4)	a = 1.000†, b = 0.145, k = 11.113, θ = 10.011	−374.75	757.51	757.89	768.27	761.87	0.064 (0.746)	0.085	0.642
BP (4)	a = 1.000†, b = 6.71e5†, k → 0†, θ = 10.011	−375.10	758.20	758.58	768.96	762.56	0.064 (0.743)	0.086	0.643
W3 (3)	c = 1.000†, σ = 14.071, μ = 10.011	−397.21	800.41	800.64	808.49	803.69	0.180 (0.001)	1.389	7.848

Throughout Tables 6 and 7, † marks a parameter restricted to the boundary of its admissible range — a shape parameter fixed at its lower bound of one (Section 2.7), or an estimate at the optimiser's upper bound for k — rather than a freely attained maximum-likelihood value; the corresponding row is better read as a degenerate Pareto, shifted-exponential, or boundary fit than as an independent competitor.

For the Danish fire-loss data (Table 7), the plain Pareto model already gives the smallest AIC, and the NBP, GaP and EP are statistically indistinguishable from it and from each other: each reduces to the Pareto sub-model, with $\hat{k} \to \infty$ for the NBP; indeed the profile likelihood for k has no finite upper bound here, its 95% profile-likelihood interval being one-sided (k ≥ 34.6), consistent with these data sitting at the Pareto boundary rather than pinning down a finite body parameter. The extra parameter is harmless here rather than helpful, costing only the usual two-unit AIC penalty for no gain in fit, and the NBP ranks third of eight, tied in practice with GaP and EP. Figure 5 shows the three curves overlapping almost exactly. This dataset illustrates the boundary case opposite to Nidd: when the tail is already close to pure power-law from the threshold outward, the NBP correctly collapses toward its Pareto sub-model instead of overfitting. A parametric bootstrap with $B = 500$ resamples gives a p-value (0.532) close to its asymptotic counterpart (0.743), as expected away from a parameter boundary.

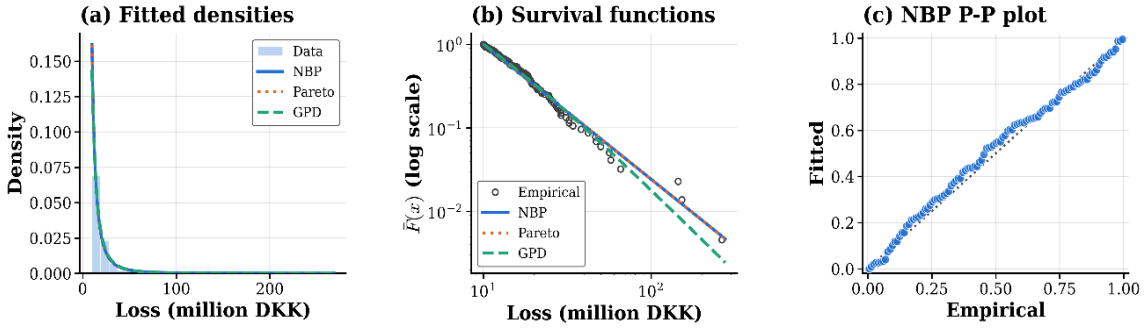

Figure 5. *Danish fire-loss exceedances: (a) histogram with fitted densities, (b) empirical and fitted survival functions on log-log axes, (c) probability-probability plot for the NBP*

3.4 Discussion

Table 8 summarises the analytical position of the NBP relative to the BP and gamma-Pareto models, with the AIC rank on each sample. The NBP trades some generality for tractability: closed-form reliability functions, moments and entropies, a fixed interpretable hazard shape, and a likelihood that is always bounded. Whether this trade pays off depends on the data. The two applications sit at opposite ends of the η spectrum: η ≈ 0.19 for River Nidd against an η indistinguishable from zero for the Danish losses, whose tail parameter k is estimated far into the Pareto limit. River Nidd, whose body decays gradually enough to sit inside the admissible range of $\mathrm{CV}\left\{\log\left(\frac{X}{\theta}\right)\right\}$, fits better than every competitor, confirmed by the parametric bootstrap of Section 3.3; the Danish losses sit at the boundary where the extra parameter is redundant, and the NBP correctly falls back to the Pareto law it contains. The bound (17) is a necessary, not a sufficient, condition: data whose log-scale dispersion falls

below it (a faster-decaying body just above the threshold) cannot be matched by any choice of (λ, k), a property of the model rather than a fitting failure -- but the bound does not by itself separate the two applications studied here, since the sample log-scale dispersion is 0.880 for River Nidd and 0.944 for the Danish losses, both inside the admissible range given by (17); which dataset the NBP actually improves on is decided by the fit reported in Section 3.3, not by the bound alone. The asymptotic $D(p)$ in Tables 6 and 7 is the standard diagnostic used throughout; the bootstrap correction is reported only for the NBP, for compute-budget reasons.

Table 8. Analytical comparison of the NBP with the beta-Pareto and gamma-Pareto models, and AIC rank (of eight fitted models) on each application.

Feature	Beta-Pareto (BP)	Gamma-Pareto (GaP)	Nested Beta-Pareto (NBP)
Parameters	4	3	3
Survival function	Incomplete beta function	Incomplete gamma function	Elementary, Eq. (6)
Hazard, cumulative hazard, odds, MRL	Special functions	Special functions	Elementary
Hazard shapes	Increasing, decreasing or unimodal	Increasing, decreasing or unimodal	Always unimodal (UBT), explicit mode
Density at threshold	0, finite or infinite	0, finite or infinite	Always 0 (bounded likelihood)
Tail index	kb	$\frac{1}{c}$	λ
Moments exist for	$r < kb$	$r < \frac{1}{c}$	$r < \lambda$
Order-statistic moments	Numerical	Numerical	Finite sums, Eq. (25)
Stress-strength R	Numerical	Numerical	Closed form, Eq. (26)
Explicit likelihood root	None	For the scale parameter	$\hat{\lambda}(k, \theta)$, Eq. (29)
Log-scale CV of $\frac{X}{\theta}$	Unrestricted	$\frac{1}{\sqrt{a}}$	Restricted: $\frac{1}{\sqrt{2}} \leq \mathrm{CV} < 1$
AIC rank (River Nidd)	3	5	**1**
AIC rank (Danish fire losses)	7	4	**3**

3.5 Limitations

The trade made explicit in Table 8 has four concrete consequences. First, the hazard is always upside-down bathtub shaped (Theorem 3): the model cannot represent monotone, bathtub, or constant hazards, unlike the four-parameter BP and KwP it nests within. Second, by (17) the log-scale dispersion of $\frac{X}{\theta}$ is confined to $\frac{1}{\sqrt{2}} \leq \mathrm{CV}\left\{\log\left(\frac{X}{\theta}\right)\right\} < 1$; data with lighter dispersion cannot be matched by any (λ, k), a property of the model, not a fitting failure. Third, inference is conditional on a single threshold chosen below the sample minimum; no selection rule is proposed here, and a different threshold would give different results. Fourth, the profile likelihood in k is flat and $\hat{k}$ is poorly determined near the Pareto ($k \to \infty$) or gamma-Pareto ($k \to 0$) boundary, as the Danish fit illustrates; interval estimates for k near these boundaries should be read as qualitative. The model should therefore be used where a soft threshold onset and power-law tail are both plausible on subject-matter grounds, not adopted by default in place of the wider beta-Pareto class.

4. CONCLUSION

We have studied the subfamily of the beta-Pareto distribution obtained by fixing the first generator parameter at two and writing the second as $\frac{\lambda}{k}$, and shown that it is not a reparameterisation of an existing elementary model (Section 2.1) but a genuine subfamily whose survival function is exactly elementary, with the reliability functions, moments, entropies, order-statistic moments and stress-strength reliability all in closed form. Its tail is an exact member of Hall's second-order class with exponent $-k$, and the hazard rate is upside-down bathtub shaped with an explicit mode and an explicit likelihood root for the tail index. The simulation study confirms accurate estimation of λ and θ, with empirical

convergence rates at least as fast as $n^{-\frac{1}{2}}$ and $n^{-0.79}$. Across two real applications, the model gives the best fit among eight competitors, confirmed by a parametric bootstrap, for the River Nidd flood exceedances, and is statistically indistinguishable from the Pareto law it contains for the Danish fire losses; the dispersion bound (17) gives a necessary, though not sufficient, condition for the model's applicable range, and Section 3.5 states the resulting limitations formally. Natural extensions include integer first generator parameters $m > 2$, censored and Bayesian inference built on the elementary survival function, and exact bias analysis of tail-index estimators within Hall's class.

Ethics Committee Approval

Not applicable. This study uses exclusively secondary, publicly available data, the River Nidd flood exceedances and the Danish fire-insurance losses distributed in the R package **evir**, and involves no new data collection from human participants or animals. Both datasets are already in the public domain and have been analysed and published by other researchers (Davison & Smith, 1990; McNeil, 1997), so no ethics committee approval was required for this work.

Data Availability and Reproducibility

The River Nidd exceedances are available as nidd.thresh in the R package **evir** and are listed in Table 3. The Danish fire-loss data are available as danish in the same package and are listed in Table 4. The R code implementing the NBP density, survival, quantile and exact random-generation functions, and the profile-likelihood estimator, is provided as Supplementary Material accompanying this article (density integrates to 1 by numerical quadrature, the quantile function inverts the survival function to machine precision, and maximum likelihood recovers known generating parameters on simulated data), together with the simulation scripts, the random seed, optimiser settings and convergence tolerances used to generate Tables 2 and 6-8.

5. REFERENCES

Aga, M. A. (2026). A new extended Pareto distribution: Statistical properties, estimation, and applications. The Scientific World Journal, 2026, Article 6572559. https://doi.org/10.1155/tswj/6572559

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723. https://doi.org/10.1109/TAC.1974.1100705

Akinsete, A., Famoye, F., & Lee, C. (2008). The beta-Pareto distribution. Statistics, 42, 547-563. https://doi.org/10.1080/02331880801983876

Alzaatreh, A., Famoye, F., & Lee, C. (2012). Gamma-Pareto distribution and its applications. Journal of Modern Applied Statistical Methods, 11, 78-94. https://doi.org/10.22237/jmasm/1335845160

Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. Metron, 71, 63-79. https://doi.org/10.1007/s40300-013-0007-y

Alzaatreh, A., Famoye, F., & Lee, C. (2013). Weibull-Pareto distribution and its applications. Communications in Statistics - Theory and Methods, 42, 1673-1691. https://doi.org/10.1080/03610926.2011.599002

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Annals of Mathematical Statistics, 23, 193-212. https://doi.org/10.1214/aoms/1177729437

Ayenigba, A. A., Ajao, O. M., & Okolie, F. A. (2025). Sum of Poisson-distributed random variables: A convolution method approach. Journal of Applied Sciences and Environmental Management, 29(2), 401-405. https://doi.org/10.4314/jasem.v29i2.8

Ayenigba, A. A., Ajao, O. M., & Okpechi, W. C. (2026). Probability generating functions: Theory and applications to distribution generation, branching processes and queueing model. Diophantine Journal of Mathematics and Its Applications, 4(2), 92-103. https://doi.org/10.33369/diophantine.v4i2.47677

Ayenigba, A. A., & Ogunkoya, C. O. (2025). Analytical properties of the uniform exponential distribution. Mikailalsys Journal of Advanced Engineering International, 2(3), 359-374. https://doi.org/10.58578/mjaei.v2i3.7188

Baharith, L. A., AL-Beladi, K. M., & Klakattawi, H. S. (2020). The odds exponential-Pareto IV distribution: Regression model and application. Entropy, 22(5), Article 497. https://doi.org/10.3390/e22050497

Bourguignon, M., Silva, R. B., Zea, L. M., & Cordeiro, G. M. (2013). The Kumaraswamy Pareto distribution. Journal of Statistical Theory and Applications, 12, 129-144. https://doi.org/10.2991/jsta.2013.12.2.1

Chhetri, S. B., Akinsete, A., Aryal, G., & Long, H. (2017). The Kumaraswamy transmuted Pareto distribution. Journal of Statistical Distributions and Applications, 4(1), Article 11. https://doi.org/10.1186/s40488-017-0065-4

Cirillo, P. (2013). Are your data really Pareto distributed? Physica A: Statistical Mechanics and its Applications, 392, 5947-5962. https://doi.org/10.1016/j.physa.2013.07.061

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. SIAM Review, 51, 661-703. https://doi.org/10.1137/070710111

Cordeiro, G. M., & de Castro, M. (2011). A new family of generalized distributions. Journal of Statistical Computation and Simulation, 81, 883-898. https://doi.org/10.1080/00949650903530745

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). Journal of the Royal Statistical Society: Series B, 52(3), 393-442. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. Communications in Statistics - Theory and Methods, 31, 497-512. https://doi.org/10.1081/STA-120003130

Ghitany, M. E., Gómez-Déniz, E., & Nadarajah, S. (2018). A new generalization of the Pareto distribution and its application to insurance data. Journal of Risk and Financial Management, 11(1), Article 10. https://doi.org/10.3390/jrfm11010010

Glaser, R. E. (1980). Bathtub and related failure rate characterizations. Journal of the American Statistical Association, 75, 667-672. https://doi.org/10.1080/01621459.1980.10477530

Gupta, R. C., Gupta, P. L., & Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. Communications in Statistics - Theory and Methods, 27, 887-904. https://doi.org/10.1080/03610929808832134

Haj Ahmad, H., & Almetwally, E. (2020). Marshall-Olkin generalized Pareto distribution: Bayesian and non-Bayesian estimation. Pakistan Journal of Statistics and Operation Research, 16(1), 21-33. https://doi.org/10.18187/pjsor.v16i1.2935

Hall, P. (1982). On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society: Series B, 44, 37-42. https://doi.org/10.1111/j.2517-6161.1982.tb01183.x

Hall, P., & Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. Annals of Statistics, 13, 331-341. https://doi.org/10.1214/aos/1176346596

Hamed, D., Famoye, F., & Lee, C. (2018). On families of generalized Pareto distributions: Properties and applications. Journal of Data Science, 16, 377-396. https://doi.org/10.6339/jds.201804_16(2).0008

Hmingthansanga, Jha, S. K., Das, B., & Kumar, D. (2026). A gamma generated new Pareto-type distribution: A comprehensive mathematical framework with inferential issues, simulations and implementation. Statistics and Computing, 36, Article 146. https://doi.org/10.1007/s11222-026-10904-x

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics, 29, 339-349. https://doi.org/10.1080/00401706.1987.10488243

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307. https://doi.org/10.1093/biomet/76.2.297

Ihtisham, S., Khalil, A., Manzoor, S., Khan, S. A., & Ali, A. (2019). Alpha-power Pareto distribution: Its properties and applications. PLoS ONE, 14, Article e0218027. https://doi.org/10.1371/journal.pone.0218027

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. Test, 13, 1-43. https://doi.org/10.1007/BF02602999

Karakaya, K., Akdoğan, Y., Saadati Nik, A., Kuş, C., & Asgharzadeh, A. (2024). A generalization of new Pareto-type distribution. Annals of Data Science, 11(1), 87-101. https://doi.org/10.1007/s40745-022-00376-x

Lee, C., Famoye, F., & Alzaatreh, A. (2013). Methods for generating families of univariate continuous distributions in the recent decades. WIREs Computational Statistics, 5, 219-238. https://doi.org/10.1002/wics.1255

Mahdavi, A., & Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. Communications in Statistics - Theory and Methods, 46, 6543-6557. https://doi.org/10.1080/03610926.2015.1130839

Mahmoudi, E. (2011). The beta generalized Pareto distribution with application to lifetime data. Mathematics and Computers in Simulation, 81, 2414-2430. https://doi.org/10.1016/j.matcom.2011.03.006

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin, 27, 117-137. https://doi.org/10.2143/AST.27.1.563210

Nadarajah, S. (2008). On the product of generalized Pareto random variables. Applied Economics Letters, 15(4), 253-259. https://doi.org/10.1080/13504850500425378

Obeid, N., & Kadry, S. (2022). Distributions of the product and ratio of two independent Pareto and exponential random variables. Journal of Statistics Applications & Probability, 11(1), 291-304. https://doi.org/10.18576/jsap/110123

Salahuddin, N., Khalil, A., Mashwani, W. K., Alrajhi, S., Al-Marzouki, S., & Shah, K. (2021). On the properties of the new generalized Pareto distribution and its applications. Mathematical Problems in Engineering, 2021, Article 6855652. https://doi.org/10.1155/2021/6855652

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464. https://doi.org/10.1214/aos/1176344136

Self, S. G., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association, 82, 605-610. https://doi.org/10.1080/01621459.1987.10478472

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika, 72, 67-90. https://doi.org/10.1093/biomet/72.1.67

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association, 69, 730-737. https://doi.org/10.1080/01621459.1974.10480196

Tahir, M. H., & Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established G families. Anais da Academia Brasileira de Ciências, 87, 539-568. https://doi.org/10.1590/0001-3765201520140299

Tahir, M. H., Cordeiro, G. M., Alzaatreh, A., Mansoor, M., & Zubair, M. (2016). A new Weibull-Pareto distribution: Properties and applications. Communications in Statistics - Simulation and Computation, 45, 3548-3567. https://doi.org/10.1080/03610918.2014.948190