Scandinavian Actuarial Journal

ISSN: 0346-1238 (Print) 1651-2030 (Online) Journal homepage: https://www.tandfonline.com/loi/sact20

Odd Pareto families of distributions for modeling loss payment data

Nonhle Channon Mdziniso & Kahadawala Cooray

To cite this article: Nonhle Channon Mdziniso & Kahadawala Cooray (2018) Odd Pareto families of distributions for modeling loss payment data, Scandinavian Actuarial Journal, 2018:1, 42-63, DOI: 10.1080/03461238.2017.1280527

To link to this article: https://doi.org/10.1080/03461238.2017.1280527

Published online: 25 Jan 2017.

SCANDINAVIAN ACTUARIAL JOURNAL, 2018
VOL. 2018, NO. 1, 42–63
https://doi.org/10.1080/03461238.2017.1280527

Odd Pareto families of distributions for modeling loss payment data

Nonhle Channon Mdziniso and Kahadawala Cooray

Department of Mathematics, Central Michigan University, Mt. Pleasant, MI, USA

ABSTRACT
A three-parameter generalization of the Pareto distribution is presented with density function having a flexible upper tail in modeling loss payment data. This generalized Pareto distribution will be referred to as the *Odd Pareto* distribution since it is derived by considering the distributions of the odds of the Pareto and inverse Pareto distributions. Basic properties of the Odd Pareto distribution (OP) are studied. Model parameters are estimated using both modified and regular maximum likelihood methods. Simulation studies are conducted to compare the OP with the exponentiated Pareto, Burr, and Kumaraswamy distributions using two different test statistics based on the ml method. Furthermore, two examples from the Norwegian fire insurance claims data-set are provided to illustrate the upper tail flexibility of the distribution. Extensions of the Odd Pareto distribution are also considered to improve the fitting of data.

ARTICLE HISTORY
Received 21 September 2016
Accepted 5 January 2017

KEYWORDS
Coverage probabilities; hazard function; moments; Pareto distribution

1. Introduction

In insurance industry, loss payment data are typically modeled using heavy-tailed distributions. In this case, the Pareto distribution is the prominent candidate though it has some limitations in its upper tail for practical applications. As a result, many other generalizations/extensions of the Pareto distribution have been developed to improve the fitting of data, see Arnold (2015) and Johnson et al. (1994). The generalized Pareto distribution (GPD) was first introduced by Pickands (1975) and has applications in the analysis of extreme events and modeling large insurance claims (Hosking & Wallis 1987). For example, the Norwegian fire insurance claims data (Beirlant et al. 1996) which has been analyzed extensively in literature.

The Norwegian fire insurance claims data consist of a record for Norwegian fire claims for the years 1972–1984 and 1986–1992 which exceed 500 thousand Norwegian krones, rounded to thousand Norwegian krones. Brazauskas & Kleefeld (2011) fitted the log-folded-normal and log-folded-t distributions to the logged 1988 Norwegian fire claims data (after taking the logarithm of the data divided by 500). Their results showed that the log-folded-t_7 distribution (LFT7D) is appropriate for the Norwegian fire claims data. However, Scollnik (2014) argued that the LFT7D is not necessarily the best fit to this data. Because of the lower limit of 500, Scollnik (2014) showed that truncated versions of simpler standard distributions such as the GPD and lognormal-Pareto distribution (LNPD) provide a better for this data, with the LNPD providing the best fit amongst the distributions compared in his paper. In addition, Scollnik (2014) suggested that the competing distributions in Brazauskas & Kleefeld (2011) should be applied to the original claims data instead of the transformed data. Brazauskas & Kleefeld (2014) agreed with Scollnik (2014) on modeling the

CONTACT Kahadawala Cooray coora1k@cmich.edu

original data-set since using the log-transformed data gave an advantage to the LFT7D over the rest of the competing distributions. However, after fitting the data using truncated versions of the distributions in Scollnik (2014), Brazauskas & Kleefeld (2014) showed that the LFT7D outperforms the LNPD in fitting this data. Furthermore, Brazauskas & Kleefeld (2016) fitted truncated versions of the GPD and several alternative models to the Norwegian fire insurance claims data for the period 1981–1992.

Motivated by Brazauskas & Kleefeld (2011), Nadarajah & Bakar (2015) fitted the folded generalized t distribution, folded Gumbel distribution, and folded exponential power distribution to the logged 1988 Norwegian fire insurance claims data using the ml method. Their results showed that the folded generalized t and folded exponential power distributions provide the best fits to the data. Beirlant et al. (2001) also fitted the generalized Burr-gamma distribution to the 1990 Norwegian fire insurance claims data. In light of the work of Scollnik (2014), analysis of the Norwegian fire insurance claims data for the years 1988 and 1990 is performed in this paper without transforming the original data.

Recent generalizations of the Pareto distribution include exponentiated Pareto distributions introduced by Nadarajah (2005) for the random variable $Y = \log X$ where the random variable X is taken from five commonly known Pareto distributions, namely, the Pareto distribution, Lomax distribution, Pareto III distribution, GPD, and the truncated version of the Pareto distribution. Another work introduced by Akinsete et al. (2008) is the beta Pareto distribution which has applications in modeling exceedances of flood peaks of the Wheaton River in Canada, and the Floyd River in Iowa, USA. Beirlant et al. (2005) introduced an extension of the GPD which has applications in modeling daily log-returns of the euro-UK pound exchange rate. Aljarrah et al. (2015) also introduced a new Weibull Pareto distribution for modeling remission times of bladder cancer patients and breaking stress of carbon fibers.

In the same line of research, we will study a Pareto extension under the idea developed in the Odd Weibull distribution (Cooray 2006). To extend the Odd family of distributions, this paper will introduce, develop, and study a new three-parameter generalization of the Pareto distribution having flexible upper tail in modeling loss payment data. This GPD given in Section 2 will be referred to as the Odd Pareto distribution (OP) since it is derived by considering the distributions of the odds of the Pareto $(F(x) = 1 - (\phi/x)^{\tau};\ x > \phi > 0,\ 0 < \tau)$ and inverse Pareto $(F(x) = (x/\delta)^{\kappa};\ 0 < x < \delta,\ 0 < \kappa)$ families. The importance of this Pareto extension is that it brought both the Pareto and inverse Pareto into a single distribution.

The rest of the article is as follows. In Section 3, we provide properties of the OP distribution. Unlike the Pareto distribution, parameters of the OP distribution are estimated under both regular and modified maximum likelihood (ml) methods in Section 4.1. A simulation study is conducted to assess the performance of the ml estimators of the OP parameters in Section 4.2. Furthermore, the OP distribution over various tail-thicknesses is compared with other Pareto-type distributions by simulation in Section 4.3. In Section 5, we fit the new distribution to the Norwegian fire insurance claims data to illustrate its upper tail flexibility. In addition, possible extensions of the OP distribution are considered when modeling the Norwegian fire insurance claims data for the purpose of obtaining a better fit. Finally, Section 6 concludes the article.

2. The OP distribution

In this section, we define the OP distribution. The OP distribution is derived based on the following two questions found in loss modeling:

(1) What are the odds that an individual loss will not exceed a certain threshold x, if the loss random variable X follows the Pareto distribution?
(2) If these odds follow some other loss distribution L, then what is the corrected distribution of X?

Obviously, the answer to the first question is very straightforward and depends on the Pareto distribution. However, the answer to the second question will vary due to the choice of both L and the Pareto distribution. Let us answer the first question by representing odds that an individual loss will not exceed x in terms of the Pareto distribution function $F_X(x)$ as

$$F_X(x)/(1 - F_X(x)) = (x/\theta)^\lambda - 1$$

where $F_X(x) = \Pr(X \le x) = 1 - (\theta/x)^\lambda; x \ge \theta, 0 < \lambda, 0 < \theta$. Here one can denote this ratio, the odds of loss, by y $(0 < y)$, and it can be considered as a random variable. Suppose that we are interested in modeling the randomness of the 'odds' using an appropriate parametric distribution, say, $F_Y(y)$. Then, we can write

$$\Pr(Y \le y) = F_Y(y) = F_Y[(x/\theta)^\lambda - 1].$$

Let us consider L as the loglogistic distribution to model this randomness with its cdf given by $F_Y(y) = 1 - (1 + y^\gamma)^{-1}; 0 < \gamma < \infty$. Perhaps, it is a desirable candidate to model this randomness, since the analog power transformation exists as follows

$$F_X(x)/(1 - F_X(x)) = [F_Y(y)/(1 - F_Y(y))]^{1/\gamma}.$$

Hence, γ can be considered as a correction parameter of the Pareto distribution. Furthermore, the parameter γ is the log odd ratio between the OP and Pareto distributions given by the following formula.

$$\gamma = \log[F_Y(y)/(1 - F_Y(y))]/\log[F_X(x)/(1 - F_X(x))].$$

Then the cdf of the corrected distribution of X is

$$F_X(x) = 1 - 1/\{1 + [(x/\theta)^\lambda - 1]^\gamma\}; \quad x > \theta, 0 < \theta, 0 < \lambda, 0 < \gamma. \tag{1}$$

If the random variable X follows the inverse Pareto distribution, or commonly called the Power distribution, with its distribution function, $F_X(x) = (\theta/x)^\lambda; 0 \le x \le \theta, 0 < \theta, \lambda < 0$. Then we can write the corrected distribution of X as

$$F_X(x) = 1 - 1/\{1 + [(x/\theta)^\lambda - 1]^{-\gamma}\}; \quad 0 < x < \theta, 0 < \gamma, \lambda < 0, 0 < \theta. \tag{2}$$

One can easily combine (1) and (2) by writing the correction parameter $\beta = \pm\gamma, 0 < \gamma$, to obtain the cdf, pdf, and quantile function of the OP distribution as

$$F(x; \lambda, \beta, \theta) = 1 - \left\{1 + \left[\left(x/\theta\right)^\lambda - 1\right]^\beta\right\}^{-1}, \tag{3}$$

$$f(x; \lambda, \beta, \theta) = \left(\lambda\beta/x\right)\left(x/\theta\right)^\lambda \left[\left(x/\theta\right)^\lambda - 1\right]^{\beta-1} \left\{1 + \left[\left(x/\theta\right)^\lambda - 1\right]^\beta\right\}^{-2}, \tag{4}$$

$$Q(u) = \theta\left\{1 + \left[u/(1-u)\right]^{1/\beta}\right\}^{1/\lambda}; 0 < u < 1, 0 < \lambda\beta, 0 < \theta. \tag{5}$$

where $0 < x < \theta$ if $\lambda < 0$ and $\beta < 0$, $\theta < x < \infty$ if $0 < \lambda$ and $0 < \beta$; $0 < \theta$. In addition, the parameter β is the log odd ratio between the OP and the Pareto (or the inverse Pareto) distributions.

Note that when $\beta = -1$ and $\beta = 1$ this density represents the inverse Pareto (or Power) and the Pareto distributions, respectively. When $\theta = 1$, $\lambda < 0$, and $\beta < 0$, the shapes of the arising densities

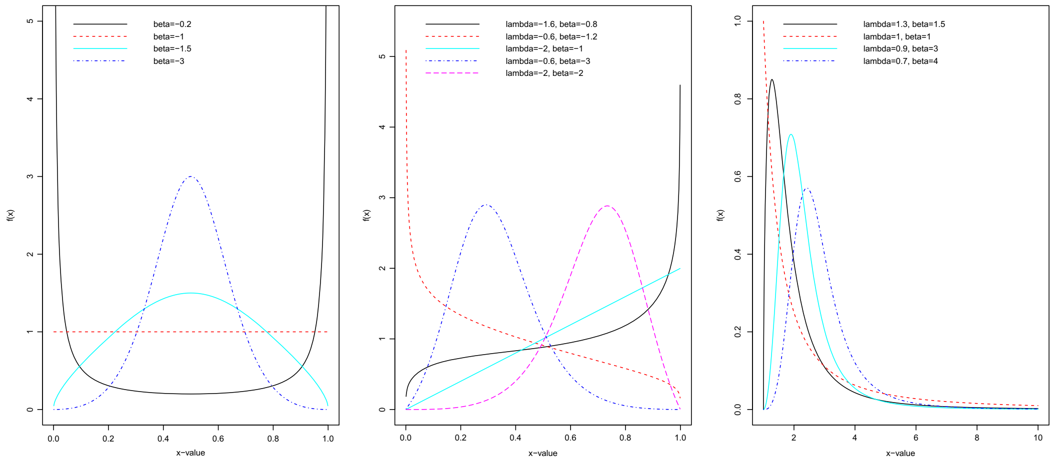

Figure 1. Typical OP densities: symmetric(left), asymmetric(middle), and positively skewed(right) densities.

are similar to the beta densities. It is clear that $1/x$ transformation of the OP distribution does not change its density form.

3. Some properties of the OP distribution

3.1. Density shapes

(1) If both shape parameters of the distribution are positive, $0 < \lambda$ and $0 < \beta$, then the OP densities are positively skewed with thick upper tails.
(2) If $\lambda = -1$ and $\beta < 0$, then the OP densities are symmetric.
(3) If $\lambda < -1$ and $\beta < 0$, then the OP densities are negatively skewed.
(4) If $-1 < \lambda < 0$ and $\beta < 0$, then the OP densities are positively skewed.
1. When $0 < \lambda\beta \leq 1$ and $-1 < \beta < 0$, the OP densities are bathtub shaped.

Figure 1 shows symmetric OP density curves for $\lambda = -1, \beta < 0, \theta = 1$; asymmetric OP density curves for $\lambda < 0, \beta < 0, \theta = 1$; and positively skewed OP density curves for $0 < \lambda, 0 < \beta$, and $\theta = 1$, respectively.

Furthermore, the following Pareto and inverse Pareto bounds can be constructed to identify the structure of the OP distribution.

Theorem 3.1: *The bounds for the OP distribution function $F(x)$ in (3) are given by*

$$\begin{aligned} F_{ip}^{-\beta}(x) \leq F(x) \leq -\beta F_{ip}(x) \quad & \text{if } -\infty < \beta \leq -1; \\ -\beta F_{ip}(x) \leq F(x) \leq F_{ip}^{-\beta}(x) \quad & \text{if } -1 \leq \beta < 0; \\ \beta F_p(x) \leq F(x) \leq F_p^{\beta}(x) \quad & \text{if } 0 < \beta \leq 1; \\ F_p^{\beta}(x) \leq F(x) \leq \beta F_p(x) \quad & \text{if } 1 \leq \beta < \infty; \end{aligned} \tag{6}$$

where $F_p(x)$, $F_{ip}(x)$, and $F(x)$ stand for the cdf of Pareto, inverse Pareto, and OP distributions, respectively.

Proof: Proofs for these results can be found in the Appendices 1–6. □

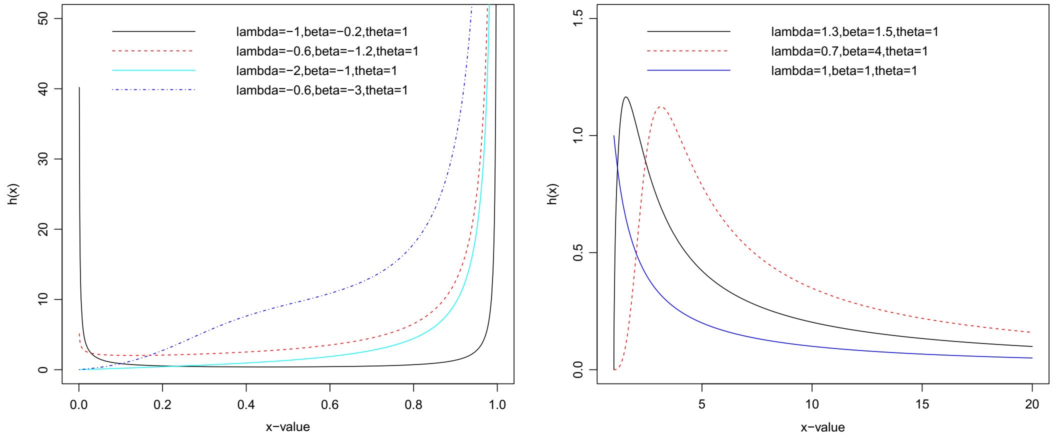

Figure 2. Typical hazard shapes: negative parameters(left) and positive parameters(right).

3.2. Hazard shapes

For the random variable X distributed as the OP distribution, the hazard function is

$$h(x) = (\lambda\beta/x)\,(x/\theta)^{\lambda}\left[(x/\theta)^{\lambda} - 1\right]^{\beta-1}\left\{1 + \left[(x/\theta)^{\lambda} - 1\right]^{\beta}\right\}^{-1},$$

where $0 < x < \theta$ if $\lambda < 0$ and $\beta < 0$, $\theta < x < \infty$ if $0 < \lambda$ and $0 < \beta$; $0 < \theta$.

Figure 2 shows graphs of the hazard function for negative parameters(left), and positive parameters(right).The hazard function is bathtub shaped for $\lambda < 0$ and $\beta < 0$ with $\lambda\beta < 1$; it is strictly increasing for $\lambda < 0$ and $\beta < 0$ with $1 < \lambda\beta$; and it is always unimodal for all $0 < \lambda$ and $0 < \beta$.

3.3. Related distributions

(1) When $\lambda = -1$, the OP distribution reduces to a submodel of the following distribution by Johnson et al. (1994).

$$F(x) = \left[1 + e^{-\gamma}\,(1/y - 1)^{-\delta}\right]^{-1}; \quad 0 \le y \le 1, 0 < \gamma, \delta < 0.$$

(2) Gilchrist (2000) presented the following quantile function by multiplying the quantile functions of Pareto and inverse Pareto distributions.

$$Q(u) = \theta u^{\gamma}/(1-u)^{\delta}; \quad 0 \le u \le 1, 0 < \gamma, \delta, \theta < \infty.$$

Note that even though the Pareto and the inverse Pareto distributions are submodels of this distribution like the OP distribution, it does not provide closed form expressions for either density or distribution functions.

(3) The OP distribution can be obtained by logarithmic transformation to a submodel of the following distribution given by Cooray (2006).

$$F(x; \lambda, \beta, \theta) = 1 - \left\{1 + \left[e^{(x/\theta)^{\lambda}} - 1\right]^{\beta}\right\}^{-1}; 0 < x < \infty, 0 < \theta, 0 < \lambda\beta.$$

This distribution is named as the Odd Weibull distribution, since it is derived by considering the distribution of the 'odds of death' of a Weibull or an inverse Weibull random variable.

3.4. Moments

The kth positive raw moments of the OP distribution are given by the following.

If $0 < \lambda$ and $0 < \beta$, let $u = (\theta/x)^{\lambda}$. Then

$$E[x^k] = \theta^k + (k/\lambda)\theta^k \int_0^1 \sum_{m=1}^{\infty} (-1)^m u^{-\beta m-(k/\lambda)-1}(1-u)^{\beta m}\, du, \tag{7}$$

provided and $|1-u| < |u|$; where $0 < \theta$.

If $\lambda < 0$ and $\beta < 0$, let $\lambda = -\tau$, $\beta = -\delta$, and $v = (x/\theta)^{\tau}$, where $0 < \tau$ and $0 < \delta$. Then

$$E[x^k] = (k/\lambda)\theta^k \sum_{m=1}^{\infty} (-1)^m B(\delta m + (k/\tau), -\delta m + 1), \tag{8}$$

provided $|1-v| < |v|$ and $\delta m < 1$, where $0 < \theta$ and $B(x, y)$ is the Beta function

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt; \quad 0 < x, 0 < y.$$

The formulas in (7) and (8) do not have closed form solutions and have to be evaluated numerically. However, the following theorem provides the condition for the existence of the moments of the OP distribution.

Theorem 3.2: *For $k \in \mathbb{R}^+$, the kth raw moments of the OP distribution exist finitely if $k < \lambda\beta$.*

Proof: The proof can be found in the Appendices 1–6. □

It should be noted that this restriction on the existence of positive raw moments, $k < \lambda\beta$, indicates that the OP distribution has a heavy right tail. Furthermore, bounds for the kth positive moments are given by the following theorem.

Theorem 3.3: *The kth positive moments of the OP distribution have the following finite bounds.*

$$\lambda/(\lambda - k) \le E\left[(x/\theta)^k\right] \le B\left(1 - k/\lambda, \beta\right),\ \textit{provided } k < \lambda;\ 0 < \lambda, 0 < \beta \le 1.$$
$$B\left(1 - k/\lambda, \beta\right) \le E\left[(x/\theta)^k\right] \le \lambda/(\lambda - k),\ \textit{provided } k < \lambda;\ 0 < \lambda, 1 < \beta < \infty.$$
$$1/(k - \lambda) \le E\left[(x/\theta)^k\right] \le 1/(k + \lambda\beta);\ \lambda < 0, -1 \le \beta < 0.$$
$$1/(k + \lambda\beta) \le E\left[(x/\theta)^k\right] \le 1/(k - \lambda);\ \lambda < 0, -\infty < \beta < -1.$$

where $B(x, y)$ is the Beta function

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt; \quad 0 < x, 0 < y.$$

Proof: The proof can be found in the Appendices 1–6. □

4. Parametric inference

In this Section we discuss parametric inference as follows. Parameter estimation is discussed in Section 4.1. In Section 4.2, a simulation study is conducted to assess the performance of the ml estimators of the OP parameters. Section 4.3 discusses a comparison by simulation of the OP distribution with other Pareto-type distributions.

4.1. Parameter estimation

This subsection presents results on ml estimation of the parameters λ, β, and θ. Modified ml method of the OP parameters is used to estimate ml estimators in Section 4.1.1, whereas regular ml method to estimate the OP parameters are given in Section 4.1.2.

4.1.1. Derivation using modified ml method

Let $U_y(\theta) = \left(x_i/\theta\right)^{\lambda} - 1$. The loglikelihood function is given by

$$\begin{aligned} l(x_1,\dots,x_n;\theta,\lambda,\beta) &= \log L(x_1,\dots,x_n;\theta,\lambda,\beta) \\ &= \log\left\{\left(\lambda^n\beta^n \Big/ \prod_{i=1}^{n} x_i\right)\prod_{i=1}^{n}\left(x_i/\theta\right)^{\lambda}\prod_{i=1}^{n}\left(U_y(\theta)\right)^{\beta-1} \Big/ \prod_{i=1}^{n}\left[1+\left(U_y(\theta)\right)^{\beta}\right]^2\right\} \end{aligned} \tag{9}$$

Denote $l(\underline{\theta}) = l(x_1,\dots,x_n;\lambda,\beta,\theta)$. Finding the derivatives with respect to λ and β of the loglikelihood function in Equation (9) gives the following.

$$\partial l(\underline{\theta})/\partial\lambda = T_{21}(\underline{\theta}) + T_{22}(\underline{\theta}), \text{ where} \tag{10}$$

$$T_{21}(\underline{\theta}) = n/\lambda + \sum_{i=1}^{n}\log\left(x_i/\theta\right) + (\beta-1)\sum_{i=1}^{n}\left(x_i/\theta\right)^{\lambda}\log\left(x_i/\theta\right) \Big/ U_y(\theta), \text{ and}$$

$$T_{22}(\underline{\theta}) = -2\sum_{i=1}^{n}\beta\left(x_i/\theta\right)^{\lambda}\left(U_y(\theta)\right)^{\beta-1}\log\left(x_i/\theta\right) \Big/ \left[1+\left(U_y(\theta)\right)^{\beta}\right];$$

$$\partial l(\underline{\theta})/\partial\beta = n/\beta + \sum_{i=1}^{n}\log\left(U_y(\theta)\right) - 2\sum_{i=1}^{n}\left(U_y(\theta)\right)^{\beta}\log\left(U_y(\theta)\right) \Big/ \left[1+\left(U_y(\theta)\right)^{\beta}\right] \tag{11}$$

We obtain the ml estimators for λ and β by solving the system of equations $\left[\partial/\partial\lambda, \partial l/\partial\beta\right]' = \underline{0}$. Solving this system directly is not possible. However, the following theorem provides the existence of the ml estimators.

Theorem 4.1: *The ml estimators for λ and β based on the modified ml method exist for a given parameter θ.*

Proof: The proof can be found in the Appendices 1–6. □

4.1.2. Derivation using regular ml method

The partial derivative of the loglikelihood function in Equation (9) with respect to θ is

$$\partial l(\underline{\theta})/\partial\theta = (1/\theta)\sum_{i=1}^{n}\left[-\lambda + \lambda(1-\beta)(x_i/\theta)^{\lambda}/U_y(\theta) + T_{71}(\underline{\theta})\right], \text{ where} \tag{12}$$

$$T_{71}(\underline{\theta}) = 2\lambda\beta(x_i/\theta)^{\lambda}(U_y(\theta))^{(\beta-1)} \Big/ \left[1+(U_y(\theta))^{\beta}\right].$$

The following results were obtained on the ml estimators for λ, β, and θ using the regular ml method.

Lemma 4.1: *The solution to $\partial l(\underline{\theta})/\partial\theta = 0$ always exists when $1 < |\beta|$.*

Proof: The proof can be found in the Appendices 1–6. □

The following theorem shows existence of the ml estimators for λ, β, and θ when $\partial l(\underline{\theta})/\partial\theta = 0$ has a solution.

Theorem 4.2: *The ml estimators for λ, β, and θ based on the regular ml method exist.*

Proof: The proof can be found in the Appendices 1–6. □

Table 1. Approximate coverage probabilities under modified ml method based on 10,000 simulations.

		90% intended			95% intended		
(λ, β)	$n \rightarrow$	25	50	100	25	50	100
$(0.5, 0.5)$	λ :	0.924	0.915	0.907	0.958	0.961	0.950
	β :	0.899	0.905	0.901	0.952	0.951	0.955
$(0.5, 1.5)$	λ :	0.901	0.897	0.901	0.948	0.946	0.949
	β :	0.903	0.898	0.903	0.952	0.953	0.953
$(1.5, 0.5)$	λ :	0.935	0.915	0.898	0.969	0.957	0.948
	β :	0.909	0.913	0.901	0.953	0.956	0.949
$(0.5, 5.0)$	λ :	0.883	0.890	0.894	0.931	0.940	0.943
	β :	0.900	0.896	0.902	0.953	0.952	0.953
$(5.0, 0.5)$	λ :	0.928	0.919	0.903	0.967	0.958	0.954
	β :	0.913	0.910	0.907	0.958	0.959	0.955
$(-2.0, -1.0)$	λ :	0.920	0.905	0.894	0.962	0.951	0.945
	β :	0.898	0.897	0.897	0.951	0.952	0.948
$(-1.6, -0.8)$	λ :	0.932	0.929	0.899	0.969	0.965	0.948
	β :	0.908	0.905	0.899	0.953	0.956	0.949
$(-0.9, -0.5)$	λ :	0.929	0.921	0.906	0.966	0.965	0.951
	β :	0.910	0.901	0.902	0.954	0.952	0.954
$(-0.6, -3.0)$	λ :	0.881	0.897	0.900	0.934	0.947	0.949
	β :	0.900	0.905	0.903	0.951	0.952	0.949
$(-0.5, -0.5)$	λ :	0.932	0.915	0.904	0.964	0.958	0.948
	β :	0.903	0.902	0.902	0.953	0.951	0.953

4.2. Coverage probabilities

The approximate coverage probabilities for the modified ml method for complete data with intended confidence levels $\alpha = 0.1$ and $\alpha = 0.05$ are given in Table 1. These coverage probabilities are based on 10000 simulated random samples from the density given in (4). The random samples are generated by plugging the known values of parameters λ, β, and θ (say $\lambda = 0.5, \beta = 0.5, \theta = 1$) to the quantile function given in (5). In addition, n (say $n = 25$) number of ordered uniform random sample from the uniform distribution, $u \sim U(0, 1)$ is required to substitute as u in (5). In that way, one random sample with size n (say $n = 25$) from the odd Pareto distribution with parameters λ, β, and θ (say $\lambda = 0.5$, $\beta = 0.5, \theta = 1$) can be generated. In this simulation study, ten thousand such samples are generated to get a single cell value in Table 1. The approximate $100\left(1 - \alpha\right)\%$ confidence intervals for parameters, λ, β, and θ are calculated by using $\left(\widehat{\lambda} - Z_{\alpha/2}SE_{\widehat{\lambda}}, \widehat{\lambda} + Z_{\alpha/2}SE_{\widehat{\lambda}}\right)$, $\left(\widehat{\beta} - Z_{\alpha/2}SE_{\widehat{\beta}}, \widehat{\beta} + Z_{\alpha/2}SE_{\widehat{\beta}}\right)$, and $\left(\widehat{\theta} - Z_{\alpha/2}SE_{\widehat{\theta}}, \widehat{\theta} + Z_{\alpha/2}SE_{\widehat{\theta}}\right)$, respectively, where $SE_{\widehat{\lambda}}$, $SE_{\widehat{\beta}}$, and $SE_{\widehat{\theta}}$ are, respectively, asymptotic standard errors of $\widehat{\lambda}$, $\widehat{\beta}$, and $\widehat{\theta}$, which are taken from the observed information matrix.

From Table 1, one can clearly see that when the sample size increases, the approximate coverage probabilities for the parameters under the ml method are getting closer to the intended coverage probabilities. When scale parameter, θ, increases from 1 to 2, we cannot see significant changes of the coverage probabilities. Furthermore, the coverage probabilities of parameter β and θ are lower than the intended level for small samples. But, one can obtain a desired confidence level by appropriately adjusting the confidence coefficient α. Moreover, the procedure gives an extremely over coverage for parameter λ for too small samples. For example, if we keep the intended level 90%, the coverage probabilities for λ when $n = 10$, $\lambda = 1$, $\beta = 1$, and $\theta = 1$, is approximately 99%. Therefore, in the construction of Table 1, we kept the minimal sample size as 25 to make sense for the coverage probability values. Hence, the values in Table 1 predict that the estimated parameter values under the ml method are not substantially biased either to overestimate or underestimate. Note that in Table 1, the shape parameter values, λ and β, are chosen to represent different hazard and density shapes. Similarly, the approximate coverage probabilities using the regular ml method for complete data with intended confidence levels $\alpha = 0.1$ and $\alpha = 0.05$, based on a small simulation study are

Table 2. Approximate coverage probabilities under regular ml method based on 1000 simulations.

		90% intended			95% intended		
(λ, β, θ)	$n \rightarrow$	25	50	100	25	50	100
$(1.0, 2.0, 1.0)$	$\lambda:$	.974	0.932	0.911	0.993	0.979	0.937
	$\beta:$	0.974	0.911	0.902	0.995	0.969	0.931
	$\theta:$	0.860	0.803	0.825	0.897	0.844	0.873
$(0.5, 5.0, 1.0)$	$\lambda:$	0.952	0.920	0.884	0.964	0.951	0.937
	$\beta:$	0.846	0.751	0.751	0.885	0.810	0.803
	$\theta:$	0.901	0.827	0.812	0.926	0.871	0.856
$(2.0, 5.0, 1.0)$	$\lambda:$	0.919	0.878	0.889	0.958	0.916	0.938
	$\beta:$	0.763	0.743	0.765	0.803	0.807	0.811
	$\theta:$	0.768	0.739	0.779	0.809	0.778	0.824
$(1.5, 3.0, 1.0)$	$\lambda:$	0.924	0.864	0.841	0.963	0.920	0.879
	$\beta:$	0.851	0.794	0.810	0.881	0.866	0.849
	$\theta:$	0.826	0.775	0.773	0.868	0.807	0.821
$(1.0, 3.0, 1.0)$	$\lambda:$	0.975	0.941	0.900	0.991	0.965	0.948
	$\beta:$	0.912	0.874	0.870	0.954	0.912	0.917
	$\theta:$	0.900	0.871	0.842	0.928	0.903	0.884
$(-1.5, -2.0, 1.0)$	$\lambda:$	0.963	0.951	0.825	0.998	0.979	0.906
	$\beta:$	0.915	0.847	0.786	0.975	0.899	0.861
	$\theta:$	0.741	0.727	0.603	0.778	0.793	0.671
$(-0.25, -3.0, 1.0)$	$\lambda:$	0.939	0.890	0.892	0.982	0.930	0.923
	$\beta:$	0.862	0.850	0.834	0.903	0.905	0.882
	$\theta:$	0.643	0.734	0.772	0.678	0.759	0.803
$(-1.0, -5.0, 1.0)$	$\lambda:$	0.806	0.868	0.855	0.856	0.900	0.909
	$\beta:$	0.624	0.736	0.757	0.671	0.773	0.800
	$\theta:$	0.545	0.698	0.739	0.603	0.742	0.767
$(-2.0, -5.0, 1.0)$	$\lambda:$	0.941	0.915	0.894	0.965	0.952	0.934
	$\beta:$	0.816	0.778	0.807	0.841	0.829	0.847
	$\theta:$	0.813	0.777	0.751	0.839	0.820	0.791

given in Table 2. Results show that coverage probabilities do not approach intended probabilities when $|\beta| \rightarrow 1$, a consequence of Lemma 4.1.

4.3. A comparison of the OP with Kum, EP, and Burr distributions

As mentioned in the introduction, the OP distribution has a flexible upper tail in modeling loss payment data. To see whether this upper tail flexibility is practically significant, it is a common practice to compare with existing competitive distributions possibly through a simulation study than analyze a huge number of real data-sets. For this purpose, we consider three well-known three-parameter distributions: the exponentiated Pareto distribution (EP) ($F(x;\gamma,\tau,\delta) = \left[1 - \left(\delta/x\right)^{\gamma}\right]^{\tau}$; $x \geq \delta,\ 0 < \delta, \gamma, \tau$) (Gupta et al. 1998), the Kumaraswamy distribution (Kum) ($F(x;\lambda,\beta,\theta) = 1 - \left[1 - \left(x/\theta\right)^{\lambda}\right]^{\beta}$; $0 < x < \theta,\ 0 < \lambda, \beta, \theta$) (Kumaraswamy 1980), and the Burr distribution ($F(x;\lambda,\beta,\theta) = 1 - \left\{1\Big/\left[1 + (x/\theta)^{\lambda}\right]\right\}^{\beta}; x > 0, 0 < \lambda, \beta, \theta$) (Johnson et al. 1994). Using the modified ml method, we compare the OP distribution with positive parameters and the EP distribution; the OP distribution with negative parameters and the Kum distribution. We compare the OP distribution with the Burr distribution using the regular ml method. Comparisons are done through a simulation study by considering the logarithm of the RML (ratio between ml's) test statistic, T_1 (Gupta & Kundu 2003), and an alternative test statistic, T_2 (Cooray 2015). The comparison criteria based on T_1 and T_2 are defined as follows.

Comparison criterion based on T_1 test statistic

If the data are coming from the OP distribution with positive parameters, we choose the OP distribution as a correct model if $T_1 > 0$, otherwise, we choose the EP distribution, where $T_1 =$

$l_{\mathrm{OP}}(\widehat{\boldsymbol{\theta}}) - l_{\mathrm{EP}}(\widehat{\boldsymbol{\theta}})$, $l_{\mathrm{OP}}(\widehat{\boldsymbol{\theta}}) = \log L\left(\widehat{\lambda}, \widehat{\beta}, \widehat{\theta}\right)$, $l_{\mathrm{EP}}(\widehat{\boldsymbol{\theta}}) = \log L\left(\widehat{\gamma}, \widehat{\tau}, \widehat{\delta}\right)$, and we denote $\widehat{\lambda}, \widehat{\beta}, \widehat{\theta}, \widehat{\gamma}, \widehat{\tau}$, and $\widehat{\delta}$ as the ml estimators for $\lambda, \beta, \theta, \gamma, \tau$, and δ, respectively. Similarly, if the data are coming from the EP distribution, we choose the EP distribution as a correct model if $T_1 > 0$ such that $T_1 = l_{\mathrm{EP}}(\widehat{\boldsymbol{\theta}}) - l_{\mathrm{OP}}(\widehat{\boldsymbol{\theta}})$. A similar procedure is done to compare the OP distribution having negative parameters with the Kum distribution.

Comparison criterion based on T_2 test statistic

If the data are coming from the OP distribution with positive parameters, we choose the OP distribution as a correct model if $T_2 > 0$, otherwise, we choose the EP distribution, where $T_2 = d'_{\mathrm{EP}} - d_{\mathrm{OP}}$. Similarly, if the data are coming from the EP distribution, we choose the EP distribution as a correct model if $T_2 > 0$ such that $T_2 = d'_{\mathrm{OP}} - d_{\mathrm{EP}}$. Where d_{EP}, d_{OP}, d'_{EP}, and d'_{OP} are Cramér-von Mises type distances defined by $d_{EP} = \sum_{i=1}^{n}\left(F_{EP}(x_i;\boldsymbol{\theta}_0) - F_{EP}(x_i;\widehat{\boldsymbol{\theta}})\right)^2$, $d_{OP} = \sum_{i=1}^{n}\left(F_{OP}(x_i;\boldsymbol{\theta}_0) - F_{OP}(x_i;\widehat{\boldsymbol{\theta}})\right)^2$, $d'_{EP} = \sum_{i=1}^{n}\left(F_{OP}(x_i;\boldsymbol{\theta}_0) - F_{EP}(x_i;\widehat{\boldsymbol{\theta}})\right)^2$ and $d'_{OP} = \sum_{i=1}^{n}\left(F_{EP}(x_i;\boldsymbol{\theta}_0) - F_{OP}(x_i;\widehat{\boldsymbol{\theta}})\right)^2$, respectively.

$F_{EP}(x_i;\boldsymbol{\theta}_0)$, $F_{OP}(x_i;\boldsymbol{\theta}_0)$, $F_{EP}(x_i;\widehat{\boldsymbol{\theta}})$, and $F_{OP}(x_i;\widehat{\boldsymbol{\theta}})$ are, respectively, cdf of the original EP, original OP, estimated OP, and estimated EP distributions. The test statistic T_2 measures the comparative closeness to the known original distribution. Hence, T_2 serves the need in our simulation but cannot be computed if the original distribution is unknown. In such a case, empirical cdf, $F_n\left(x\right)$ can be used to replace the original cdf to compute the T_2 values. A similar procedure is done to compare the OP distribution having negative parameters with the Kum distribution.

We simulate data from OP, EP, and Kum distributions with sample sizes $n \in \{25, 50, 100, 200, 400\}$ to produce different density and hazard shapes such that $f(x) \in$ {unimodal (U), increasing (I), decreasing (D), bathtub (B)} and $h(x) \in$ {increasing (I), decreasing (D), bathtub (B), unimodal (U)}. For each of these setups we compute $P\left(T_1 > 0\right)$ and $P\left(T_2 > 0\right)$ based on 1000 replications and the results obtained using modified ml method are summarized in Table 3. Note that we did not consider any variation of the scale parameter when we simulate data from both distributions. Since, it does not affect the shape of the density or hazard functions. From Table 3, it is clear that the two distributions are different in many cases that we considered based on $P\left(T_1 > 0\right)$ and $P\left(T_2 > 0\right)$ values. Observe that it is not reasonable to compare EP distribution with Kum distribution since EP distribution is bounded below with a thick upper tail while Kum distribution is not.

Using a similar procedure we compute $P\left(T_1 > 0\right)$ and $P\left(T_2 > 0\right)$ based on 1000 replications in order to compare the OP distribution with the Burr distribution using the regular ml method. Results are also summarized in Table 3. From the results, it is clear that the two distributions are different in many cases that we considered based on $P\left(T_1 > 0\right)$ and $P\left(T_2 > 0\right)$ values.

5. Applications

In this Section, the OP distribution is applied to model the Norwegian fire insurance claims data (Beirlant et al. 1996) for the years 1988 and 1990. We use the 1988 and 1990 data-sets since several authors have analyzed these data using different distributions (see Brazauskas & Kleefeld (2011), Brazauskas & Kleefeld (2016), Scollnik (2014), Nadarajah & Bakar (2015), and Beirlant et al. (2001)). In addition, analysis for other years using Odd Pareto families can be obtained from the authors on request.

The 1988 Norwegian fire claims data consists of 827 fire insurance losses in thousand Norwegian krones, ranging from 500 to 465365 thousand Norwegian krones. From preliminary data analysis, the mean, standard deviation, and coefficient of skewness are calculated as 3175.87, 17676.54, and 22.53, respectively. The skewness measure indicates that the data is highly positively skewed.

The 1990 Norwegian fire claims data consist of 628 fire insurance losses in thousand Norwegian krones, ranging from 500 to 78537 thousand Norwegian krones. From preliminary data analysis, the mean, standard deviation, and coefficient of skewness are calculated as 1973.52, 4257.27, and 11.68, respectively. The skewness measure indicates that the data is highly positively skewed.

Table 3. Acceptance probability between OP and EP; OP and Kum; OP and Burr distributions based on 1000 replications.

	$n \rightarrow$	25	50	100	200	400	25	50	100	200	400
Parameters		$P(T_1 > 0)$					$P(T_2 > 0)$				
						Results obtained using modified ml method					
β, θ, λ	Data are coming from the OP distribution compared with EP distribution										
1.0, 1.0, 1.0		0.81	0.72	0.66	0.62	0.53	0.63	0.60	0.52	0.44	0.44
1.5, 1.0, 1.0		0.67	0.60	0.64	0.76	0.79	0.61	0.52	0.52	0.59	0.65
1.5, 1.0, 0.5		0.66	0.58	0.63	0.60	0.79	0.61	0.53	0.52	0.69	0.63
1.2, 1.0, 0.8		0.74	0.61	0.61	0.58	0.66	0.61	0.54	0.48	0.45	0.49
β, θ, λ	Data are coming from the OP distribution compared with Kum distribution										
−0.5, 1.0, −1.0		0.80	0.78	0.83	0.90	0.94	0.87	0.87	0.84	0.87	0.92
−0.8, 1.0, −1.6		0.86	0.78	0.66	0.67	0.72	0.75	0.71	0.71	0.70	0.70
−1.2, 1.0, −0.6		0.74	0.63	0.59	0.58	0.59	0.62	0.53	0.44	0.47	0.49
−1.5, 1.0, −1.0		0.66	0.59	0.65	0.72	0.80	0.60	0.51	0.52	0.58	0.67
τ, δ, γ	Data are coming from the EP distribution compared with OP distribution										
1.0, 1.0, 1.0		0.18	0.29	0.39	0.41	0.46	0.04	0.09	0.14	0.21	0.27
1.5, 1.0, 1.0		0.30	0.50	0.48	0.48	0.54	0.08	0.19	0.46	0.71	0.90
1.5, 1.0, 0.5		0.29	0.48	0.52	0.49	0.47	0.08	0.18	0.40	0.67	0.89
1.2, 1.0, 0.8		0.25	0.39	0.45	0.53	0.54	0.05	0.13	0.27	0.43	0.64
β, θ, λ	Data are coming from the Kum distribution compared with OP distribution										
2.0, 1.0, 2.0		0.38	0.46	0.45	0.48	0.46	0.42	0.46	0.50	0.56	0.55
3.0, 1.0, 1.0		0.40	0.49	0.41	0.36	0.39	0.43	0.49	0.49	0.48	0.46
1.0, 1.0, 5.0		0.19	0.27	0.35	0.42	0.45	0.32	0.43	0.47	0.54	0.58
0.5, 1.0, 0.5		0.26	0.45	0.55	0.69	0.80	0.22	0.38	0.51	0.62	0.75
						Results obtained using regular ml method					
β, θ, λ	Data are coming from the OP distribution compared with Burr distribution										
1.5, 1.0, 0.5		0.89	0.88	0.95	0.97	1.00	0.67	0.69	0.69	0.67	0.76
3.0, 1.0, 0.5		0.60	0.74	0.85	0.88	0.93	0.69	0.71	0.73	0.76	0.81
1.5, 1.0, 1.0		0.84	0.92	0.94	0.98	0.99	0.63	0.68	0.67	0.74	0.83
3.0, 1.0, 1.0		0.84	0.83	0.86	0.91	0.94	0.65	0.72	0.71	0.78	0.81
1.5, 1.0, 2.0		0.97	0.98	0.98	0.99	1.00	0.82	0.89	0.91	0.95	0.95
β, θ, λ	Data are coming from the Burr distribution compared with OP distribution										
0.5, 1.0, 1.5		0.53	0.59	0.69	0.77	0.88	0.34	0.36	0.43	0.42	0.48
1.0, 1.0, 1.5		0.68	0.75	0.87	0.96	1.00	0.46	0.54	0.64	0.73	0.78
2.0, 1.0, 1.5		0.84	0.93	0.99	1.00	1.00	0.60	0.64	0.78	0.85	0.94
1.0, 1.0, 3.0		0.77	0.75	0.84	0.89	0.96	0.40	0.50	0.57	0.68	0.73
2.0, 1.0, 3.0		0.88	0.92	0.96	1.00	1.00	0.60	0.64	0.79	0.85	0.94

Therefore in Section 5.2 we analyze these two data-sets using the OP distribution to see whether there is any significant improvement in the fit of the model by this distribution. However, the OP distribution is not compared to the distributions from Brazauskas & Kleefeld (2011), Brazauskas & Kleefeld (2016), Scollnik (2014), Nadarajah & Bakar (2015), and Beirlant et al. (2001), since our analysis is performed without transforming the data. Furthermore, we analyze these data-sets by considering OP extensions in order to improve fitting of data. For this purpose, we consider the following extensions.

5.1. OP extensions

Exponentiated OP

The distribution function and probability density function of the exponentiated Odd Pareto distribution (EOP) are

$$F(x) = \left(1 - 1\Big/\left\{1 + \left[\left((x/\theta)^{\lambda} - 1\right\}^{\beta}\right]\right)^{\gamma} \tag{13}$$

$$f(x) = (\lambda\beta\gamma/x)\,(x/\theta)^{\lambda}\left[(x/\theta)^{\lambda}-1\right]^{\beta\gamma-1}\Big/\left\{1+\left[(x/\theta)^{\lambda}-1\right]^{\beta}\right\}^{\gamma+1}, \tag{14}$$

where $x \geq \theta$, $0 < \lambda$, $0 < \beta$, $0 < \theta$, and $0 < \gamma$.

The quantile function is given by

$$Q(u) = \theta\left\{1+\left[u^{1/\gamma}\Big/\left(1-u^{1/\gamma}\right)\right]^{1/\beta}\right\}^{1/\lambda}, \quad 0 \leq u \leq 1.$$

Substituting $\gamma = 1$ in Equation (13) gives the cdf of the OP distribution.

Equation (14) is a valid probability density function when the parameters $\lambda < 0$ and $\beta < 0$. (However, we cannot identify the densities.)

Odd generalized Pareto distribution

The distribution function and probability density function of the Odd generalized Pareto distribution (OGP) are

$$F(x) = 1 - 1\Big/\left[1+\left(\left\{1+\lambda\left[(x-\mu)/\sigma\right]\right\}^{1/\lambda}-1\right)^{\beta}\right]$$

$$f(x) = (\beta/\sigma)\left(\left\{1+\lambda\left[(x-\mu)/\sigma\right]\right\}^{1/\lambda}-1\right)^{\beta-1}\left\{1+\lambda\left[(x-\mu)/\sigma\right]\right\}^{1/\lambda-1}\Big/T_{c1}(x),$$

where $T_{c1}(x) = \left[1+\left(\left\{1+\lambda\left[(x-\mu)/\sigma\right]\right\}^{1/\lambda}-1\right)^{\beta}\right]^{2}$, $x \geq \mu$, $0 < \lambda$, $0 < \beta$, $0 < \mu$, and $0 < \sigma$.

The quantile function is given by

$$Q(u) = (\sigma/\lambda)\left(\left\{\left[u/(1-u)\right]^{1/\beta}+1\right\}^{\lambda}-1\right)+\mu, \quad 0 \leq u \leq 1.$$

5.2. Norwegian fire insurance claims data examples

Analysis of the Norwegian fire insurance claims data is performed by comparing the OP distribution and its extensions with the Pareto, 3-parameter GP (Coles 2001), and Burr distributions with parameters estimated using the modified and regular ml methods. The results for each data-set are as follows.

5.2.1. Analysis of 1988 Norwegian fire insurance claims data

In this example, we analyze the complete 1988 Norwegian fire insurance claims data-set. Under the modified ml method we ignore fourteen 500 values at the beginning of the data-set. The fire claims data-set is also analyzed using the regular ml method which accounted for all the data points. This is only possible with the OP distribution but not the other Pareto distributions like the Pareto and 3-parameter GP distributions studied in this paper. Results from the analysis are given in Table 4. The standard errors (S.E.) are given in parenthesis. It is clear from Table 4 that the EOP distribution gives the lowest AIC, BIC, K-S, and A-D values. Thus it is the best distribution for fitting the 1988 Norwegian fire claims data under the modified ml method compared to other distributions investigated in this paper. It is also followed closely by the other OP extension and distribution. Among the OP distribution and its extensions, the OGP distribution has the second smallest values for AIC, BIC, K-S, and A-D. This shows that the two OP extensions improved the fitting of the data significantly. It is also clear from the results that the OP distribution is the best distribution for fitting the data under the regular ml method.

Figure 3 shows the quantile–quantile plots for the OP, 3-parameter GP, OGP, and EOP distributions with estimators obtained using the modified ml method (left); and OP vs. Burr distribution with

Table 4. Norwegian fire insurance claims 1988: Estimated values for fitted distributions.

Distribution	Parameters (S.E.)	-2 LOGLIKE	AIC	BIC	K-S	A-D
Pareto	$\hat{\lambda} = 0.9410$ (0.03300) $\hat{\theta} = 500$ (0.6441584)	13558	13560	13565	0.09739	12.23401
3-parameter -GP	$\hat{\psi} = 0.6673$ (0.05481) $\hat{\sigma} = 766.86$ (46.5581) $\hat{\mu} = 500$ (0.9287401)	13511	13515	13525	0.03829	1.43132
OP	$\hat{\lambda} = 0.8421$ (0.02711) $\hat{\beta} = 1.2707$ (0.04028) $\hat{\theta} = 500$ (2.216186)	13507	13511	13521	0.02821	0.48677
OP*	$\hat{\lambda} = 0.8666$ (0.03264) $\hat{\beta} = 1.2054$ (0.05130) $\hat{\theta} = 495.76$ (2.9910)	**13716**	**13722**	**13736**	**0.03689**	**2.05270**
Burr*	$\hat{\lambda} = 20.3353$ (12.1148) $\hat{\gamma} = 0.04929$ (0.03161) $\hat{\theta} = 525.92$ (26.2910)	13774	13780	13794	0.07034	5.62666
EP	$\hat{\gamma} = 1.1556$ (0.04837) $\hat{\tau} = 1.3994$ (0.06782) $\hat{\delta} = 500$ (2.390424)	13512	13516	13526	0.03578	1.10853
EOP	$\hat{\lambda} = 0.5780$ (0.05607) $\hat{\beta} = 1.7731$ (0.1760) $\hat{\gamma} = 0.5920$ (0.09007) $\hat{\theta} = 500$ (1.343803)	**13500**	**13506**	**13520**	**0.02295**	**0.10740**
OGP	$\hat{\lambda} = 0.8998$ (0.1198) $\hat{\beta} = 1.1558$ (0.05969) $\hat{\sigma} = 693.45$ (51.8910) $\hat{\mu} = 500$ (1.710622)	13503	13509	13523	0.02683	0.26614

Note: Bold values are the best values.

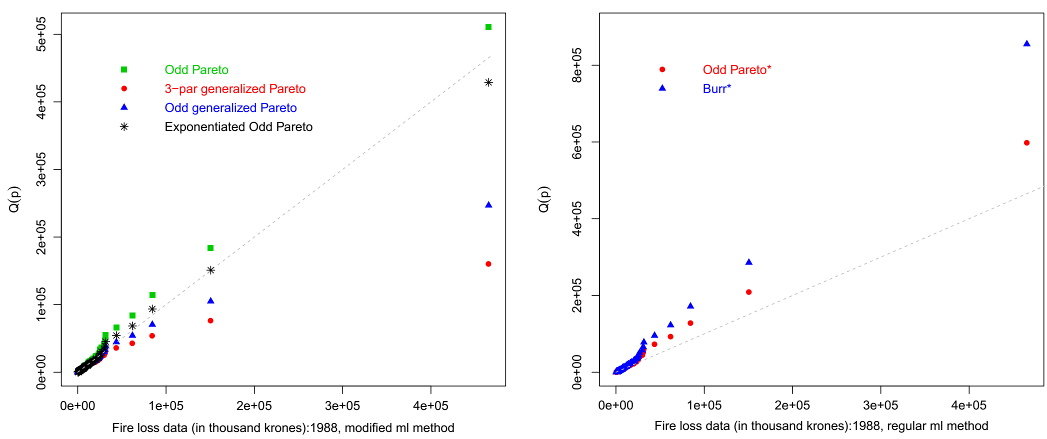

Figure 3. Q-Q plots of 1988 Norwegian fire insurance claims data with estimators obtained using modified ml method (left), regular ml method (right).

estimators obtained using the regular ml method (right). In this figure, the estimated quantiles are plotted against ordered observations. Here the pth quantile $\hat{Q}(p)$ is estimated from the pth quantile of the fitted distribution and $p = (r - 0.5)/n$, $r = 1, \ldots, n$. The Pareto distribution was left out because of its poor performance in fitting the data. The q-q plots show that the OP distribution and its extensions provide a better fit for the data, especially in the upper tail area, compared to

Table 5. Norwegian fire insurance claims 1990: Estimated values for fitted distributions.

Distribution	Parameters (S.E.)	-2 LOGLIKE	AIC	BIC	K-S	A-D
Pareto	$\hat{\lambda} = 1.0405$ (0.04162) $\hat{\theta} = 500$ (0.7675377)	10170	10172	10176	0.13689	19.99709
3-parameter -GP	$\hat{\psi} = 0.4201$ (0.04971) $\hat{\sigma} = 776.12$ (47.7746) $\hat{\mu} = 500$ (1.237515)	10093	10097	10106	0.05216	2.77242
OP	$\hat{\lambda} = 0.8859$ (0.02969) $\hat{\beta} = 1.4216$ (0.05119) $\hat{\theta} = 500$ (3.954075)	10090	10094	10103	0.04790	2.16890
OP*	$\hat{\lambda} = 0.8646$ (0.03653) $\hat{\beta} = 1.4640$ (0.07699) $\hat{\theta} = 490.57$ (7.3209)	**10137**	**10143**	**10156**	**0.04437**	**1.91996**
Burr*	$\hat{\lambda} = 7.8031$ (1.5263) $\hat{\gamma} = 0.1721$ (0.04305) $\hat{\theta} = 637.04$ (34.5818)	10189	10195	10208	0.05818	3.13120
EP	$\hat{\gamma} = 1.4022$ (0.06390) $\hat{\tau} = 1.6525$ (0.09361) $\hat{\delta} = 500$ (4.171434)	10097	10101	10110	0.06443	3.25531
EOP	$\hat{\lambda} = 0.5542$ (0.03918) $\hat{\beta} = 2.3415$ (0.2260) $\hat{\gamma} = 0.4583$ (0.06521) $\hat{\theta} = 500$ (2.022844)	**10071**	**10077**	**10091**	**0.02396**	**0.30838**
OGP	$\hat{\lambda} = 0.6540$ (0.1079) $\hat{\beta} = 1.2155$ (0.06454) $\hat{\sigma} = 717.44$ (49.9820) $\hat{\mu} = 500$ (2.775365)	10078	10084	10098	0.03323	0.85834

Note: Bold values are the best values.

the other distributions under both regular and modified ml methods. Among the OP distribution and its extensions under modified ml method, the EOP is the best fit, followed by the OP, and then OGP distributions, respectively. Unlike other distributions used by Brazauskas & Kleefeld (2011), Brazauskas & Kleefeld (2016), Scollnik (2014), and Nadarajah & Bakar (2015) to fit this data-set, the OP distribution and its extensions provide a good fit for this data without requiring any transformations on the data.

5.2.2. Analysis of 1990 Norwegian fire insurance claims data

In this example, we analyze the complete 1990 Norwegian fire insurance claims data-set. Under the modified ml method we ignore three 500 values at the beginning of the data-set. As done in the first example, the fire claims data-set is also analyzed using the regular ml method which accounted for all the data points, which was only possible with the OP distribution but not the other Pareto distributions like the Pareto and 3-parameter GP distributions studied in this paper. The results from the analysis are given in Table 5. The standard errors (S.E.) are given in parenthesis. Results show that the EOP distribution is the best distribution for fitting the data under the modified ml method since it has the smallest AIC, BIC, K-S, and A-D values. It is followed closely by the OGP distribution, which shows that the two OP extensions improved the fitting of the data significantly.

Figure 4 shows the q-q plots for the OP, 3-parameter GP, OGP, and EOP distributions with estimators obtained using the modified ml method (left); and OP vs. Burr distribution with estimators

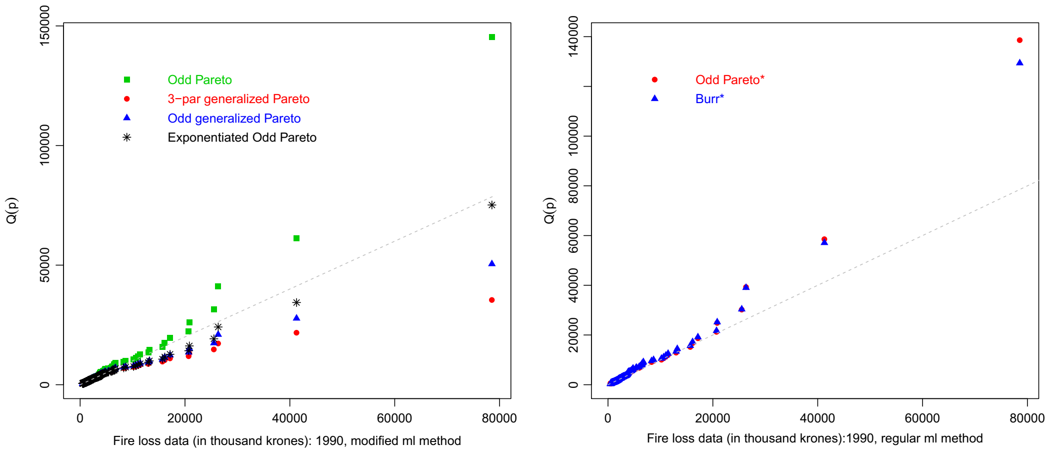

Figure 4. Q-Q plots of 1990 Norwegian fire insurance claims data with estimators obtained using modified ml method (left), regular ml method (right).

obtained using the regular ml method (right). Quantiles for this plot were calculated using the same procedure used in the first example. The Pareto distribution was left out because of its poor performance in fitting the data. The q-q plot shows that the OP distribution and its extensions provide a better fit for the data, especially in the upper tail area, compared to the other distributions investigated in this paper under the modified ml method. Among the OP distribution and its extensions, the EOP is the best fit, followed by the OGP, and then OP distributions, respectively. In comparison with the Burr distribution under the regular ml method, the OP gives a better fit in terms of AIC, BIC, K-S, and A-D values given in the table. Though the last four data points on the q-q plot seem to be a better fit with the Burr distribution.

6. Conclusions

A generalization of the Pareto distribution, the OP distribution for modeling loss payment data has been defined and studied. Various properties of this new distribution are studied. Two extensions of the OP distribution, the EOP and OGP distributions are also presented to improve the fitting of data. All these distributions were fitted to two Norwegian fire claims data-sets and compared to other Pareto-type distributions using the modified and regular ml methods. Based on the distributions that we compared in this paper, the EOP distribution provides the best fit to the data under the modified ml method, showing that the OP extensions improved fitting of data significantly. The OP distribution gave a better fit of the data under the regular ml method, outperforming the Burr distribution which is known to capture highly skewed loss data (Abu Bakar et al. 2015). Unlike other distributions in literature, the OP distribution and its extensions provide a good fit for this data without requiring any transformations on the data. These examples illustrate the usefulness of the OP distribution and its extensions as perfect fits for modeling heavy-tail data.

As a future work, when the OP parameters are negative, one may be able to find useful applications specifically in survival and reliability analysis. In addition, exploring the properties of the EOP and OGP are important tasks.

Acknowledgements

The authors wish to acknowledge an anonymous referee whose constructive remarks improved the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abu Bakar, S. A., Hamzar, N. A., Maghsoudi, M., & Nadarajah, S. (2015). Modeling loss data using composite models. *Insurance: Mathematics and Economics* **61**, 146–154.

Akinsete, A. A., Famoye, F. & Lee, C. (2008). The beta-Pareto distribution. *Statistics* **42**(6), 547–563.

Aljarrah, M. A., Famoye, F. & Lee, C. (2015). A new Weibull-Pareto distribution. *Communications in Statistics – Theory and Methods* **44**(19), 4077–4095.

Arnold, B. C. (2015). *Pareto distributions*. Boca Raton: CRC Press.

Beirlant, J., Joossens, E. & Segers, J. (2005). *Unbiased tail estimation by an extension of the generalized Pareto distribution. (CentER Discussion Paper; Vol. 2005–112)*, Tilburg: Econometrics.

Beirlant, J., Matthys, G. & Dierckx, G. (2001). Heavy-tailed distributions and rating. *Astin Bulletin* **31**(1), 37–58.

Beirlant, J., Teugels, J. L. & Vynckie, P. (1996). *Practical analysis of extreme values*. Leuven: Leuven University Press.

Brazauskas, V. & Kleefeld, A. (2011). Folded- and log-folded-t distributions as models for insurance loss data. *Scandinavian Acturial Journal* **1**, 59–74.

Brazauskas, V. & Kleefeld, A. (2014). Authors' reply to 'Letter to the Editor regarding folded models and the paper by Brazauskas and Kleefeld (2011)'. *Scandinavian Acturial Journal* **8**, 753–757.

Brazauskas, V. & Kleefeld, A. (2016). Modeling severity and measuring tail risk of Norwegian fire claims. *North American Actuarial Journal* **20**(1), 1–16.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*, London: Springer.

Cooray, K. (2006). Generalization of the Weibull distribution: the Odd Weibull family. *Statistical Modeling* **6**, 265–277.

Cooray, K. (2015). A study of moments and likelihood estimators of the Odd Weibull distribution. *Statistical Methodology* **26**, 72–83.

Gilchrist, W. G. (2000). *Statistical modelling with quantile functions*. London: Chapman & Hall.

Gupta, R. C., Gupta, R. D. & Gupta, P. L. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics: Theory and Methods* **27**, 887–904.

Gupta, R. D. & Kundu, D. (2003). Discriminating between Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis* **43**, 179–196.

Hosking, J. R. M. & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339–349.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions*. New York: Wiley.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology* **46**(1–2), 79–88.

Nadarajah, S. (2005). Exponentiated Pareto distributions. *Statistics: A Journal of Theoritical and Applied Statistics* **39**(3), 255–260.

Nadarajah, S. & Bakar, S. A. A. (2015). New folded models for the log-transformed Norwegian fire claim data. *Communications in Statistics - Theory and Methods* **44**(20), 4408–4440.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 119–131.

Scollnik, D. P. (2014). Regarding folded models and the paper by Brazauskas and Kleefeld (2011). *Scandinavian Actuarial Journal* **3**, 278–281.

Appendix 1. Proof for Theorem 3.1

Proof: Consider the following cases.

Case 1 For $0 < \lambda$, and $0 < \beta$, since $F_p(x) = 1 - \left(\theta/x\right)^{\lambda}$, it is straightforward to show that

$$1 - 1\Big/\left\{1 + \left[\left(x/\theta\right)^{\lambda} - 1\right]^{\beta}\right\} = F_p^{\beta}(x)\Big/\left[(1 - F_p(x))^{\beta} + F_p^{\beta}(x)\right] \tag{A1}$$

We have the following subcases.

Case 1a Since $0 \le F_p(x) \le 1$ and $0 \le 1 - F_p(x) \le 1$, then, for $0 < \beta < 1$ we have $1 \le (1 - F_p(x))^{\beta} + F_p^{\beta}(x)$. Thus it follows from Equation (A1) that $1 - 1\Big/\left\{1 + \left[\left(x/\theta\right)^{\lambda} - 1\right]^{\beta}\right\} \le F_p^{\beta}(x)$ i.e.

$$F(x) \le F_p^{\beta}(x)\,; \quad 0 < \beta < 1. \tag{A2}$$

Case 1b Similarly, for $1 < \beta$ we have $(1 - F_p(x))^\beta + F_p^\beta(x) < 1$. Thus it follows from Equation (A1) that

$$F_p^\beta(x) \leq F(x)\,;\, 1 < \beta. \tag{A3}$$

Case 2 Now, for $\lambda < 0$, and $\beta < 0$, since $F_{ip}(x) = (\theta/x)^\lambda$, it is straightforward to show that

$$F(x) = F_{ip}^{-\beta}(x)\left\{(1 - F_{ip}(x))^\beta F_{ip}^\beta(x) \Big/ \left[F_{ip}^\beta(x) + (1 - F_{ip}(x))^\beta\right]\right\} \tag{A4}$$

Since $0 \leq F_{ip}(x), 1 - F_{ip}(x) \leq 1$, and $a^x \leq 1$ is a decreasing function for $0 \leq a \leq 1$, we have the following.
Case 2a If $-1 \leq \beta < 0$ then

$$(1 - F_{ip}(x))^\beta F_{ip}^\beta(x) \Big/ \left[F_{ip}^\beta(x) + (1 - F_{ip}(x))^\beta\right] \leq 1.$$

Thus it follows from Equation (A4) that

$$F(x) \leq F_{ip}^{-\beta}(x)\,; \quad -1 \leq \beta < 0. \tag{A5}$$

Case 2b If $-\infty < \beta < -1$ then

$$1 \leq (1 - F_{ip}(x))^\beta F_{ip}^\beta(x) \Big/ \left[F_{ip}^\beta(x) + (1 - F_{ip}(x))^\beta\right].$$

Thus it follows from Equation (A4) that

$$F_{ip}^{-\beta}(x) \leq F(x)\,; \quad -\infty < \beta < -1. \tag{A6}$$

Case 3 Now differentiating both sides of Equation (A1) gives

$$f(x) = \beta f_p(x) \cdot F_p^{\beta-1}(x)(1 - F_p(x))^{\beta-1} \Big/ \left[(1 - F_p(x))^\beta + F_p^\beta(x)\right]^2 \tag{A7}$$

Rewriting the second part of the right-hand side of Equation (A7), we get the following.

$$F_p^{\beta-1}(x)(1 - F_p(x))^{\beta-1} \Big/ \left[(1 - F_p(x))^\beta + F_p^\beta(x)\right]^2 = 1 \Big/ \left(T_{b1}(x) + T_{b2}(x)\right)^2 \tag{A8}$$

where $T_{b1}(x) = (F_p(x))^{(1-\beta)/2}(1 - F_p(x))^{(1+\beta)/2}$ and $T_{b2}(x) = (F_p(x))^{(1+\beta)/2}(1 - F_p(x))^{(1-\beta)/2}$. Then we have the following subcases.
Case 3a $0 < \beta < 1$
The denominator of Equation (A8) is monotone for each fixed value of $F_p(x)$. Thus it achieves its minimum and maximum values at the end points when $\beta = 0$ and $\beta = 1$. It is straightforward to show that the denominator is ≤ 1 in this case. Thus

$$1 \leq F_p^{\beta-1}(x)(1 - F_p(x))^{\beta-1} \Big/ \left[(1 - F_p(x))^\beta + F_p^\beta(x)\right]^2 \,; 0 < \beta < 1.$$

This implies that $\beta f_p(x) \leq f(x)$, hence integrating on both sides gives

$$\beta F_p(x) \leq F(x)\,; \quad 0 < \beta < 1. \tag{A9}$$

Case 3b $1 < \beta$
Using a similar argument, it is straightforward to show that the denominator of Equation (A8) ≥ 1. Therefore, $f(x) \leq \beta f_p(x)$. Hence integrating on both sides gives

$$F(x) \leq \beta F_p(x)\,; \quad 1 < \beta. \tag{A10}$$

Case 4 Now, differentiating both sides of Equation (A4) gives

$$f(x) = -\beta f_{ip}(x) \cdot \left[F_{ip}(x)(1 - F_{ip}(x))\right]^{\beta-1} \Big/ \left[F_{ip}^\beta(x) + (1 - F_{ip}(x))^\beta\right]^2 \tag{A11}$$

Observe that the second factor of Equation (A11) can be written as follows.

$$\left[F_{ip}(x)(1-F_{ip}(x))\right]^{\beta-1}\Big/\left[F_{ip}^{\beta}(x)+(1-F_{ip}(x))^{\beta}\right]^{2}=1\Big/\left(T_{b3}(x)+T_{b4}(x)\right)^{2}$$

where $T_{b3}(x)=(F_{ip}(x))^{(1-\beta)/2}(1-F_{ip}(x))^{(1+\beta)/2}$ and $T_{b4}(x)=(F_{ip}(x))^{(1+\beta)/2}(1-F_{ip}(x))^{(1-\beta)/2}$. Using the same argument as in case 3, we have the following subcases.

Case 4a $\beta\le -1$

Clearly, the denominator approaches ∞ as $\beta\to-\infty$ with a minimum value of 1 achieved when $\beta=-1$ in this case. From Equation (A11), this implies that $f(x)\le-\beta f_{ip}(x)$. Thus

$$F(x)\le-\beta F_{ip}(x)\,;\quad \beta\le-1 \tag{A12}$$

Case 4b $-1<\beta<0$

Again, using monotonicity it is clear that the denominator ≤ 1 in this case. From Equation (A11), this implies that $-\beta f_{ip}(x)\le f(x)$. Thus

$$-\beta F_{ip}(x)\le F(x)\,;\quad -1<\beta<0. \tag{A13}$$

Combining the results obtained in Equations (A2)–(A13) completes the proof for Theorem 3.1 □

Appendix 2. Proof for Theorem 3.2

Proof: Consider the following two cases.

Case 1 $0<\lambda$ and $0<\beta$.

Using the transformation $t=\left(x/\theta\right)^{\lambda}$, we have

$$E\left[\left(x/\theta\right)^{k}\right]=\int_{1}^{\infty}\beta t^{(k/\lambda)}(t-1)^{\beta-1}\Big/\left[1+(t-1)^{\beta}\right]^{2}\,\mathrm{d}t \tag{B1}$$

$$=\beta\int_{1}^{\infty}t^{(k/\lambda)}(t-1)^{-\beta-1}\Big/\left[(t-1)^{-\beta}+1\right]^{2}\,\mathrm{d}t \tag{B2}$$

Since the integrand is absolutely continuous, we only need to check convergence when $t\to1$ and $t\to\infty$. Consider the following two subcases.

Case 1a When $t\to1$.

In this case we consider the integral in Equation (B1) when $t\to1$. We ignore the upper limit and denote it by a star. After using the transformation $u=t-1$, Equation (B1) becomes

$$\int_{0}^{*}\beta(u+1)^{(k/\lambda)}u^{\beta-1}\Big/(1+u)^{2}\,\mathrm{d}u$$

which does exist finitely when $u\to0$.

Case 1b When $t\to\infty$.

In this case we consider the integral in Equation (B1) when $t\to\infty$. We ignore the lower limit and denote it by star, i.e.

$$=\int_{*}^{\infty}\beta t^{(k/\lambda)}(t-1)^{\beta-1}\Big/\left[1+(t-1)^{\beta}\right]^{2}\,\mathrm{d}t$$

which converges only if $k<\lambda\beta$ when $t\to\infty$.

Hence $E\left[\left(x/\theta\right)^{k}\right]$ exists finitely when $0<\lambda$ and $0<\beta$ only if $k<\lambda\beta$.

Case 2 $\lambda<0$ and $\beta<0$.

Let $\lambda=-\gamma$ and $\beta=-\delta$, where $0<\gamma,\delta$. Using the transformation $v=\left(\frac{x}{\theta}\right)^{-\gamma}-1$, we have

$$E\left[\left(x/\theta\right)^{k}\right]=\delta\int_{0}^{\infty}(v+1)^{(-k/\gamma)}v^{-(\delta+1)}\Big/\left(1+v^{-\delta}\right)^{2}\,\mathrm{d}v$$

$$=\delta\int_{0}^{\infty}1/(v+1)^{(k/\gamma)}v^{(\delta+1)}\left(1+v^{-\delta}\right)^{2}\,\mathrm{d}v \tag{B3}$$

Since the integrand is absolutely continuous for $v\in(0,\infty)$, we only need to check for convergence when $v\to0$ and when $v\to\infty$. Consider the following two subcases.

Case 2a When $v \to \infty$.
We consider Equation (B3) when $v \to \infty$. We ignore the lower limit and denote it by star.

$$E\left[(x/\theta)^k\right] = \delta \int_*^\infty 1/(v+1)^{(k/\gamma)} v^{(\delta+1)} \left(1+v^{-\delta}\right)^2 \, \mathrm{d}v$$

Since the integrand converges when $v \to \infty$, the integral does exist finitely when $v \to \infty$.
Case 2B When $v \to 0$.
We consider Equation (B3) when $v \to 0$. We ignore the upper limit and denote it by star.

$$E\left[(x/\theta)^k\right] = \delta \int_0^* 1/(v+1)^{(k/\gamma)} v^{(\delta+1)} \left(1+v^{-\delta}\right)^2 \, \mathrm{d}v$$

The integrand converges only if $k < \delta\gamma$, i.e. $k < \lambda\beta$, when $v \to 0$.
Therefore $E\left[(x/\theta)^k\right]$ exists finitely when $\lambda < 0$ and $\beta < 0$ only if $k < \lambda\beta$. This completes the proof for Theorem 3.2. □

Appendix 3. Proof for Theorem 3.3

Proof: We apply Theorem 3.1 which is on the bounds of the OP distribution. Consider the following cases.
Case 1 $0 < \lambda, 0 < \beta$.
Case 1a $0 < \lambda, 0 < \beta \le 1$
For the upper bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_\theta^\infty (x/\theta)^k f(x)\, \mathrm{d}x \\ &\le \int_\theta^\infty (x/\theta)^k\, d\left(F_p^\beta(x)\right) \quad \text{by Theorem 3.1} \\ &= \beta \cdot B\left[1-(k/\lambda), \beta\right], \text{ where } 0 < 1-k/\lambda, \text{ i.e. } k < \lambda \end{aligned} \tag{C1}$$

after using the transformation $u = (\theta/x)^\lambda$ and applying the Beta function.
Now for the lower bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_\theta^\infty (x/\theta)^k f(x)\, \mathrm{d}x \\ &\ge \int_\theta^\infty (x/\theta)^k \cdot \beta \cdot d\left(F_p(x)\right) \quad \text{by Theorem 3.1} \\ &= \lambda\beta/\left(k-\lambda\right)\ ; k < \lambda. \end{aligned} \tag{C2}$$

Case 1b $0 < \lambda, 1 < \beta \le \infty$
For the upper bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_\theta^\infty (x/\theta)^k f(x)\, \mathrm{d}x \\ &\le \int_\theta^\infty (x/\theta)^k \cdot \beta \cdot d\left(F_p(x)\right) \quad \text{by Theorem 3.1} \\ &= \lambda\beta/\left(k-\lambda\right)\ ; k < \lambda. \end{aligned} \tag{C3}$$

For the lower bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_\theta^\infty (x/\theta)^k f(x)\, \mathrm{d}x \\ &\ge \int_\theta^\infty (x/\theta)^k\, d\left(F_p^\beta(x)\right) \quad \text{by Theorem 3.1.} \\ &= \beta \cdot B\left[1-(k/\lambda), \beta\right]\ ; k < \lambda \end{aligned} \tag{C4}$$

where $B(x, y)$ is the Beta function.

Case 2 $\lambda < 0, \beta < 0$.
Case 2a $\lambda < 0, -1 \leq \beta < 0$
For the lower bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_0^\theta (x/\theta)^k f(x)\,\mathrm{d}x \\ &\geq \int_0^\theta (x/\theta)^k \cdot (-\beta) \cdot d\left(F_{ip}(x)\right) \quad \text{by Theorem 3.1} \\ &= \lambda\beta/\left(k-\lambda\right). \end{aligned} \tag{C5}$$

For the upper bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_0^\theta (x/\theta)^k f(x)\,\mathrm{d}x \\ &\leq \int_0^\theta (x/\theta)^k d\left(F_{ip}^{-\beta}(x)\right) \quad \text{by Theorem 3.1} \\ &= \lambda\beta/\left(k+\lambda\beta\right). \end{aligned} \tag{C6}$$

Case 2b $\lambda < 0, -\infty < \beta < -1$
For the upper bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_0^\theta (x/\theta)^k f(x)\,\mathrm{d}x \\ &\leq \int_0^\theta (x/\theta)^k (-\beta) \cdot d\left(F_{ip}(x)\right) \quad \text{by Theorem 3.1} \\ &= \lambda\beta/\left(k-\lambda\right). \end{aligned} \tag{C7}$$

For the lower bound;

$$\begin{aligned} E\left[(x/\theta)^k\right] &= \int_0^\theta (x/\theta)^k f(x)\,\mathrm{d}x \\ &\geq \int_0^\theta (x/\theta)^k d\left(F_{ip}^{-\beta}(x)\right) \quad \text{by Theorem 3.1} \\ &= \lambda\beta/\left(k+\lambda\beta\right). \end{aligned} \tag{C8}$$

Combining results (C1)–(C7) completes the proof for Theorem 3.3. □

Appendix 4. Proof for Theorem 4.1

Proof: Consider the following cases.
Case 1 ml estimator for β

If $\lambda < 0$ and $\beta < 0$, estimate the ml estimator for θ using the largest order statistic $X_{(N)}$ and denote $\hat{\theta} = X_{(N)} = y_n$. Using Equation (11), let

$$0 = g(\beta) = n + \beta \sum_{i=1}^{n} \log\left(U_y(y_n)\right) - 2\beta \sum_{i=1}^{n} \left(U_y(y_n)\right)^\beta \log\left(U_y(y_n)\right) \Big/ \left[1 + \left(U_y(y_n)\right)^\beta\right]$$

After applying the transformation $\lambda = -\tau$ and $\beta = -\delta$ where $0 < \tau$ and $0 < \delta$, taking the limit as $\delta \to 0$ of $g(-\delta)$ gives $\lim_{\delta\to 0} g(-\delta) = n > 0 \Rightarrow \delta = 0$ is not a solution of $g(-\delta) = 0. \Rightarrow \beta = 0$ is not a solution of $g(\beta) = 0$.

It is straightforward to show that taking the limit as $\delta \to \pm\infty$ gives the results $\lim_{\beta\to-\infty} g(\beta) = -\infty$, and $\lim_{\beta\to\infty} g(\beta) = n$.

Therefore, this proves the existence of the ml estimator for β when $\lambda < 0$ and $\beta < 0$.

If $0 < \lambda$ and $0 < \beta$, estimate the ml estimator for θ using the smallest order statistic, $X_{(1)}$ and denote $\hat{\theta} = X_{(1)} = y_1$. Let

$$g(\beta) = n + \beta \sum_{i=1}^{n} \log\left(U_y(y_1)\right) - 2\beta \sum_{i=1}^{n} \left(U_y(y_1)\right)^{\beta} \log\left(U_y(y_1)\right) \Big/ \left[1 + \left(U_y(y_1)\right)^{\beta}\right]$$

Taking the limit as $\beta \to 0$ of $g(\beta)$ gives $\lim_{\beta\to 0} g(\beta) = n > 0 \Rightarrow \beta = 0$ is not a solution to $g(\beta) = 0$. Taking the limit as $\beta \to \pm\infty$ gives $\lim_{\beta\to\infty} g(\beta) = -\infty$ and $\lim_{\beta\to -\infty} g(\beta) = n$. Therefore, this proves the existence of the ml estimator for β when $0 < \lambda$ and $0 < \beta$.

Case 2 ml estimator for λ

If $\lambda < 0$ and $\beta < 0$, estimate the ml estimator for θ using the largest order statistic $X_{(N)}$ and denote $\hat{\theta} = X_{(N)} = y_n$. Using Equation (10), let $0 = g(\lambda) = T_{41}(\underline{\theta}) + T_{42}(\underline{\theta})$, where

$$T_{41}(\underline{\theta}) = n + \lambda \sum_{i=1}^{n} \log\left(x_i/y_n\right) + \lambda(\beta - 1) \sum_{i=1}^{n} \left(x_i/y_n\right)^{\lambda} \log\left(x_i/y_n\right) \Big/ U_y(y_n), \text{ and}$$

$$T_{42}(\underline{\theta}) = -2\lambda \sum_{i=1}^{n} \beta \left(x_i/y_n\right)^{\lambda} \left(U_y(y_n)\right)^{\beta-1} \log\left(x_i/y_n\right) \left[1 + \left(U_y(y_n)\right)^{\beta}\right]$$

After applying the transformation $\lambda = -\tau$ and $\beta = -\delta$, where $0 < \tau$ and $0 < \delta$, it is straightforward to show that $\lim_{\lambda\to 0} g(\lambda) = n > 0$, which implies that $\lambda = 0$ is not a solution of $g(\lambda) = 0$.

Taking the limit as $\tau \to \pm\infty$ results in $\lim_{\lambda\to -\infty} g(\lambda) = -\infty$ and $\lim_{\lambda\to\infty} g(\lambda) = n$.

Therefore, this proves the existence of the ml estimator for λ when $\lambda < 0$ and $\beta < 0$.

If $0 < \lambda$ and $0 < \beta$, estimate the ml estimator for θ using the smallest order statistic, $X_{(1)}$ and denote $\hat{\theta} = X_{(1)} = y_1$. Let $g(\lambda) = \lambda\left(T_{61}(\underline{\theta}) + T_{62}(\underline{\theta})\right)$, where

$$T_{61}(\underline{\theta}) = n/\lambda + \sum_{i=1}^{n} \log\left(x_i/y_1\right) + (\beta - 1) \sum_{i=1}^{n} \left(x_i/y_1\right)^{\lambda} \log\left(x_i/y_1\right) \Big/ U_y(y_1), \text{ and}$$

$$T_{62}(\underline{\theta}) = -2 \sum_{i=1}^{n} \beta \left(x_i/y_1\right)^{\lambda} \left(U_y(y_1)\right)^{\beta-1} \log\left(x_i/y_1\right) \Big/ \left[1 + \left(U_y(y_1)\right)^{\beta}\right];$$

Taking the limit as $\lambda \to 0$ of $g(\lambda)$ gives $\lim_{\lambda\to 0} g(\lambda) = n > 0. \Rightarrow \lambda = 0$ is not a solution to $g(\lambda) = 0$. Taking the limit as $\lambda \to \pm\infty$ gives $\lim_{\lambda\to\infty} g(\lambda) = \infty$, and $\lim_{\lambda\to -\infty} g(\lambda) = n$.

This proves the existence of the ml estimator for λ when $0 < \lambda$ and $0 < \beta$. □

Appendix 5. Proof for Lemma 4.1

Proof: Consider the following cases.

Case 1 $0 < \lambda$ and $0 < \beta$:

Using Equation (12), rewrite $\partial l(\underline{\theta})/\partial\theta = 0$ as

$$\sum_{i=1}^{n} -\lambda\left[1 - \beta(x_i/\theta)^{\lambda}\right]/U_y(\theta) = \sum_{i=1}^{n} 2\lambda\beta(x_i/\theta)^{\lambda} \left(U_y(\theta)\right)^{\beta-1} \Big/ \left[1 + \left(U_y(\theta)\right)^{\beta}\right] \tag{E1}$$

It is straightforward to show that the left-hand side of Equation (E1) is always positive. The right-hand side of Equation (E1) is always positive only if $1 < \beta$. Thus the solution to Equation (E1) always exists when $1 < \beta$.

Case 2 $\lambda < 0$ and $\beta < 0$:

Let $\lambda = -\tau$ and $\beta = -\delta$, where $0 < \tau$ and $0 < \delta$. Using Equation (12), rewrite $\partial l(\underline{\theta})/\partial\theta = 0$ as

$$\sum_{i=1}^{n} \left[1 + \delta(x_i/\theta)^{-\tau}\right] = \sum_{i=1}^{n} \left[\delta(x_i/\theta)^{-\tau} - 1\right]\left[(x_i/\theta)^{-\tau} - 1\right]^{-\delta} \tag{E2}$$

It is straightforward to show that the left-hand side of Equation (E2) is always positive. The right-hand side of Equation (E2) is always positive only if $1 < \delta$. Thus the solution to Equation (E2) always exists when $\beta < -1$. □

Appendix 6. Proof for Theorem 4.2

Proof: Proof of ml estimators for λ and β is similar to that of theorem 4.1 with the proof for ml estimator for θ given as follows.

Set $g(\theta) = \theta \cdot \left(\partial l(\underline{\theta})/\partial\theta\right)$. It is straightforward to show that $\lim_{\theta\to 0} g(\theta) = n\lambda\beta/(\beta+1)$ when $0 < \lambda$ and $0 < \beta$, and $\lim_{\theta\to 0} g(\theta) = -n\lambda$ when $\lambda < 0$ and $\beta < 0$. Thus $\theta = 0$ is not a solution to $g(\theta) = 0$.

Taking the limit as $\theta \to \infty$ gives $\lim_{\theta\to\infty} = -n\lambda$ when $0 < \lambda$ and $0 < \beta$, and $\lim_{\theta\to\infty} = -\lambda\beta$ when $0 < \lambda$ and $0 < \beta$. Therefore the ml estimator for θ exists for the given parameters λ and β. □