

Article

Record-Based Transmuted Pareto Distribution: Properties, Estimation, and Applications to Medical and Hydrological Data

Özlem Arık¹, İbrahim Arık^{2,*}  and Yeliz M. Kantar³

¹ Department of Biostatistics and Medical Informatics, Faculty of Medicine, Necmettin Erbakan University, Konya 42090, Türkiye; ozlem.arik@erbakan.edu.tr

² Department of Business Administration, Faculty of Economics and Administrative Sciences, Kütahya Dumlupınar University, Kütahya 43100, Türkiye

³ Department of Statistics, Faculty of Science, Eskişehir Technical University, Eskişehir 26470, Türkiye; ymert@eskisehir.edu.tr

* Correspondence: ibrahim.arik@yandex.com or ibrahim.arik@dpu.edu.tr

Abstract

In this study, we introduce and study the record-based transmuted Pareto distribution (RBTPD), a three-parameter model obtained by applying the record-based transmutation map to the classical two-parameter Pareto distribution. Its relationship to the record-based transmuted power Lomax distribution is set out explicitly, and the contribution of this paper is located in the Pareto parameterization itself: the closed-form properties it permits, the fact that its lower endpoint is a threshold parameter with a substantive interpretation and requires non-regular inference, and a systematic comparison of five estimation methods. The additional parameter is the weight attached to the second upper record value of the baseline; it is confined to the unit interval and returns the Pareto distribution at zero. The proposed model is more flexible than the baseline and, unlike several existing extensions, has a density that remains bounded on the whole support. We derive the distribution; density; survival and hazard rate functions; the quantile function, in terms of the Lambert W function; moments and incomplete moments; mean deviations; Bonferroni and Lorenz curves; order statistics; record values; and Renyi and Shannon entropies. Five estimation methods are compared in a Monte Carlo study: maximum likelihood, least squares, weighted least squares, Anderson–Darling, and maximum product of spacings. For α and for p away from the boundaries of its range, the mean squared errors of all five estimators decrease with increasing sample size, a pattern consistent with consistency, though a finite range of sample sizes cannot establish it; the threshold is a support-boundary parameter and is treated separately, as the model is non-regular in it. Maximum likelihood performs best overall and is markedly the most stable numerically. The model is fitted to two real data sets, one medical and one hydrological, and gives a better fit than both the baseline Pareto and the transmuted Pareto distributions on every criterion considered. Compared with competing models outside the Pareto family, the proposed model gives generally comparable results and performs relatively well in some cases. Bootstrap-calibrated goodness-of-fit tests indicate adequate absolute fit for the medical data but not for the hydrological series, and the conditions under which the model is appropriate are stated explicitly.

Keywords: record-based transmutation; Pareto distribution; record values; hazard rate; order statistics; parameter estimation; survival analysis

MSC: 62E15; 62F10; 60E05



Academic Editors: Francisco Novoa-Muñoz and Bernardo M. Lagos-Álvarez

Received: 24 August 2026

Revised: 10 October 2026

Accepted: 12 October 2026

Published: 22 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The Pareto distribution, introduced by [1], is among the most widely used heavy-tailed models in survival and reliability analysis, actuarial science, hydrology, and the analysis of income and wealth. In biomedical applications, it arises wherever the quantity of interest is a time to event with a natural lower bound and a long right tail, such as the time to relief following treatment. Its appeal rests on a limit theorem as much as on convenience: by the results of [2,3], the conditional distribution of the excess over a high threshold converges to the generalized Pareto distribution for a wide class of parent distributions, which explains why the model recurs wherever extreme observations matter. In its classical form, however, the distribution has a single shape parameter and a strictly decreasing hazard rate, and it is therefore unable to accommodate data with non-monotone failure rates or with a degree of asymmetry that its two parameters cannot reach. This limitation has led to the creation of a substantial body of literature on ways of extending the model.

Two broad strategies have been followed. The first embeds the Pareto distribution in a larger family through a generator applied to its distribution function. Ref. [4] obtained the exponentiated Pareto distribution in this way; Ref. [5], the beta Pareto distribution; Ref. [6], the gamma Pareto distribution; Ref. [7], the Kumaraswamy Pareto distribution; and [8], the alpha power Pareto distribution. Within this line of work, the odd Burr generator of [9], treated systematically in [10], was applied by [11,12] to the Rayleigh and Pareto baselines, obtaining the odd Burr Rayleigh and odd Burr Pareto distributions and studying their moments, incomplete moments, generating functions, and order statistics.

The second strategy modifies the baseline distribution by a transmutation map. Ref. [13] proposed the quadratic rank transmutation map, which introduces a single parameter through a quadratic function of the baseline distribution function. Ref. [14] applied it to the Pareto distribution to obtain the transmuted Pareto distribution, and the same device has since been combined with other generators, giving the beta transmuted Pareto distribution of [15] and the Kumaraswamy transmuted Pareto distribution of [16]. Ref. [17] extended the idea to a cubic rank transmutation and obtained the cubic transmuted Pareto distribution. Transmutation maps are attractive because they add flexibility with a single bounded parameter and return the baseline distribution as a special case, so that the extension can be tested against its parent by a simple hypothesis on that parameter.

A different transmutation map was proposed by [18]. Rather than modifying the distribution function algebraically, they construct the new model as a mixture of the first two upper record values of the baseline distribution, the weight attached to the second record playing the role of the transmutation parameter. The resulting record-based transmutation map therefore carries a probabilistic interpretation that the quadratic and cubic maps do not, and it has been applied to a range of baselines: the Weibull distribution by [19], the generalized linear exponential distribution by [20], the Lindley distribution by [21], the unit omega distribution by [22], the Burr X distribution by [23], and the log-logistic distribution by [24].

We introduce the record-based transmuted Pareto distribution, abbreviated as RBTPD, and study it in detail. Three features of the RBTPD deserve mention at the outset. First, its distribution and density functions are available in closed form and involve only elementary functions. Second, the transmutation parameter is confined to the unit interval and is interpretable as a mixing weight, so that the reduction to the baseline Pareto distribution corresponds to a boundary value of a bounded parameter. Third, the density remains bounded on the whole support for every admissible parameter value, in contrast to several of the generator-based extensions cited above, whose densities diverge at the threshold when their extra shape parameter falls below one. The exponentiated Pareto distribution

of [4] is a concrete instance: its density contains the factor $(1 - (\theta/x)^\alpha)^{\beta-1}$, which tends to infinity as x approaches θ whenever $\beta < 1$.

The relationship between the proposed model and the record-based transmuted power Lomax distribution of Sakthivel and Nandhini [25] should be stated explicitly, since the two are close relatives. A random variable following the power Lomax distribution, the baseline of that model, has survival function $S(x) = (1 + x^\beta/\lambda)^{-\alpha}$ on the positive half-line and contains no location or threshold parameter. Setting $\beta = 1$ reduces it to the Lomax distribution, and the Lomax and Pareto Type I distributions are connected by a location shift: if X follows Pareto Type I with threshold θ and shape α , then $X - \theta$ follows Lomax with scale θ and shape α . Because the record-based transmutation map depends only on the baseline survival function, applying the map after that reduction and shift yields the present model: up to a location shift, the RBTPD corresponds to the $\beta = 1$ submodel of the record-based transmuted power Lomax distribution. We state the connection in these terms and rest no claim on the fact that the shift places the model formally outside a family that carries no location parameter.

What the Pareto parameterization contributes is, accordingly, the subject of this paper rather than a matter of nomenclature. The shift that connects the two families introduces a threshold, and the threshold is not an incidental reparameterization. It changes the inferential problem, since the model becomes non-regular in that parameter and standard asymptotic theory ceases to apply to it; it admits a substantive reading in the applications considered here, as a minimum time to relief or a lower endpoint for exceedances; and it determines when the model is appropriate and when it is not, as the applications in Section 6 show. This also marks the Pareto case off from the earlier applications of the same map. The baselines used there—exponential, Weibull, Rayleigh, Lindley, log-logistic, and Chen—are all supported on a half-line whose endpoint is fixed at the origin and is not a parameter, so the estimation problem they present is regular throughout and the standard asymptotic theory applies to every parameter. Introducing a free lower endpoint changes the problem rather than only the baseline: it removes one parameter from the scope of that theory, it makes the sample minimum the maximum likelihood estimate, and it makes the adequacy of the model depend on whether the data really possess a lower endpoint, as the two applications in Section 6 illustrate. The contribution of the present work is thus the explicit Pareto Type I formulation and the closed-form properties it permits, including a quantile function in terms of the Lambert W function and a closed-form Rényi entropy; the endpoint interpretation and the non-regular inference it requires; the systematic comparison of five estimation methods with full numerical diagnostics; and an assessment of absolute, as well as relative, adequacy of fit.

The contributions of this paper are as follows. We derive the distribution, density, survival, and hazard rate functions of an RBTPD random variable and establish the non-negativity of the density. We obtain the quantile function in terms of the Lambert W function, the moments and incomplete moments, the mean deviations, the Bonferroni and Lorenz curves, the densities of the order statistics and of the record values, and a closed-form expression for the Rényi entropy in terms of the upper incomplete gamma function. Five methods of estimation are developed and compared in a Monte Carlo study. Finally, the model is fitted to two data sets and compared with the Pareto and transmuted Pareto distributions.

The remainder of this paper is organized as follows. Section 2 constructs the RBTPD and derives its distribution, density, survival, and hazard rate functions, together with plots showing the shapes they can take. Section 3 develops the statistical properties of the model: the quantile function, the moments, the incomplete moments and related measures, the order statistics, the record values, and the entropy measures. Section 4 presents five

methods of parameter estimation. Section 5 reports the simulation study comparing them. Section 6 gives the applications to real data, and Section 7 concludes this paper.

2. The Record-Based Transmuted Pareto Distribution

2.1. Construction via the Record-Based Transmutation Map

Let $X_1, X_2, \dots$ be an i.i.d. sequence of random variables with cdf $G(x)$. Let $X_{U(1)}$ and $X_{U(2)}$ denote the first two upper record values of this sequence. Following [18], define the mixture random variable:

$$Y = X_{U(1)} \text{ with probability } (1 - p),$$

$$Y = X_{U(2)} \text{ with probability } p,$$

where $p \in [0, 1]$. By the total probability theorem, the cdf of Y is as follows:

$$F_Y(x) = (1 - p)P(X_{U(1)} \leq x) + pP(X_{U(2)} \leq x) \quad (1)$$

For the first term, $P(X_{U(1)} \leq x)$, by definition, the first upper record value coincides with the first observation X_1 . Therefore,

$$P(X_{U(1)} \leq x) = P(X_1 \leq x) = G(x) \quad (2)$$

For the second term of Equation (1), the cdf of the n -th upper record value, $X_{U(n)}$, is given by the well-known result of Arnold, Balakrishnan and Nagaraja [26]:

$$P(X_{U(n)} \leq x) = 1 - \bar{G}(x) \sum_{j=0}^{n-1} \frac{(-\log \bar{G}(x))^j}{j!}$$

where $\bar{G}(x) = 1 - G(x)$, setting $n = 2$:

$$P(X_{U(2)} \leq x) = 1 - \bar{G}(x)(1 - \log \bar{G}(x)) = G(x) + \bar{G}(x) \log \bar{G}(x) \quad (3)$$

Substituting (2) and (3) into (1), the general form of the cumulative distribution function of the record-based transmuted (RBTM) family is obtained as follows:

$$F_Y(x) = (1 - p)G(x) + p[G(x) + \bar{G}(x) \log \bar{G}(x)] = G(x) + p\bar{G}(x) \log \bar{G}(x) \quad (4)$$

where $0 \leq p \leq 1$. The corresponding general forms of the probability density function and hazard rate function are given, respectively, as follows:

$$f_Y(x) = g(x)(1 - p - p \log \bar{G}(x)), \quad 0 \leq p \leq 1 \quad (5)$$

$$h_Y(x) = \frac{g(x)(1 - p - p \log \bar{G}(x))}{\bar{G}(x)(1 - p \log \bar{G}(x))} = h_x(x) \frac{(1 - p - p \log \bar{G}(x))}{(1 - p \log \bar{G}(x))} \quad (6)$$

where $h_x(x)$ is the baseline hazard function.

2.2. Baseline Distribution

Let X be a random variable following the two-parameter Pareto distribution with a shape parameter of $\alpha > 0$ and a threshold parameter of $\theta > 0$. The pdf of the baseline distribution is as follows:

$$g(x; \alpha, \theta) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}, \quad x \geq \theta \quad (7)$$

And the corresponding cdf is

$$G(x; \alpha, \theta) = 1 - \left(\frac{\theta}{x}\right)^\alpha, \quad x \geq \theta \quad (8)$$

The hazard function is as follows:

$$h(x) = \frac{\alpha}{x} \quad (9)$$

2.3. Record-Based Transmuted Pareto Distribution (RBTPD)

Substituting the cdf, pdf, and hazard function of the two-parameter Pareto distribution into the general RBTM expressions given in (4)–(6), the corresponding functions of the RBTPD are obtained as follows:

$$F(x; \alpha, \theta, p) = 1 - \left(\frac{\theta}{x}\right)^\alpha \left[1 + p\alpha \log\left(\frac{x}{\theta}\right)\right] \quad (10)$$

$$f(x; \alpha, \theta, p) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \left[1 - p + p\alpha \log\left(\frac{x}{\theta}\right)\right] \quad (11)$$

$$h(x; \alpha, \theta, p) = \frac{\alpha \left[1 - p + p\alpha \log\left(\frac{x}{\theta}\right)\right]}{x \left[1 + p\alpha \log\left(\frac{x}{\theta}\right)\right]} \quad (12)$$

where $\theta > 0$, $x \geq \theta$, $\alpha > 0$, and $p \in [0, 1]$. It is easily seen that if $p = 0$, RBTPD reduces to the baseline Pareto distribution.

Equations (10)–(12) give the distribution, density, and hazard rate functions in turn. The survival function follows directly from (10) and is used repeatedly below, so we record it here as well: $S(x; \alpha, \theta, p) = (\theta/x)^\alpha [1 + p\alpha \log(x/\theta)]$ for $x \geq \theta$, and $S(x; \alpha, \theta, p) = 1$ otherwise. Setting $p = 0$ returns the baseline Pareto survival function, $(\theta/x)^\alpha$.

The density in (11) is non-negative for every admissible parameter value, as the following argument shows. For $x \geq \theta$, the logarithm $\log(x/\theta)$ is non-negative, and $\alpha > 0$ and $p \in [0, 1]$, so the bracketed factor satisfies $1 - p + p\alpha \log(x/\theta) \geq 1 - p \geq 0$. The factor is therefore non-negative on the whole support and is strictly positive, except in the single case $p = 1$ at $x = \theta$, where it vanishes. Since the remaining factor, $\alpha\theta^\alpha / x^{\alpha+1}$, is strictly positive for $x \geq \theta$, the density is well defined throughout.

The flexibility introduced by the transmutation parameter, p , is best appreciated graphically. Figures 1 and 2 display the density and hazard rate functions given in (11) and (12), respectively, for various combinations of α and p with $\theta = 1$. In each figure, panel (a) illustrates the effect of the shape parameter, α , with p held fixed, while panel (b) illustrates the effect of the transmutation parameter, p , with α held fixed.

Figure 1 illustrates the flexibility of the RBTPD density. Panel (a) shows that the shape parameter, α , governs the concentration of probability mass near the threshold and the rate of tail decay: larger values of α yield a higher density at $x = \theta$ and a more rapid decline. Panel (b) reveals a qualitatively different role for the transmutation parameter, p . At $p = 0$, the density reduces to that of the classical Pareto distribution, attaining its maximum at the threshold and decreasing monotonically thereafter. As p increases, the density at $x = \theta$ decreases, reaching zero at $p = 1$, and for sufficiently large values of p , the mode shifts into the interior of the support. Consequently, the RBTPD accommodates both monotonically decreasing and unimodal density shapes, whereas the baseline Pareto admits only the former.

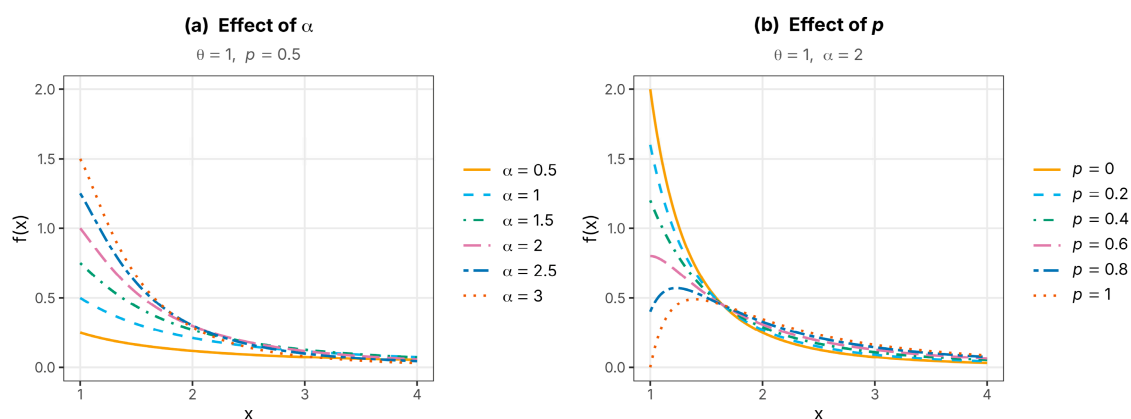


Figure 1. Probability density function of the RBTPD for selected parameter values with $\theta = 1$: (a) $p = 0.5$ fixed and $\alpha \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$; and (b) $\alpha = 2.0$ fixed and $p \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$.

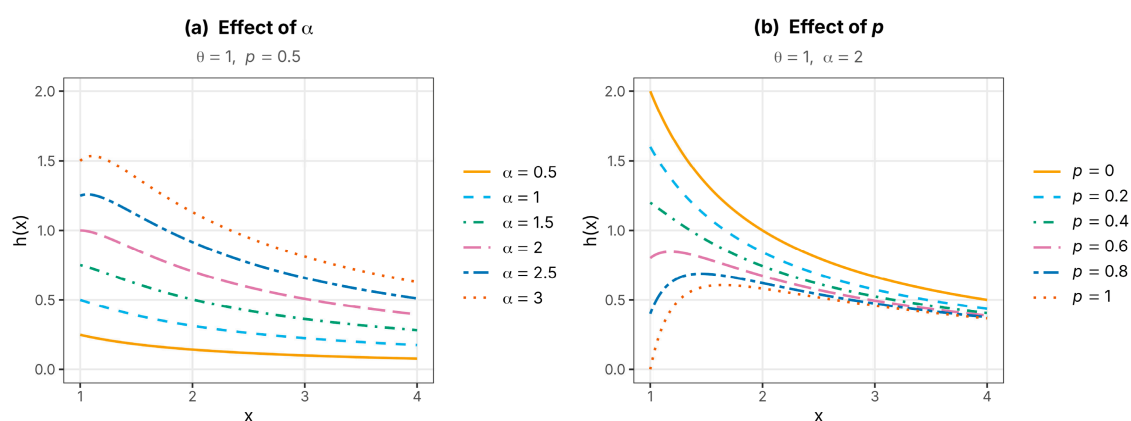


Figure 2. Hazard rate function of the RBTPD for the same parameter settings as in Figure 1.

Figure 2 displays the corresponding hazard rate functions. The RBTPD can exhibit either a decreasing failure rate (DFR) or an upside-down bathtub (UBT)-shaped hazard rate, depending on the values of α and p . In panel (a), where $p = 0.5$ is held fixed, the hazard rate is decreasing for smaller values of α , whereas for sufficiently large α , it initially increases and subsequently decreases. In panel (b), where $\alpha = 2$ is fixed, $p = 0$ recovers the classical Pareto hazard, $h(x) = \alpha/x$, which is strictly decreasing. As p increases, the initial hazard level decreases, and for $p > 0.5$, the hazard becomes non-monotone, increasing initially to an interior maximum before eventually decreasing. At the limiting case, $p = 1$, the hazard rate approaches zero at the lower endpoint, $x = \theta$. Thus, the record-based transmutation extends the classical Pareto model by allowing both DFR and UBT-shaped hazard functions, providing additional flexibility for lifetime data in which the failure risk may initially increase before declining.

3. Statistical Properties

In this section, we derive several statistical properties of the RBTPD, including the quantile function, moments, incomplete moments, order statistics, record values, and Renyi entropy.

3.1. Quantile Function

The quantile function of a continuous random variable is obtained by inverting its cumulative distribution function. Let $u \in (0, 1)$. Setting $F(x) = u$ in (10) and rearranging gives

$$\left(\frac{\theta}{x}\right)^\alpha \left[1 + p\alpha \log\left(\frac{x}{\theta}\right)\right] = 1 - u \quad (13)$$

Introducing the substitution $t = (\theta/x)^\alpha$, so that $\log(x/\theta) = -\log(t)/\alpha$, Equation (13) becomes

$$t[1 - p\log t] = 1 - u \quad (14)$$

Equation (14) does not admit an elementary closed-form solution. It can, however, be solved in terms of the Lambert W function, defined implicitly by $W(z)e^{W(z)} = z$. Writing $t = v \cdot e^{1/p}$ and rearranging, (14) reduces to the standard Lambert form, from which the quantile function of the RBTPD is obtained as

$$Q(u) = \theta \exp\left\{-\frac{1}{\alpha} \left[W_{-1}\left(-\frac{1-u}{p} e^{-\frac{1}{p}}\right) + \frac{1}{p}\right]\right\} \quad (15)$$

where $W_{-1}(\cdot)$ denotes the lower (negative) branch of the Lambert W function. The argument of W_{-1} lies in $(-1/e, 0)$ for all $u \in (0, 1)$ and $p \in (0, 1]$, so the branch is real-valued over the entire range. The principal branch, W_0 , yields values below the threshold, θ , and is therefore inadmissible.

For $p = 0$, Equation (13) reduces to $(\theta/x)^\alpha = 1 - u$, which can be inverted directly to give the quantile function of the baseline Pareto distribution,

$$Q(u) = \theta(1 - u)^{-\frac{1}{\alpha}} \quad (16)$$

In particular, the median, M , of the RBTPD is obtained by setting $u = 1/2$ in (15).

Because the RBTPD is heavy-tailed, its higher-order moments may fail to exist for small values of α . Quantile-based measures of shape are therefore particularly useful, as they remain well defined throughout the parameter space. The Bowley coefficient of skewness [27], B , and the Moors coefficient of kurtosis [28], K , are respectively given by

$$B = \frac{Q\left(\frac{3}{4}\right) - 2Q\left(\frac{1}{2}\right) + Q\left(\frac{1}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)} \quad (17)$$

and

$$K = \frac{Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right) + Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)} \quad (18)$$

These measures are well defined for all admissible parameter values, irrespective of the existence of moments. Table 1 reports the quartiles together with B and K for selected parameter combinations with $\theta = 1$.

Table 1 shows that the RBTPD is positively skewed for all parameter combinations considered, as expected for a heavy-tailed model. The Bowley coefficient decreases as α increases, indicating that larger values of the shape parameter produce a less asymmetric distribution. The transmutation parameter, p , shifts all three quartiles upward, while leaving the shape measures relatively stable, confirming that p primarily controls location and tail weight rather than asymmetry.

Table 1. Quartiles, Bowley skewness, and Moors kurtosis of the RBTPD for $\theta = 1$.

α	p	Q(1/4)	Median	Q(3/4)	B	K
1.5	0.00	1.2114	1.5874	2.5198	0.4253	1.7975
1.5	0.25	1.2866	1.8162	3.2092	0.4491	1.8959
1.5	0.50	1.4125	2.1471	4.0700	0.4471	1.9090
1.5	0.75	1.6149	2.5725	5.0253	0.4385	1.9064
1.5	1.00	1.8981	3.0615	6.0200	0.4355	1.9061
2.5	0.00	1.1220	1.3195	1.7411	0.3619	1.5650
2.5	0.25	1.1632	1.4305	2.0130	0.3709	1.5961
2.5	0.50	1.2302	1.5817	2.3214	0.3559	1.5815
2.5	0.75	1.3332	1.7628	2.6345	0.3397	1.5739
2.5	1.00	1.4689	1.9569	2.9360	0.3348	1.5738
3.5	0.00	1.0857	1.2190	1.4860	0.3338	1.4809
3.5	0.25	1.1140	1.2914	1.6483	0.3360	1.4914
3.5	0.50	1.1595	1.3875	1.8250	0.3149	1.4704
3.5	0.75	1.2280	1.4992	1.9975	0.2952	1.4633
3.5	1.00	1.3161	1.6153	2.1583	0.2894	1.4639

Random Number Generation

Random samples from the RBTPD can be generated using the inverse transform method. If $U \sim \text{Uniform}(0, 1)$, then $X = Q(U)$ follows the RBTPD with parameters α , θ , and p , where $Q(\cdot)$ is given in (15). This procedure is used in the simulation study of Section 5.

3.2. Moments

Moments play a central role in describing the location, dispersion, and shape of a distribution. In this subsection, we derive the raw and central moments of the RBTPD and examine their existence conditions.

Let X follow the RBTPD with parameters α , θ , and p . The r -th raw moment is obtained from the density in (11) as

$$E[X^r] = \int_{\theta}^{\infty} x^r f(x) dx = (1-p)I_1 + p\alpha I_2 \quad (19)$$

where the two component integrals are

$$I_1 = \int_{\theta}^{\infty} \alpha \theta^{\alpha} x^{r-\alpha-1} dx = \frac{\alpha \theta^r}{\alpha - r}$$

and, using the standard result, $\int_1^{\infty} t^{(-s-1)} \log t dt = 1/s^2$ with $s = \alpha - r > 0$,

$$I_2 = \int_{\theta}^{\infty} \alpha \theta^{\alpha} x^{r-\alpha-1} \log\left(\frac{x}{\theta}\right) dx = \frac{\alpha \theta^r}{(\alpha - r)^2}$$

Substituting these expressions into (19) yields the r -th raw moment of the RBTPD:

$$E[X^r] = \frac{\alpha \theta^r}{\alpha - r} \left[1 - p + \frac{p\alpha}{\alpha - r} \right], \quad \alpha > r \quad (20)$$

It is worth emphasizing that the r -th moment exists only when $\alpha > r$. This restriction is inherited from the baseline Pareto distribution and reflects the heavy-tailed nature of the

model. In particular, the mean requires $\alpha > 1$, the variance requires $\alpha > 2$, the skewness requires $\alpha > 3$, and the kurtosis requires $\alpha > 4$.

For notational convenience, we define

$$A_r = 1 - p + \frac{p\alpha}{\alpha - r}, E[X^r] = \frac{\alpha\theta^r A_r}{\alpha - r}$$

so that the r -th raw moment can be written compactly. Setting $r = 1$ in (20) gives the mean of the RBTPD,

$$\mu = E[X] = \frac{\alpha\theta}{\alpha - 1} \left[1 - p + \frac{p\alpha}{\alpha - 1} \right], \alpha > 1 \quad (21)$$

and combining the first two raw moments gives the variance,

$$Var(X) = \frac{\alpha\theta^2 A_2}{\alpha - 2} - \left(\frac{\alpha\theta A_1}{\alpha - 1} \right)^2, \alpha > 2 \quad (22)$$

More generally, the k -th central moment is obtained from the raw moments by the binomial expansion

$$\mu_k = \sum_{i=0}^k \binom{k}{i} (-1)^i \mu^i E[X^{k-i}] \quad (23)$$

from which the coefficients of skewness and kurtosis follow as

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}, \gamma_2 = \frac{\mu_4}{\mu_2^2} \quad (24)$$

When $p = 0$, expression (20) reduces to $E[X^r] = \alpha\theta^r / (\alpha - r)$, the r -th raw moment of the baseline Pareto distribution, and (21) reduces to the familiar Pareto mean $\alpha\theta / (\alpha - 1)$. This confirms the nested structure of the proposed model.

Table 2 reports the mean, variance, skewness and kurtosis of the RBTPD for selected parameter combinations with $\theta = 1$. Only values of α greater than four are considered, so that all four quantities exist.

Table 2. Mean, variance, skewness and kurtosis of the RBTPD for $\theta = 1$.

α	p	Mean	Variance	Skewness	Kurtosis
4.5	0.00	1.2857	0.1469	5.4659	149.4444
4.5	0.25	1.3776	0.2624	5.9658	232.3781
4.5	0.50	1.4694	0.3609	5.7033	218.0074
4.5	0.75	1.5612	0.4426	5.5081	206.5540
4.5	1.00	1.6531	0.5074	5.4413	202.8031
5.0	0.00	1.2500	0.1042	4.6476	73.8000
5.0	0.25	1.3281	0.1805	4.8955	91.5328
5.0	0.50	1.4062	0.2447	4.6419	83.9459
5.0	0.75	1.4844	0.2966	4.4625	78.9075
5.0	1.00	1.5625	0.3364	4.4009	77.3004
6.0	0.00	1.2000	0.0600	3.8103	38.6667
6.0	0.25	1.2600	0.0999	3.8638	41.3856
6.0	0.50	1.3200	0.1326	3.6230	37.2027
6.0	0.75	1.3800	0.1581	3.4582	34.6754
6.0	1.00	1.4400	0.1764	3.4009	33.8770

Table 2. Cont.

α	p	Mean	Variance	Skewness	Kurtosis
8.0	0.00	1.1429	0.0272	3.1177	22.7250
8.0	0.25	1.1837	0.0434	3.0564	22.2217
8.0	0.50	1.2245	0.0562	2.8277	19.6550
8.0	0.75	1.2653	0.0657	2.6729	18.1709
8.0	1.00	1.3061	0.0718	2.6178	17.7017

Several features emerge from Table 2. First, the mean increases monotonically with p for every value of α , confirming that the transmutation parameter shifts probability mass towards the right tail. Second, the variance likewise increases with p , indicating greater dispersion under transmutation. Third, both the skewness and the kurtosis decrease as α increases, so that larger values of the shape parameter produce a distribution that is less asymmetric and less heavy-tailed. Finally, for fixed α , the quantile-based shape measures of Table 1 vary relatively modestly with p , whereas the moment-based kurtosis in Table 2 is more sensitive to it in some settings. The transmutation parameter therefore acts principally on location and dispersion, but its effect on the standardized fourth central moment should not be dismissed as negligible.

All entries in Table 2 are positive and substantially larger than the corresponding normal values of zero and three, confirming that the RBTPD is strongly right-skewed and leptokurtic throughout the parameter space considered.

Remark on the Moment Generating Function

Since the RBTPD is a heavy-tailed distribution whose survival function decays polynomially, the moment-generating function, $M_X(t) = E[e^{tX}]$, does not exist for any $t > 0$. Moments of order r exist only for $r < \alpha$, as noted above. This behaviour is characteristic of the Pareto family and is shared by its transmuted extensions.

3.3. Incomplete Moments and Related Measures

Incomplete moments describe the contribution of the lower part of the support to the moments of a distribution. They are the natural building blocks for mean deviations, for the Bonferroni and Lorenz curves, and for measures of inequality and residual lifetime. For the RBTPD, the r -th incomplete moment is defined as

$$\varphi_r(z) = \int_{\theta}^z x^r f(x) dx \quad (25)$$

for $z > \theta$ and $r < \alpha$. Splitting the transmutation factor, as in Section 3.2, and integrating each component separately gives the closed form

$$\varphi_r(z) = \frac{\alpha\theta^r}{\alpha-r} \left\{ (1-p) \left[1 - \left(\frac{\theta}{z} \right)^{\alpha-r} \right] + \frac{p\alpha}{\alpha-r} \left[1 - \left(\frac{\theta}{z} \right)^{\alpha-r} \left(1 + (\alpha-r) \log\left(\frac{z}{\theta} \right) \right) \right] \right\} \quad (26)$$

Letting $z \rightarrow \infty$ recovers the r -th raw moment given in (20), since both $(\theta/z)^{(\alpha-r)}$ and $(\theta/z)^{(\alpha-r)} \log(z/\theta)$ vanish in the limit whenever $\alpha > r$:

$$\varphi_r(\infty) = E[X^r]$$

When $p = 0$ the second term disappears and (26) reduces to the incomplete moment of the baseline Pareto distribution,

$$\varphi_r(z) = \frac{\alpha\theta^r}{\alpha-r} \left[1 - \left(\frac{\theta}{z} \right)^{\alpha-r} \right]$$

3.3.1. Mean Deviations

The mean deviation about the mean and the mean deviation about the median measure the dispersion of a distribution and are less sensitive to extreme observations than the variance. Writing μ for the mean given in (21) and M for the median, they are obtained from the first incomplete moment as

$$\delta_1 = E[|X - \mu|] = 2\mu F(\mu) - 2\varphi_1(\mu) \quad (27)$$

and

$$\delta_2 = E[|X - M|] = \mu - 2\varphi_1(M) \quad (28)$$

where $F(\cdot)$ is the cumulative distribution function in (10). Both quantities require $\alpha > 1$ for the mean to exist. They are particularly informative for the RBTPD, since the variance itself requires the stronger condition $\alpha > 2$.

3.3.2. Bonferroni and Lorenz Curves

The Bonferroni and Lorenz curves are widely used in economics, actuarial science and reliability to describe the concentration of a distribution. For $\pi \in (0, 1)$ they are defined in terms of the first incomplete moment and the quantile function as

$$B(\pi) = \frac{\varphi_1(Q(\pi))}{\pi\mu}, \quad L(\pi) = \frac{\varphi_1(Q(\pi))}{\mu} \quad (29)$$

where $Q(\cdot)$ is the quantile function given in (15). Because the Pareto distribution is the classical model for income and wealth concentration, these curves provide a natural interpretation of the transmutation parameter: increasing p redistributes probability mass toward the upper tail and therefore lowers the Lorenz curve, indicating greater concentration.

Table 3 reports the mean, median and the two mean deviations for selected parameter combinations with $\theta = 1$.

Table 3. Mean, median, and mean deviations of the RBTPD for $\theta = 1$.

α	p	μ	Median	δ_1	δ_2
2.5	0.00	1.6667	1.3195	0.6197	0.5325
2.5	0.25	1.9444	1.4305	0.9010	0.7645
2.5	0.50	2.2222	1.5817	1.1396	0.9725
2.5	0.75	2.5000	1.7628	1.3385	1.1501
2.5	1.00	2.7778	1.9569	1.5036	1.2954
3.0	0.00	1.5000	1.2599	0.4444	0.3899
3.0	0.25	1.6875	1.3476	0.6207	0.5405
3.0	0.50	1.8750	1.4653	0.7660	0.6723
3.0	0.75	2.0625	1.6039	0.8824	0.7807
3.0	1.00	2.2500	1.7497	0.9744	0.8645
4.0	0.00	1.3333	1.1892	0.2812	0.2523
4.0	0.25	1.4444	1.2508	0.3763	0.3368
4.0	0.50	1.5556	1.3318	0.4517	0.4083

Table 3. Cont.

α	p	μ	Median	δ_1	δ_2
4.0	0.75	1.6667	1.4252	0.5087	0.4639
4.0	1.00	1.7778	1.5213	0.5499	0.5031
5.0	0.00	1.2500	1.1487	0.2048	0.1859
5.0	0.25	1.3281	1.1960	0.2679	0.2433
5.0	0.50	1.4062	1.2576	0.3167	0.2907
5.0	0.75	1.4844	1.3277	0.3521	0.3261
5.0	1.00	1.5625	1.3989	0.3759	0.3493

Two patterns are evident in Table 3. First, for every value of α , both mean deviations increase monotonically with p , confirming that the record-based transmutation widens the distribution. Second, for fixed p , both deviations decrease as α increases, since a larger shape parameter produces a lighter tail and, hence, less dispersion. The mean deviation about the median is smaller than that about the mean in every case, as expected, since the median minimizes the expected absolute deviation. The gap between the two grows with p , which reflects the increasing separation between the mean and the median as the distribution becomes more strongly right skewed.

3.4. Order Statistics

Order statistics arise naturally in reliability analysis, where the smallest and largest observations correspond to the failure times of series and parallel systems, and in the construction of goodness-of-fit statistics. Let $X_1, X_2, \dots, X_n$ be a random sample of size n from the RBTPD, and let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denote the corresponding order statistics. The density of the k -th order statistic is

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [S(x)]^{n-k} f(x) \quad (30)$$

where $F(\cdot)$ and $S(\cdot)$ are the distribution and survival functions. Writing $F(x) = 1 - S(x)$ and applying the binomial theorem gives

$$[F(x)]^{k-1} = [1 - S(x)]^{k-1} = \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j [S(x)]^j.$$

Substituting the survival function and the density of the RBTPD, and setting $m = j + n - k$, the density of the k -th order statistic can be written in the explicit form

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} \sum_{j=0}^{k-1} \binom{k-1}{j} (-1)^j \frac{\alpha \theta^{\alpha(m+1)}}{x^{\alpha(m+1)+1}} \left[1 - p + p \alpha \log\left(\frac{x}{\theta}\right) \right] \left[1 + p \alpha \log\left(\frac{x}{\theta}\right) \right]^m \quad (31)$$

Two special cases are of particular interest. For $k = 1$, the sum contains a single term, and the density of the sample minimum simplifies to

$$f_{(1)}(x) = \frac{n \alpha \theta^{n \alpha}}{x^{n \alpha + 1}} \left[1 - p + p \alpha \log\left(\frac{x}{\theta}\right) \right] \left[1 + p \alpha \log\left(\frac{x}{\theta}\right) \right]^{n-1} \quad (32)$$

whose leading factor is of Pareto form with shape parameter $n\alpha$. The density is not, however, a member of the RBTPD family for $n > 1$, because of the additional factor raised to the power $n - 1$. Two cases are exceptions. The first is $n = 1$, where the factor reduces to unity and the density returns the parent distribution. The second is $p = 0$, where the model is the classical Pareto distribution itself: the additional factor is then identically one and

the sample minimum is again Pareto, with shape parameter $n\alpha$ and the same threshold. Outside of these two cases, the statement stands. For $k = n$, we have $m = j$, and the density of the sample maximum becomes

$$f_{(n)}(x) = n\alpha \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^j \frac{\theta^{\alpha(j+1)}}{x^{\alpha(j+1)+1}} \left[1 - p + p\alpha \log\left(\frac{x}{\theta}\right)\right] \left[1 + p\alpha \log\left(\frac{x}{\theta}\right)\right]^j \quad (33)$$

3.5. Record Statistics

Record values are the successive maxima of a sequence of observations and play a central role in the analysis of extreme events, in reliability testing, and in hydrology. They are of particular interest for the model proposed here, since the RBTPD is itself constructed from the first two upper record values of a Pareto population. The present subsection therefore closes the circle by describing the record values of the RBTPD.

Let $X_1, X_2, \dots$ be a sequence of independent observations from the RBTPD, and let $X_{U(1)}, X_{U(2)}, \dots$ denote the associated sequence of upper record values. The distribution of records is governed by the cumulative hazard function, which, for the RBTPD, follows from (12) as

$$R(x) = -\log S(x) = \alpha \log\left(\frac{x}{\theta}\right) - \log\left(1 + p\alpha \log\left(\frac{x}{\theta}\right)\right) \quad (34)$$

The density of the n -th upper record value is then given by the standard result [26],

$$f_{U(n)}(x) = \frac{[R(x)]^{n-1}}{(n-1)!} f(x) \quad (35)$$

Substituting (11) and (34) gives the explicit form

$$f_{U(n)}(x) = \frac{\alpha \theta^\alpha}{(n-1)! x^{\alpha+1}} \left[1 - p + p\alpha \log\left(\frac{x}{\theta}\right)\right] \left[\alpha \log\left(\frac{x}{\theta}\right) - \log\left(1 + p\alpha \log\left(\frac{x}{\theta}\right)\right)\right]^{n-1} \quad (36)$$

for $x > \theta$ and $n = 1, 2, \dots$. Setting $n = 1$ reduces (36) to the parent density in (11), as it must, since the first upper record coincides with the first observation.

The joint density of the first n upper records is obtained from the hazard rate function in (12) as

$$f(x_1, \dots, x_n) = \left[\prod_{i=1}^{n-1} h(x_i)\right] f(x_n) \quad (37)$$

This expression holds on the ordered region $\theta < x_1 < x_2 < \dots < x_n$ and is zero elsewhere, the ordering being inherent in the definition of successive upper records and the lower bound being the threshold of the underlying distribution.

3.6. Entropy Measures

Entropy quantifies the uncertainty associated with a random variable and is widely used in information theory, reliability, and model selection. For a continuous random variable, X , with density, f , the Renyi entropy [29] of order δ is defined as

$$I_R(\delta) = \frac{1}{1-\delta} \log \int_{\theta}^{\infty} [f(x)]^{\delta} dx \quad (38)$$

for $\delta > 0$ and $\delta \neq 1$. Substituting the density of the RBTPD from (11) and applying the transformation $s = \log(x/\theta)$, the integral reduces to

$$\int_{\theta}^{\infty} [f(x)]^{\delta} dx = \alpha^{\delta} \theta^{1-\delta} \int_0^{\infty} e^{-\kappa s} [1 - p + p\alpha s]^{\delta} ds \quad (39)$$

where

$$\kappa = \delta(\alpha + 1) - 1$$

Convergence of the integral in (39) requires $\kappa > 0$, that is, $\delta > 1/(\alpha + 1)$. A second substitution, $u = 1 - p + p\alpha s$, transforms the remaining integral into an upper incomplete gamma integral and yields, for $p > 0$,

$$\int_{\theta}^{\infty} [f(x)]^{\delta} dx = \frac{\alpha^{\delta} \theta^{1-\delta}}{p\alpha} e^{c(1-p)} c^{-(\delta+1)} \Gamma(\delta + 1, c(1-p)) \quad (40)$$

with

$$c = \frac{\kappa}{p\alpha}$$

where $\Gamma(.,.)$ denotes the upper incomplete gamma function. The Renyi entropy of the RBTPD is therefore available in closed form as

$$I_R(\delta) = \frac{1}{1-\delta} \log \left\{ \frac{\alpha^{\delta} \theta^{1-\delta}}{p\alpha} e^{c(1-p)} c^{-(\delta+1)} \Gamma(\delta + 1, c(1-p)) \right\} \quad (41)$$

valid for $\delta > 1/(\alpha + 1)$, $\delta \neq 1$ and $0 < p \leq 1$. No series truncation is required.

When $p = 0$, the transmutation factor equals unity, the integral in (39) becomes elementary, and (41) reduces to the Renyi entropy of the baseline Pareto distribution,

$$I_R(\delta) = \frac{1}{1-\delta} \log \left(\frac{\alpha^{\delta} \theta^{1-\delta}}{\kappa} \right) \quad (42)$$

Shannon Entropy

The Shannon entropy, $H(X) = -E[\log f(X)]$, is recovered as the limit of the Renyi entropy as $\delta \rightarrow 1$. Expanding $\log f(X)$ from (11) and using the closed-form result, $E[\log X] = \log \theta + (1 + p)/\alpha$, gives the decomposition,

$$H(X) = \log \left(\frac{\theta}{\alpha} \right) + \frac{(\alpha + 1)(1 + p)}{\alpha} - E \left[\log \left(1 - p + p\alpha \log \left(\frac{X}{\theta} \right) \right) \right] \quad (43)$$

The first two terms are elementary. The remaining expectation involves the logarithm of the transmutation factor and does not admit an elementary closed form for general p ; it is expressed, with the integration variable written in lower case throughout, as the one-dimensional integral of $\log(1 - p + p\alpha s)$ against the density $\alpha e^{(-\alpha s)}(1 - p + p\alpha s)$ on $(0, \infty)$ and evaluated by adaptive quadrature, using the integrate function of R, which is the routine employed for every numerical integral reported in this paper. For $p = 0$, that term vanishes and (43) collapses to the classical Pareto result,

$$H(X) = \log \left(\frac{\theta}{\alpha} \right) + 1 + \frac{1}{\alpha} \quad (44)$$

Table 4 reports the Renyi entropy at four values of δ , together with the Shannon entropy for selected parameter combinations with $\theta = 1$.

Two regularities emerge from Table 4. For every fixed α and δ , entropy increases monotonically with p : the transmutation spreads probability mass over a wider effective range and therefore raises uncertainty. Conversely, for fixed p and δ , entropy decreases as α increases, since a larger shape parameter concentrates the distribution near the threshold, θ . Within each row, entropy decreases in δ , which is the expected behaviour of the Renyi family, with the Shannon value lying between the entries at $\delta = 0.5$ and $\delta = 2$ in every case.

Negative entries are not anomalous. Differential entropy is not bounded below, and negative values can arise when the density exceeds unity over a sufficiently influential region. In Table 4, they occur at the larger values of α and δ , combined with a smaller p ,

that is, precisely for the combinations in which the density is most sharply peaked at the threshold; no negative entry appears at $p = 1$.

Table 4. Renyi and Shannon entropies of the RBTPD for $\theta = 1$.

α	p	$\delta=0.5$	$\delta=2$	$\delta=3$	$\delta=5$	Shannon
2.0	0.00	2.0794	0.2231	0.0000	−0.2067	0.8069
2.0	0.25	2.5732	0.5344	0.3022	0.0892	1.1540
2.0	0.50	2.8618	0.8580	0.6412	0.4457	1.4518
2.0	0.75	3.0684	1.1584	0.9722	0.8132	1.7005
2.0	1.00	3.2242	1.3626	1.1836	1.0288	1.8841
2.5	0.00	1.4917	−0.0408	−0.2488	−0.4445	0.4837
2.5	0.25	1.8953	0.2597	0.0464	−0.1528	0.8059
2.5	0.50	2.1423	0.5659	0.3714	0.1930	1.0787
2.5	0.75	2.3210	0.8409	0.6763	0.5333	1.3024
2.5	1.00	2.4541	1.0170	0.8578	0.7174	1.4609
3.0	0.00	1.0986	−0.2513	−0.4490	−0.6372	0.2347
3.0	0.25	1.4536	0.0415	−0.1588	−0.3485	0.5402
3.0	0.50	1.6758	0.3355	0.1559	−0.0106	0.7964
3.0	0.75	1.8374	0.5930	0.4427	0.3104	1.0034
3.0	1.00	1.9557	0.7501	0.6041	0.4733	1.1453
4.0	0.00	0.5754	−0.5754	−0.7599	−0.9384	−0.1363
4.0	0.25	0.8788	−0.2928	−0.4767	−0.6540	0.1484
4.0	0.50	1.0731	−0.0150	−0.1757	−0.3271	0.3837
4.0	0.75	1.2143	0.2201	0.0873	−0.0317	0.5699
4.0	1.00	1.3146	0.3533	0.2234	0.1045	0.6909

4. Parameter Estimation

In this section, five methods are used to estimate the parameter vector, $\Psi = (\alpha, \theta, p)^T$, of the RBTPD: maximum likelihood, least squares, weighted least squares, Anderson–Darling, and maximum product of spacings. Throughout, $x_1, \dots, x_n$ is an observed random sample, $x_{(1)} \leq \dots \leq x_{(n)}$ is the corresponding order statistics, and $F(\cdot; \Psi)$ denotes the distribution function given in (10). Except for the threshold component of the maximum likelihood estimator, which is shown, in Section 4.1, to equal $x_{(1)}$, none of the resulting systems admits a closed-form solution, so all are solved numerically over $\alpha > 0$, $0 < \theta \leq x_{(1)}$, and $0 \leq p \leq 1$. The finite-sample behaviour of the five estimators is compared in Section 5.

Before the individual criteria are described, one structural feature of the model must be recorded, because it governs what can and cannot be claimed about the estimators that follow. The support of the RBTPD is the half-line $x \geq \theta$, so the density does not have common support across the parameter space, and the standard regularity conditions fail in θ . Two consequences follow. Because the support depends on θ , the regular Fisher-information theory is not applicable to the threshold, and the usual information-based asymptotic variance and standard error are not valid for it; θ cannot therefore appear in an information matrix alongside the remaining parameters. The threshold is estimated by a non-regular endpoint estimator whose rate is faster than $n^{1/2}$. Its rate follows from the behaviour of the density at the lower endpoint, which, for this model, is $f(\theta+) = \alpha(1 - p)/\theta$. For every $p < 1$, this limit is finite and strictly positive, so the distribution function is locally

linear at θ , the correct normalization is $n(X_{(1)} - \theta)$, and the limiting law is exponential with mean $\theta/\{\alpha(1 - p)\}$. The threshold is thus estimated at rate n rather than the usual $n^{1/2}$, and its limiting distribution is exponential rather than normal, so symmetric intervals built on a standard error are not meaningful for it. The boundary case $p = 1$ must be treated separately: there, $f(\theta+) = 0$, the distribution function is locally quadratic, and the rate falls to $n^{1/2}$. All statements below refer to the interior case $p < 1$ unless the boundary case is named. The parameter α is regular, as is p in the interior of its range; the values $p = 0$ and $p = 1$ are boundary points of the unit interval, at which the usual asymptotic theory does not apply either, and the likelihood ratio test of Section 6 is adjusted accordingly. The standard theory is therefore invoked for α and for p with $0 < p < 1$, once θ is held fixed. Accordingly, every asymptotic statement made in this paper is restricted to α and to p in the open interval; inference for the threshold is carried out by bootstrap in Section 6, and the simulation results for θ in Section 5 are interpreted separately from those for the regular parameters.

4.1. Maximum Likelihood Estimation

The log-likelihood function of the RBTPD follows from the density in (11) as

$$\uparrow(\Psi) = n \log \alpha + n \alpha \log \theta - (\alpha + 1) \sum_{i=1}^n \log x_i + \sum_{i=1}^n \log \left[1 - p + p \alpha \log \left(\frac{x_i}{\theta} \right) \right] \quad (45)$$

For the classical Pareto distribution, the log-likelihood is increasing in θ , so the maximum likelihood estimator of the threshold parameter is the sample minimum, $\hat{\theta} = x_{(1)}$; the remaining parameters are obtained by maximizing (45) numerically. Conditional standard errors for α and p , given $\theta = x_{(1)}$, are computed from the inverse of the corresponding 2×2 negative Hessian of (45) evaluated at the estimates.

The corresponding argument for the RBTPD is as follows. Differentiating the log-likelihood (45) with respect to θ gives $\partial \ell / \partial \theta = (\alpha / \theta) \{n - p \sum 1/u_i\}$, where $u_i = 1 - p + p \alpha \log(x_i / \theta)$. The sum appearing here is determined by the score equation for p . Writing $u_i = 1 + p(\alpha L_i - 1)$ with $L_i = \log(x_i / \theta)$, so that $\alpha L_i - 1 = (u_i - 1)/p$, the derivative of the log-likelihood with respect to p is $\partial \ell / \partial p = \sum (\alpha L_i - 1)/u_i = (1/p) \sum (1 - 1/u_i)$. At any interior optimum in p , this vanishes, which gives the identity $\sum 1/u_i = n$. Substituting it into the derivative above yields $\partial \ell / \partial \theta = n \alpha (1 - p) / \theta$, which is strictly positive for every $p < 1$. The log-likelihood is therefore strictly increasing in θ along the curve of interior maximizers in p , the constraint $\theta \leq x_{(1)}$ binds, and $\hat{\theta} = x_{(1)}$ exactly. The case $p = 0$ is immediate, since $u_i = 1$ for all i and the derivative reduces to $n \alpha / \theta > 0$, recovering the baseline Pareto result. The argument therefore establishes $\hat{\theta} = x_{(1)}$ for $0 \leq p < 1$. The boundary value $p = 1$ is non-standard and is not covered by it: the Kuhn–Tucker condition, $\partial \ell / \partial p \geq 0$, required there reads $\sum 1/u_i \leq n$, which makes $\partial \ell / \partial \theta = (\alpha / \theta)(n - \sum 1/u_i)$ non-negative and so drives θ upward, while at $\theta = x_{(1)}$, the quantity $u_i = \alpha \log(x_i / \theta)$ vanishes at the smallest observation, so that the log-likelihood tends to $-\infty$ there. The two requirements can be reconciled only in one way: an interior maximum in θ needs $\partial \ell / \partial \theta = 0$, which, given the expression above, means $\sum 1/u_i = n$ exactly, and that single configuration satisfies the Kuhn–Tucker inequality with equality. We do not attempt to rule it out in general. The numerical work reflects this: the optimizations reported below are constrained to $p < 1$, so that value cannot be returned by them, and the boundary was therefore examined separately. Estimating α and θ jointly with p held at 1 and θ free below $x_{(1)}$ gives a log-likelihood of -15.910 for the relief-time data and -384.452 for the Floyd River data, against -15.672 and -384.379 for the fits reported in Section 6; the boundary therefore does not improve on the interior solution in either application. The supplementary code includes this comparison. Numerical verification agrees with the analytic result: the profile log-likelihood, maximized over α and p at each point of a grid of θ values spanning $(0, x_{(1)})$,

attained its maximum at the upper endpoint for both data sets of Section 6 and for samples generated with $p = 0.3, 0.7$ and 0.95 . Because θ is a threshold and the model is non-regular in it, the standard errors obtained from the observed information matrix are those of α and p conditional on $\theta = x_{(1)}$, and they are reported as such throughout; interval estimates that account for the sampling variability of the threshold are obtained by parametric bootstrap in Section 6.

All five criteria are optimized numerically with the L-BFGS-B algorithm, subject to the box constraints $\alpha > 0$, $0 < \theta \leq x_{(1)}$ and $0 \leq p \leq 1$, with the gradients approximated numerically. Each optimization is started from three dispersed initial values and the best of the three solutions is retained; the convergence tolerance is the default $\text{factr} = 1 \times 10^7$ with a maximum of 500 iterations. All computations were performed in R 4.4.2, with the Lambert W function evaluated through the lamW package (version 2.2.7). The complete code, the random seeds and the session information are supplied as Supplementary Material.

4.2. Least Squares Estimation

If $X_{(1)} \leq \dots \leq X_{(n)}$ are the order statistics of a random sample from a continuous distribution, then $F(X_{(i)})$ is a $Beta(i, n - i + 1)$ variate, so that $E[F(X_{(i)})] = i/(n + 1)$. The least squares estimators of Swain, Venkatraman, and Wilson [30] minimize

$$S(\Psi) = \sum_{i=1}^n \left[F(x_{(i)}; \Psi) - \frac{i}{n+1} \right]^2 \quad (46)$$

4.3. Weighted Least Squares Estimation

The variance of $F(X_{(i)})$ is $i(n - i + 1) / [(n + 1)^2(n + 2)]$ and is largest in the middle of the sample. Weighting each term by its reciprocal gives the weighted least squares estimators, which minimize

$$W(\Psi) = \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[F(x_{(i)}; \Psi) - \frac{i}{n+1} \right]^2 \quad (47)$$

4.4. Anderson–Darling Estimation

The Anderson–Darling statistic [31] is an integrated squared difference between the empirical and fitted distribution functions, weighted to emphasize both tails. The corresponding estimators minimize

$$A(\Psi) = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) \left\{ \log F(x_{(i)}; \Psi) + \log(1 - F(x_{(n+1-i)}; \Psi)) \right\} \quad (48)$$

4.5. Maximum Product of Spacings Estimation

Let $D_i(\Psi) = F(x_{(i)}; \Psi) - F(x_{(i-1)}; \Psi)$ for $i = 1, \dots, n + 1$, with the conventions $F(x_{(0)}; \Psi) = 0$ and $F(x_{(n+1)}; \Psi) = 1$, so that the $n + 1$ spacings sum to unity. The method of Cheng and Amin [32] and Ranney [33] maximizes the geometric mean of the spacings or, equivalently,

$$H(\Psi) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i(\Psi) \quad (49)$$

If two observations coincide, the corresponding spacing vanishes and the criterion is undefined; the adjustment of Cheng and Amin [32], in which a block of k tied observations contributes the density raised to the power k , is applied in that case. The method of maximum product of spacings is implemented by minimizing the negative of its objective function, so that all five criteria are handled by the same minimization routine. Each of the four criteria above is minimized with respect to the unknown parameters, and the resulting

estimators are referred to as the least squares, weighted least squares, Anderson–Darling, and maximum product of spacings estimators.

5. Monte Carlo Simulation Study

This section reports on a Monte Carlo study comparing the finite-sample performance of the five estimation methods derived in Section 4. The design follows the standard practice in the literature: samples are generated from the RBTPD at known parameter values, each of the five criteria is optimized, and the resulting estimates are summarized by their bias and mean squared error.

5.1. Design

Random samples are generated by the inverse transform method of Section 3.1, using the closed-form quantile function (15) evaluated through the lower branch of the Lambert W function. Four parameter settings are considered, chosen to span the range of behavior identified in Sections 2 and 3:

- (i) $(\alpha, \theta, p) = (2.5, 1.0, 0.3)$, a moderately heavy tail with weak transmutation;
- (ii) $(\alpha, \theta, p) = (2.5, 1.0, 0.7)$, the same tail with strong transmutation;
- (iii) $(\alpha, \theta, p) = (3.5, 1.0, 0.5)$, a lighter tail with intermediate transmutation;
- (iv) $(\alpha, \theta, p) = (2.0, 1.0, 0.9)$, a heavy tail with transmutation close to its upper limit.

Five sample sizes are used, $n = 20, 50, 100, 200$, and 500 , and each configuration is replicated $N = 1000$ times. All optimizations are performed with the L-BFGS-B algorithm, with the gradients approximated numerically, subject to the constraints $\alpha > 0$, $0 < \theta \leq x_{(1)}$, and $0 \leq p \leq 1$, and are started from three dispersed initial values, one of which is placed at the true parameter values of the setting being simulated. Since the true values are unknown in an application, the disagreement between starting values reported in Section 5.2 should be read as a lower bound on what a practitioner would encounter. Consistent with Section 4.1, the maximum likelihood optimization returned $\hat{\theta} = x_{(1)}$ in every replicate; the remaining four criteria estimate θ freely within the same region. Convergence rates are reported in Section 5.2, together with the remaining numerical diagnostics; the proportion of replicates in which at least one of the three optimizations converged ranges from 91.9 per cent to 100 per cent across the cells of the design, the lower rates occurring for the distance-based criteria at the smallest sample sizes. In total, 300,000 optimizations were performed, with five criteria from three starting values in each of the 20,000 generated samples. All computations were carried out in R 4.4.2; the Lambert W function was evaluated with the lamW package (version 2.2.7), and the results were written with openxlsx (version 4.2.8). The random seeds are set at the head of the script supplied as Supplementary Material, and the session information, including all package versions, is recorded in the accompanying session_info.txt file.

Because each entry of Tables 5–8 is itself an estimate based on 1000 replications, Monte Carlo standard errors were computed for every bias and mean squared error, and are reported in full in the Supplementary Results File, alongside the number of replications actually used in each cell. They are given there rather than in the tables below, which would otherwise carry twelve numerical columns. Their magnitude is small relative to the quantities being compared: the Monte Carlo standard error of a bias estimate has median 0.005 and never exceeds 0.036, with the largest values occurring for $\hat{\alpha}$ at the smallest sample sizes; meanwhile, for the mean squared errors, it is a median of 4.9 per cent, and, at most, 18.5 per cent of the value reported. Differences between methods that are of the order of these figures should not be read as real, and the comparisons drawn in Section 5.2 are confined to differences comfortably larger.

Table 5. Bias and MSE of the five estimators for $(\alpha, \theta, p) = (2.5, 1.0, 0.3)$.

n	Method	Bias α	Bias θ	Bias p	MSE α	MSE θ	MSE p
20	MLE	0.2018	0.0286	−0.0356	0.6881	0.0016	0.0679
20	LSE	0.0727	−0.0355	0.0920	0.8847	0.0042	0.1860
20	WLSE	0.0798	−0.0249	0.0723	0.8245	0.0027	0.1724
20	ADE	0.1535	−0.0142	0.0756	0.7975	0.0021	0.1597
20	MPSE	−0.1618	−0.0097	−0.0335	0.6239	0.0015	0.1219
50	MLE	0.0449	0.0115	−0.0345	0.3452	0.0003	0.0567
50	LSE	0.0465	−0.0199	0.0487	0.5047	0.0013	0.1149
50	WLSE	0.0361	−0.0098	0.0208	0.4536	0.0005	0.0978
50	ADE	0.0530	−0.0074	0.0224	0.4160	0.0005	0.0912
50	MPSE	−0.1894	−0.0019	−0.0775	0.3851	0.0002	0.0799
100	MLE	−0.0212	0.0059	−0.0326	0.2027	6.95×10^{-5}	0.0433
100	LSE	−0.0112	−0.0115	0.0186	0.2824	0.0005	0.0725
100	WLSE	−0.0153	−0.0041	−0.0023	0.2439	0.0001	0.0566
100	ADE	0.0052	−0.0040	0.0070	0.2333	0.0001	0.0557
100	MPSE	−0.1937	−0.0002	−0.0815	0.2441	3.93×10^{-5}	0.0555
200	MLE	−0.0182	0.0030	−0.0232	0.1424	1.72×10^{-5}	0.0315
200	LSE	0.0212	−0.0065	0.0188	0.2046	0.0002	0.0478
200	WLSE	−0.0006	−0.0017	−0.0024	0.1727	2.36×10^{-5}	0.0388
200	ADE	0.0130	−0.0021	0.0049	0.1691	3.39×10^{-5}	0.0379
200	MPSE	−0.1585	0.0000	−0.0725	0.1721	8.38×10^{-6}	0.0403
500	MLE	−0.0126	0.0012	−0.0127	0.0713	2.61×10^{-6}	0.0169
500	LSE	−0.0044	−0.0033	0.0020	0.1128	3.76×10^{-5}	0.0253
500	WLSE	−0.0156	−0.0007	−0.0088	0.0902	4.63×10^{-6}	0.0209
500	ADE	−0.0036	−0.0012	−0.0019	0.0862	7.51×10^{-6}	0.0199
500	MPSE	−0.1109	−0.0002	−0.0482	0.0899	1.33×10^{-6}	0.0218

Table 6. Bias and MSE of the five estimators for $(\alpha, \theta, p) = (2.5, 1.0, 0.7)$.

n	Method	Bias α	Bias θ	Bias p	MSE α	MSE θ	MSE p
20	MLE	−0.1532	0.0593	−0.2698	0.3962	0.0064	0.1412
20	LSE	−0.2711	−0.0218	−0.1081	0.5884	0.0065	0.1827
20	WLSE	−0.2441	−0.0169	−0.1030	0.5287	0.0052	0.1744
20	ADE	−0.1720	0.0008	−0.1041	0.4517	0.0043	0.1566
20	MPSE	−0.4259	−0.0060	−0.2179	0.5652	0.0045	0.1999
50	MLE	−0.1113	0.0255	−0.1442	0.1667	0.0012	0.0628
50	LSE	−0.1337	−0.0197	−0.0215	0.2454	0.0027	0.0960
50	WLSE	−0.1007	−0.0134	−0.0209	0.2017	0.0016	0.0800
50	ADE	−0.0802	−0.0065	−0.0271	0.1894	0.0015	0.0742
50	MPSE	−0.2566	−0.0034	−0.1284	0.2564	0.0010	0.1008
100	MLE	−0.0656	0.0126	−0.0813	0.0893	0.0003	0.0297

Table 6. Cont.

n	Method	Bias α	Bias θ	Bias p	MSE α	MSE θ	MSE p
100	LSE	−0.0714	−0.0195	0.0118	0.1222	0.0018	0.0547
100	WLSE	−0.0524	−0.0102	−0.0027	0.1020	0.0007	0.0413
100	ADE	−0.0396	−0.0088	0.0007	0.0969	0.0008	0.0398
100	MPSE	−0.1403	−0.0021	−0.0698	0.1217	0.0003	0.0445
200	MLE	−0.0265	0.0063	−0.0360	0.0379	8.15×10^{-5}	0.0125
200	LSE	−0.0091	−0.0163	0.0394	0.0520	0.0011	0.0286
200	WLSE	−0.0069	−0.0063	0.0140	0.0424	0.0003	0.0186
200	ADE	−0.0026	−0.0065	0.0171	0.0418	0.0003	0.0190
200	MPSE	−0.0646	−0.0009	−0.0289	0.0463	5.87×10^{-5}	0.0166
500	MLE	−0.0079	0.0026	−0.0125	0.0160	1.33×10^{-5}	0.0042
500	LSE	0.0110	−0.0123	0.0394	0.0231	0.0006	0.0149
500	WLSE	0.0053	−0.0021	0.0092	0.0182	3.54×10^{-5}	0.0056
500	ADE	0.0087	−0.0032	0.0143	0.0183	7.75×10^{-5}	0.0064
500	MPSE	−0.0239	−0.0002	−0.0097	0.0168	6.84×10^{-6}	0.0045

Table 7. Bias and MSE of the five estimators for $(\alpha, \theta, p) = (3.5, 1.0, 0.5)$.

n	Method	Bias α	Bias θ	Bias p	MSE α	MSE θ	MSE p
20	MLE	−0.0300	0.0280	−0.1558	0.9447	0.0015	0.0906
20	LSE	−0.2181	−0.0257	−0.0086	1.2820	0.0030	0.1716
20	WLSE	−0.1980	−0.0179	−0.0249	1.1990	0.0020	0.1641
20	ADE	−0.0855	−0.0080	−0.0206	1.0900	0.0016	0.1518
20	MPSE	−0.4848	−0.0067	−0.1440	1.1600	0.0012	0.1562
50	MLE	−0.0791	0.0110	−0.0918	0.4985	0.0002	0.0649
50	LSE	−0.1186	−0.0167	0.0080	0.7208	0.0011	0.1235
50	WLSE	−0.0943	−0.0095	−0.0090	0.6437	0.0005	0.1051
50	ADE	−0.0552	−0.0060	−0.0074	0.5710	0.0004	0.0951
50	MPSE	−0.3529	−0.0026	−0.1168	0.6817	0.0002	0.1051
100	MLE	−0.0905	0.0057	−0.0653	0.3100	6.38×10^{-5}	0.0419
100	LSE	−0.1053	−0.0111	−0.0000	0.4592	0.0005	0.0794
100	WLSE	−0.0817	−0.0041	−0.0181	0.3835	0.0001	0.0579
100	ADE	−0.0628	−0.0036	−0.0126	0.3691	0.0002	0.0567
100	MPSE	−0.2744	−0.0004	−0.0978	0.4437	3.98×10^{-5}	0.0649
200	MLE	−0.0625	0.0030	−0.0391	0.1461	1.79×10^{-5}	0.0219
200	LSE	−0.0415	−0.0069	0.0093	0.2107	0.0003	0.0383
200	WLSE	−0.0461	−0.0014	−0.0127	0.1719	2.64×10^{-5}	0.0255
200	ADE	−0.0289	−0.0018	−0.0050	0.1629	3.90×10^{-5}	0.0245
200	MPSE	−0.1849	−0.0000	−0.0652	0.2162	1.08×10^{-5}	0.0352
500	MLE	−0.0080	0.0012	−0.0104	0.0547	2.69×10^{-6}	0.0081
500	LSE	0.0117	−0.0032	0.0143	0.0830	5.08×10^{-5}	0.0138

Table 7. Cont.

n	Method	Bias α	Bias θ	Bias p	MSE α	MSE θ	MSE p
500	WLSE	0.0055	−0.0007	0.0029	0.0625	5.07×10^{-6}	0.0091
500	ADE	0.0136	−0.0011	0.0077	0.0618	8.69×10^{-6}	0.0092
500	MPSE	−0.0172	−0.0003	0.0029	0.0250	1.49×10^{-6}	0.0054

Table 8. Bias and MSE of the five estimators for $(\alpha, \theta, p) = (2.0, 1.0, 0.9)$.

n	Method	Bias α	Bias θ	Bias p	MSE α	MSE θ	MSE p
20	MLE	−0.1049	0.1241	−0.3861	0.2352	0.0252	0.2130
20	LSE	−0.1692	0.0222	−0.2016	0.3276	0.0142	0.1851
20	WLSE	−0.1510	0.0238	−0.1938	0.2978	0.0132	0.1732
20	ADE	−0.1129	0.0426	−0.2006	0.2648	0.0136	0.1665
20	MPSE	−0.3042	0.0213	−0.2958	0.3234	0.0147	0.2368
50	MLE	−0.0861	0.0625	−0.2139	0.0834	0.0065	0.0734
50	LSE	−0.0829	0.0083	−0.0829	0.1195	0.0051	0.0649
50	WLSE	−0.0623	0.0063	−0.0665	0.0955	0.0040	0.0524
50	ADE	−0.0573	0.0146	−0.0777	0.0881	0.0039	0.0515
50	MPSE	−0.1717	0.0044	−0.1400	0.1147	0.0042	0.0813
100	MLE	−0.0392	0.0376	−0.1260	0.0348	0.0024	0.0276
100	LSE	−0.0388	0.0026	−0.0366	0.0484	0.0024	0.0281
100	WLSE	−0.0227	0.0024	−0.0277	0.0396	0.0017	0.0224
100	ADE	−0.0223	0.0040	−0.0298	0.0378	0.0018	0.0222
100	MPSE	−0.0774	−0.0001	−0.0631	0.0405	0.0016	0.0270
200	MLE	−0.0224	0.0202	−0.0714	0.0164	0.0007	0.0102
200	LSE	−0.0163	−0.0024	−0.0125	0.0224	0.0012	0.0133
200	WLSE	−0.0078	−0.0043	−0.0036	0.0186	0.0008	0.0111
200	ADE	−0.0084	−0.0034	−0.0054	0.0179	0.0008	0.0110
200	MPSE	−0.0393	−0.0031	−0.0265	0.0175	0.0006	0.0100
500	MLE	−0.0117	0.0094	−0.0322	0.0061	0.0002	0.0031
500	LSE	−0.0013	−0.0071	0.0121	0.0081	0.0007	0.0060
500	WLSE	0.0002	−0.0059	0.0106	0.0067	0.0004	0.0050
500	ADE	0.0000	−0.0077	0.0144	0.0066	0.0004	0.0055
500	MPSE	−0.0179	−0.0014	−0.0097	0.0063	0.0001	0.0034

For each parameter, the two performance measures are

$$\text{Bias}(\hat{\Psi}_j) = \frac{1}{N} \sum_{k=1}^N (\hat{\Psi}_{j,k} - \Psi_j) \quad (50)$$

$$\text{MSE}(\hat{\Psi}_j) = \frac{1}{N} \sum_{k=1}^N (\hat{\Psi}_{j,k} - \Psi_j)^2 \quad (51)$$

where Ψ_j denotes the j -th component of the true parameter vector, and the second subscript indexes the replication.

5.2. Discussion of the Simulation Results

Three features of Tables 5–8 deserve comment.

First, the mean squared errors decrease monotonically with increasing sample size. Every one of the sixty MSE sequences, taken across the five sample sizes for each combination of method, parameter, and setting, shows this pattern without exception. For α , and for p away from the boundaries of the unit interval, this is the familiar picture and provides empirical behavior consistent with statistical consistency. The threshold requires separate comment. Its mean squared error also decreases, and does so faster. The rate derived above applies to the maximum likelihood threshold estimator, $\hat{\theta} = X_{(1)}$, for which the mean squared error decreases at the endpoint rate, n^{-2} , rather than the usual n^{-1} ; the four distance and spacings criteria estimate θ freely below the sample minimum, and their rates are not covered by that argument. Because $\hat{\theta}$ does not converge at the ordinary rate, its sequences and those of the regular parameters are not comparable and should not be read as evidence of the same kind. The simulation bears the derived rate out. Regressing the logarithm of the mean squared error of the maximum likelihood estimates of θ on the logarithm of n gives slopes of -2.01 , -1.93 , and -1.95 in settings (i), (ii), and (iii), against the value of -2 predicted above. In setting (iv) the slope is -1.58 , which is the expected consequence of $p = 0.9$ lying close to the boundary value at which the endpoint density vanishes and the rate changes; the finite-sample behavior there is transitional between the two regimes. The tables report them together for compactness, but the interpretation below keeps them apart.

Second, the ranking of the methods depends on which parameter is of interest. Averaging the mean squared error ranks across all cells of the study gives the overall ordering of MLE (1.72), MPSE (2.58), ADE (2.70), WLSE (3.19), and LSE (4.82), with smaller values indicating better performance. Maximum likelihood is therefore the best overall performer, and it generally becomes more competitive as the sample size grows: its mean rank falls from 2.08 at $n = 20$ to 1.42 at $n = 200$. At $n = 500$, it is 1.50, the ranking having effectively stabilized. Even at the smallest sample size, it is ahead of the Anderson–Darling estimator, though only narrowly (mean rank 2.08 against 2.17), the margin being consistent with the known robustness of minimum distance methods in small samples.

Third, and most striking, the picture reverses for the threshold parameter. For θ , the ranking is led by the maximum product of spacings estimator (mean rank: 1.30), ahead of maximum likelihood (2.50). This pattern reflects the boundary behavior of the threshold parameter. The likelihood estimate of θ is pushed toward $x_{(1)}$ and therefore carries a systematic positive bias: averaged over the five sample sizes, it is $+0.010$, $+0.021$, $+0.010$, and $+0.051$ in the four settings, and it is the only method for which the bias in θ is positive in every one of the twenty cells. The maximum product of spacings estimator, which penalizes the boundary by construction, has the smallest bias in absolute value in all four settings: below 0.003 in magnitude in the first three, and a mean absolute bias of approximately 0.006 in the fourth, where the signed values change sign across sample sizes.

The same mechanism explains the one apparent weakness of the maximum product of spacings method. Its estimates of α carry a persistent negative bias, which decays slowly with n ; in setting (iii), for example, the bias falls from -0.485 at $n = 20$ to -0.185 at $n = 200$. Because the criterion refuses to place θ at the sample minimum, it compensates by shrinking α . The trade-off is thus explicit: the method buys an accurate threshold at the cost of a less accurate shape parameter. The behaviour across the five sample sizes bears on whether the bias stabilizes or continues to fall. The extension of the design to $n = 500$ suggests that, at least in this setting, much of the cost is a finite-sample phenomenon rather than evidence of persistent asymptotic bias: the bias falls to -0.017 at $n = 500$, and the mean squared error falls from 0.216 to 0.025, a reduction by a factor of nine relative to $n = 200$.

The decline is therefore faster than linear in n over this range and gives no sign of levelling off at a non-zero value, which is what one would expect of a finite-sample effect rather than an asymptotic one. A bias-corrected version of the estimator, obtained, for instance, by bootstrapping the bias at the observed sample size, would be a natural refinement for samples at the smaller end of this range; we have not pursued it here, since the correction would itself be estimated and the comparison between methods is our object rather than the construction of a new estimator.

A fourth feature concerns the numerical behavior of the criteria rather than the statistical properties of the estimators, and it qualifies the comparison above. Convergence was not universal. Each criterion is optimized from three starting values, and we report the proportion of replicates in which at least one of the three optimizations reported convergence; across the cells of the design, this ranges from 91.9 per cent to 100 per cent, the lower figures occurring for the distance-based criteria at the smallest sample sizes. This figure should be distinguished from the convergence of the solution that is actually retained, which is the one with the smallest criterion value among the three and need not be the one that converged; the Supplementary Results report both proportions. The bias and mean squared error summaries include every replicate that returned a finite estimate, whether or not convergence was confirmed, so the reported performance is that of the procedure as a practitioner would apply it rather than of a subset selected on numerical grounds. Boundary solutions are common in small samples. At $n = 20$, the least squares criterion returns $\hat{p} = 0$ in 29.1 per cent of replicates and $\hat{p} = 1$ in a further 36.0 per cent, so that roughly two thirds of its estimates lie on a boundary of the parameter space; the frequencies fall steadily with n to 2.9 and 7.6 per cent, respectively, at $n = 500$. Maximum likelihood never returns $\hat{p} = 1$, which is structural rather than accidental: the log-likelihood contains the term $\log(1 - p + p\alpha \log(x(i)/\theta))$, which diverges to $-\infty$ at the smallest observation as p approaches one.

Sensitivity to starting values differs sharply between the methods and is reported here because a single-start comparison would have concealed it. Each optimization was run from three dispersed starting values, and the proportion of replicates in which the starts disagreed beyond tolerance was recorded. For maximum likelihood, this proportion falls from 3.7 per cent at $n = 20$ to 0.1 per cent at $n = 500$, the pattern one expects of a well-behaved criterion. For maximum product of spacings, it is 61.5 per cent at $n = 20$ and rises to 88.7 per cent at $n = 500$, indicating a criterion surface with multiple local optima that becomes more, not less, difficult as the sample grows; weighted least squares shows occasional very large spreads in the attained criterion values. Two conclusions follow. Results reported for the distance and spacings criteria from a single starting value should be treated with caution, and the advantage of maximum likelihood recorded above is reinforced once numerical reliability is taken into account alongside bias and mean squared error.

Taken together, the results support a practical recommendation. When interest centers on the tail behavior of the fitted model, and hence on α and p , maximum likelihood is preferable for samples of moderate size and above, with the Anderson–Darling estimator a reasonable alternative for small samples. When the threshold itself is the quantity of interest, or when the fitted distribution function is to be used near its lower endpoint, the maximum product of spacings estimator has the smaller bias, but it should be used only with multiple starting values in view of the sensitivity reported above. The ordinary least squares criterion has the poorest overall mean rank and is the worst of the five in 52 of the 60 cells of the design; it is retained mainly for completeness and for comparability with the existing literature on record-based transmuted models.

6. Applications

In this section, the RBTPD is applied to two data sets and compared with the baseline Pareto distribution and with the transmuted Pareto distribution of [14]. Two comparisons are made. The first is within the Pareto family, where the question at issue is whether record-based transmutation is an effective way of extending the Pareto model; the baseline Pareto and the transmuted Pareto of [14] are the relevant comparators, and because the latter has the same number of parameters as the RBTPD, the comparison between them requires no penalty for model size. The second is against models built on a different baseline, and is included because a three-parameter extension of a two-parameter model cannot fail to improve on it, so that the first comparison alone would establish little. For this purpose, the Burr XII distribution and the exponentiated Pareto distribution of Nadarajah [4] are fitted to both data sets. The record-based transmuted power Lomax distribution of Sakthivel and Nandhini [25] is fitted alongside them, but it belongs to neither group: as explained in the Introduction, up to a location shift it contains the present model as its $\beta = 1$ submodel, so the comparison with it measures what the additional power parameter buys rather than what a different baseline would. Goodness of fit is assessed in absolute, as well as relative, terms, by a bootstrap-calibrated Kolmogorov–Smirnov test.

The first data set consists of the relief times, in hours, of twenty patients receiving an analgesic, reported by [34]. The second consists of the annual flood discharge rates of the Floyd River at James, Iowa, for the years 1935 to 1973, reported by [35]. The second of these was analyzed by [14], so the present results are directly comparable with theirs. Descriptive statistics are given in Table 9.

Table 9. Descriptive statistics of the two data sets.

Data Set	n	Mean	Median	SD	Min	Max
Relief times	20	1.9000	1.7000	0.7041	1.10	4.10
Floyd River	39	6771.1026	3570.0000	11,695.6833	318.00	71,500.00

In each case, the threshold parameter is taken to be the sample minimum, which is the maximum likelihood estimator for the Pareto family and the convention adopted by [14]. The remaining parameters are estimated by maximum likelihood. Tables 10–13 report the estimates, the standard errors obtained from the observed information matrix, interval estimates, and the values of the Kolmogorov–Smirnov statistic, together with the Akaike, corrected Akaike, and Bayesian information criteria. Following Section 4, the standard errors are conditional on $\theta = x_{(1)}$ and are labelled accordingly; no standard error is quoted for the threshold itself, for which the regular information-based theory is not applicable. Interval estimates are obtained by parametric bootstrap with 2000 replications, with the threshold being re-estimated as the bootstrap sample minimum in each replicate so that its sampling variability is carried through rather than conditioned away. For α and p , the percentile interval is reported. For θ , the basic, or reflected, interval is reported instead. The reason for this is inferential rather than cosmetic: since $\hat{\theta} = x_{(1)} \geq \theta$ with probability one, the bootstrap minima all lie above the observed endpoint, so the percentile interval is located on the side of $\hat{\theta}$ on which the parameter cannot lie, whereas the admissible values of θ all lie below it. Reflecting the interval about the estimate places it where the parameter can be. We note that a confidence interval is not required to contain its point estimate, and we do not rest the argument on that fact. For p , a profile-likelihood interval is given as well, because the two can disagree when $\hat{p}$ lies near a boundary and the disagreement is itself informative.

Table 10. Parameter estimates for the relief-time data.

Parameter	Estimate	SE (Conditional)	95% Bootstrap CI	CI Type	95% Profile CI
A	3.6923	0.6815	(1.9898, 5.4793)	percentile	—
Θ	1.1000	—	(0.9153, 1.0980)	basic (reflected)	—
P	0.8238	0.1714	(0.0000, 0.8370)	percentile	(0.1203, 0.9875)

Table 11. Criteria for comparison, relief-time data.

Model	K–S	–2LL	AIC	AICC	BIC	Note
Pareto	0.2363	35.357	39.357	40.063	41.348	
Transmuted Pareto	0.1699	31.855	37.855	39.355	40.843	
Exponentiated Pareto	0.1020	30.817	36.817	38.317	39.805	not identified (ridge)
Burr XII	0.0894	31.064	37.064	38.564	40.051	
RBT power Lomax	0.0978	30.925	38.925	41.592	42.908	$\hat{p} = 1$, on boundary
RBTPD	0.1636	31.343	37.343	38.843	40.330	

Table 12. Parameter estimates for the Floyd River data.

Parameter	Estimate	SE (Conditional)	95% Bootstrap CI	CI Type	95% Profile CI
A	0.7988	0.0942	(0.4994, 1.0112)	percentile	—
Θ	318.0000	—	(81.2898, 313.9157)	basic (reflected)	—
P	0.9366	0.0631	(0.0000, 0.8972)	percentile	(0.7243, 0.9950)

Table 13. Criteria for comparison, Floyd River data.

Model	K–S	–2LL	AIC	AICC	BIC	Note
Pareto	0.3380	785.620	789.620	789.953	792.947	
Transmuted Pareto	0.2351	770.698	776.698	777.384	781.689	
Exponentiated Pareto	0.0862	755.974	761.974	762.659	766.964	not identified (ridge)
Burr XII	0.0724	753.201	759.201	759.887	764.192	
RBT power Lomax	0.0766	753.366	761.366	762.543	768.021	$\hat{p} = 1$, on boundary
RBTPD	0.2306	768.758	774.758	775.444	779.749	

Discussion

Within the Pareto family, the pattern is the same in both applications. The RBTPD attains the smallest value of every criterion considered, followed by the transmuted Pareto distribution and then by the baseline. For the relief-time data, the Kolmogorov–Smirnov statistic falls from 0.2363 for the Pareto model to 0.1699 for the transmuted Pareto and 0.1636 for the RBTPD; the corresponding values are 0.3380, 0.2351, and 0.2306 for the Floyd River data. The improvement over the baseline can also be tested formally, since the Pareto model is the RBTPD with $p = 0$. This null lies on the boundary of the parameter space, so the usual chi-squared limit does not apply; according to the results of Chernoff [36] and Self and Liang [37], the likelihood ratio statistic converges to the equal mixture of a point mass at zero and a chi-squared distribution with one degree of freedom, and the p -value is half the naive one. The statistics are 4.014 and 16.862 for the two data sets, giving boundary-corrected p -values of 0.023 and 2.0×10^{-5} , respectively. One qualification is

needed. The mixture limit is derived for a boundary null in an otherwise regular problem, whereas the present model is non-regular in θ ; the test is applied here, with θ held at $x_{(1)}$, so that the comparison is between two models sharing the same threshold and the irregularity affects both numerator and denominator identically. Subject to that reading, record-based transmutation represents a significant improvement over the baseline in both applications.

Since the transmuted Pareto distribution and the RBTPD have the same number of parameters, the comparison between them is direct and needs no penalty for model size. The RBTPD gives the smaller negative log-likelihood in both cases, by 0.26 and 0.97, respectively, and therefore the smaller value of each information criterion.

The estimated transmutation parameters are 0.8238 and 0.9366 in the two applications, respectively, and are substantially above zero, the value at which the RBTPD reduces to the baseline Pareto distribution. The parameter is confined to the unit interval and admits a direct interpretation as the weight attached to the second upper record value, so the fitted values indicate that both data sets are described better by the more extreme of the two record components than by the baseline model.

For the Floyd River data, the estimates obtained here agree with those reported by [14] to three decimal places, for the Pareto model, as well as for the transmuted Pareto model, which provides an external check on the computations.

Against the wider set of competitors, the position is different, and three points should be recorded. The first two concern the fitted models themselves, and in both cases, the difficulty is one of likelihood geometry rather than of the models as such: a maximum attained on a boundary or along a flat ridge does not satisfy the conditions under which the information criteria are derived, so the corresponding entries in Tables 11 and 13 should be read with caution rather than taken at face value. The exponentiated Pareto distribution cannot be fitted to threshold data in the form in which it is usually written: its density contains the factor $(1 - (\theta/x)^\alpha)^{\beta-1}$, which vanishes at $x = \theta$ whenever $\beta > 1$, so fixing θ at the sample minimum sends the likelihood to $-\infty$ for every $\beta > 1$ and the optimizer collapses to $\beta = 1$, at which point the model is the baseline Pareto and the fitted log-likelihood coincides with it. Allowing θ to be free permits a fit, but the likelihood then possesses a flat ridge along $\beta \rightarrow \infty$, $\theta \rightarrow 0$: on the relief-time data, the value is 15.40883 at $(\beta, \theta) = (1005, 0.279)$ and 15.40872 at $(93,000, 0.091)$. The model is not identified on these data, and its information criteria are reported in Tables 11 and 13 for completeness rather than as a basis for comparison. The record-based transmuted power Lomax distribution attains $\hat{p} = 1.0000$ on both data sets, from every starting value tried, so that its maximum lies on the boundary of the parameter space; the information criteria presume an interior maximum with four free parameters and their reading is correspondingly less straightforward here. Despite the additional parameter, it is beaten on AIC by the RBTPD on the relief-time data.

The third point is that the Burr XII distribution fits both data sets better than the proposed model on the distance and information criteria, and does so stably: three dispersed starting values converge to the same solution in each case. This is reported without qualification, but the size of the difference matters and differs sharply between the two applications. On the relief-time data, the Burr XII distribution leads by only 0.28 in AIC (37.06 against 37.34), and by the same margin in BIC; according to the usual guideline of Burnham and Anderson [38], differences below two indicate that the data give the two models essentially equal support, so that, on the data set for which the RBTPD is adequate, the two are practically indistinguishable. On the Floyd River data, the difference is 15.6 and is substantial, but that is the series for which, as shown below, the RBTPD is not an appropriate model. It is also worth stating what the comparison does and does not show. The Burr XII distribution is supported on the whole positive half-line and contains no

threshold parameter, so it does not provide what the RBTPD is constructed to provide, namely a fitted lower endpoint with a substantive reading and no probability mass below it. The two models answer different questions, and where the threshold is part of the question, the Burr XII distribution is not an alternative. Among the threshold models considered here that yielded a well-identified and stable fit, the RBTPD provides the best fit on both data sets; the exponentiated Pareto also carries a threshold, but its fit is not identified on these data, as noted above.

Relative superiority does not establish that a model fits, and the Kolmogorov–Smirnov statistics in Tables 11 and 13 were therefore calibrated by parametric bootstrap. The asymptotic null distribution is not applicable, because all parameters are estimated from the same data and the threshold in particular is estimated by the sample minimum, which places the fitted distribution function in contact with the smallest observation by construction. Samples were accordingly generated from the fitted model, the parameters re-estimated on each, and the observed statistic referred to the resulting null distribution; the calibration used $B = 2000$ parametric bootstrap replications, the same number as the interval estimates and the value set in the supplied code. For the relief-time data, the statistic is 0.1636 with a bootstrap p -value of 0.207, so the test fails to reject the fitted model at any conventional level. That is the most that such a test can establish: it shows the data to be compatible with the model rather than proving the model correct, and with twenty observations its power is limited in any case. For the Floyd River data, the statistic is 0.2306, and in none of the 2000 bootstrap replications did the simulated statistic reach that value, so the p -value is below 0.0005. The proposed model does not provide an adequate absolute fit to the hydrological series, and we report this rather than confine the assessment to the relative comparison.

Figures 3 and 4 show the fitted densities superimposed on the histogram of each data set, together with the corresponding empirical and fitted cumulative distribution functions. The location of the failure is informative. The largest departure between the empirical and fitted distribution functions occurs at the fifth of the thirty-nine observations, near $x = 1300$, where the empirical value is 0.128 against a fitted 0.333. The largest Kolmogorov–Smirnov departure therefore occurs near the lower end of the distribution, close to the threshold. The tail diagnostics in Figures 5 and 6 also indicate some departure for the most extreme observations, although this is not where the maximum discrepancy between the empirical and fitted distribution functions occurs. Allowing θ to be estimated below the sample minimum does not repair it: the smallest Kolmogorov–Smirnov statistic obtainable across the four alternative criteria is 0.178, still beyond the simulated null distribution. The explanation lies in the data rather than in the numerical procedure. The Floyd River series records annual maxima, not exceedances over a threshold, so its sample minimum is the smallest of thirty-nine annual peaks rather than a physical lower endpoint, and a model whose density is identically zero below that point is constrained in a way the data do not support. The medical data, where the threshold is the minimum time to relief and has a direct clinical reading, present no such difficulty.

The interval estimates in Tables 10 and 12 require one comment. On both data sets the percentile bootstrap interval for p includes zero, while the likelihood ratio test rejects $p = 0$ and the profile-likelihood interval excludes it: the intervals are (0.000, 0.837) and (0.120, 0.987) respectively for the relief-time data, and (0.000, 0.897) and (0.724, 0.995) for the Floyd River data. These are not in conflict. The bootstrap estimate of p is driven to the boundary value zero in 10.1 and 2.8 per cent of replicates respectively, which is the behavior of a constrained estimator in small samples and is what the percentile interval reflects; the profile interval is constructed from the same criterion as the test and so agrees with it in direction, but the two are not equivalent and we do not present them as such. Both are

conditional on $\theta = x_{(1)}$, as the standard errors are, and both rest on the same chi-squared approximation. That approximation is the one justified by Chernoff [36] and Self and Liang [37] for a boundary null in an otherwise regular problem, and the present model is non-regular in θ ; we apply it with θ held fixed, so that the two hypotheses compared share the same threshold and the irregularity enters both in the same way, but we have not derived the limiting distribution for this model and the resulting inference about p should be read as approximate rather than exact. Both intervals are reported, with the distinction made explicit, rather than selecting whichever is the more favorable.

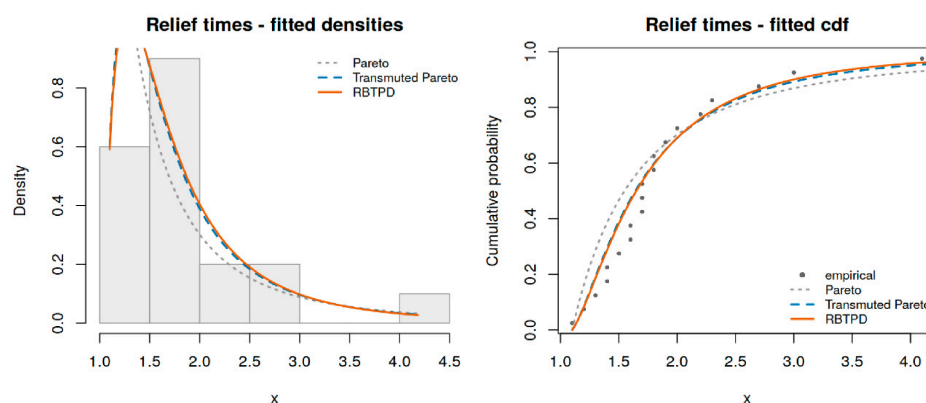


Figure 3. Fitted density and empirical/fitted cumulative distribution functions for the relief-time data.

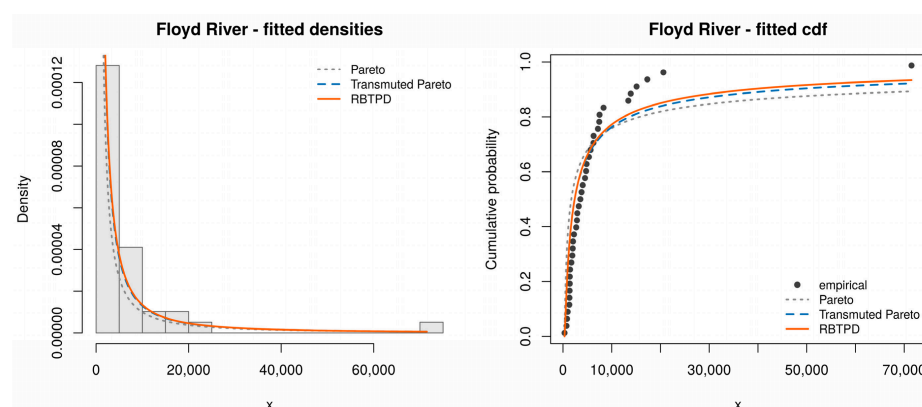


Figure 4. Fitted density and empirical/fitted cumulative distribution functions for the Floyd River data.

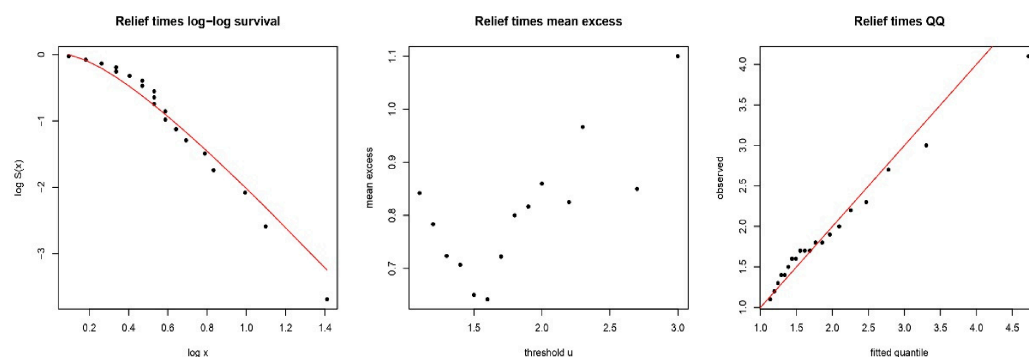


Figure 5. Tail diagnostics for the relief-time data: Empirical and fitted survival functions on a double logarithmic scale (left), empirical mean excess function (center), and quantile–quantile plot against the fitted RBTPD (right).

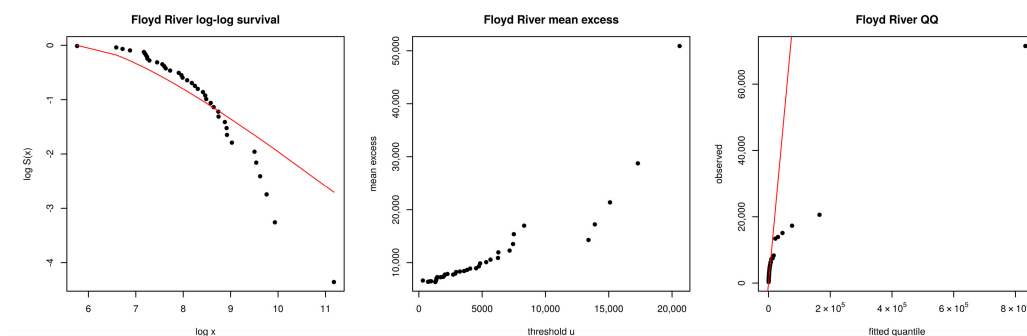


Figure 6. Tail diagnostics for the Floyd River data: Empirical and fitted survival functions on a double logarithmic scale (**left**), empirical mean excess function (**center**), and quantile–quantile plot against the fitted RBTPD (**right**).

Finally, the fitted shape parameter for the Floyd River data is $\hat{\alpha} = 0.7988$, below unity, so that, under the fitted model, neither the mean nor any higher moment exists. This is a property of the fitted model rather than an established property of the data themselves, and the distinction matters here because the goodness-of-fit assessment rejects that model for this series. What can be said is that every Pareto-type model fitted to these data places itself in the same position, which reflects how heavy the observed tail is: the fitted Pareto and transmuted Pareto models give $\hat{\alpha} = 0.4125$ and 0.5859 , more extreme values still. The practical implication is that location and dispersion should be summarized for such data by the median and by the quantile-based measures of Section 3.1 rather than by the mean and variance, which is the reason those measures were derived. Under the fitted RBTPD, therefore, the sample mean reported for this series in Table 9 should be regarded only as a descriptive statistic of the observed record, and not as an estimator of a finite model-based population mean.

Figures 5 and 6 report the tail diagnostics for the two data sets: the empirical survival function on a double logarithmic scale with the fitted survival function superimposed, the empirical mean excess function, and a quantile–quantile plot of the observations against the fitted model. Together they locate the departures discussed above: for the Floyd River series, the principal departure lies near the threshold, with some further deviation visible at the most extreme observations, while for the relief-time data, the departures are slight throughout.

7. Conclusions

This paper has introduced the record-based transmuted Pareto distribution (RBTPD), obtained by applying the record-based transmutation map of [18] to the two-parameter Pareto distribution. The resulting model has three parameters, a distribution function, and a density available in closed form and reduces to the baseline Pareto distribution when the transmutation parameter is zero.

A range of statistical properties has been derived. The quantile function was obtained in terms of the lower branch of the Lambert W function; the raw and incomplete moments, the mean deviations and the Bonferroni and Lorenz curves were given in closed form, subject to the moment condition inherited from the baseline; the densities of the order statistics and of the upper record values were obtained; and the Renyi entropy was expressed in closed form in terms of the upper incomplete gamma function.

Five estimation methods were compared in a Monte Carlo study based on twenty thousand generated samples. For all five estimators, the mean squared errors decreased as the sample size increased. For α , and for p away from the boundaries of its range, this provides empirical evidence of improved accuracy with larger samples and a pattern

consistent with statistical consistency; for the threshold, which is a support-boundary parameter estimated at a non-standard rate, the decrease is not evidence of the same kind and is interpreted separately. Maximum likelihood showed the best overall performance, with its relative advantage becoming more pronounced as the sample size increased, while the Anderson–Darling estimator showed closely comparable performance in the smallest samples. For the threshold parameter, however, the maximum product of spacings estimator performed particularly well, exhibiting the lowest average mean squared error rank and consistently small bias across the parameter settings considered. These findings indicate that the relative performance of the estimation methods depends on the parameter of primary interest: maximum likelihood is generally preferable for estimating the shape and transmutation parameters, whereas maximum product of spacings may be advantageous when accurate estimation of the threshold parameter is of particular importance. The numerical diagnostics qualify this. Convergence was not universal; boundary solutions are frequent in small samples; and the maximum product of spacings criterion proved markedly sensitive to starting values, increasingly so as the sample size grew. Maximum likelihood was by a wide margin the most stable of the five, and results obtained from the distance and spacings criteria with a single starting value should be treated with caution.

The model was fitted to two data sets and compared with the Pareto distribution and with the transmuted Pareto distribution of [14]. Within that family, the pattern was the same in both cases: the proposed model gave the smallest value of every criterion considered, followed by the transmuted Pareto distribution and then by the baseline, and a likelihood ratio test with the appropriate boundary correction rejected the baseline in both applications. Since the transmuted Pareto distribution and the model proposed here have the same number of parameters, the comparison between them requires no penalty for model size, and it indicates that record-based transmutation is at least as effective as quadratic rank transmutation for these data. The model was also compared with the Burr XII and exponentiated Pareto distributions, which are built on different baselines, and with the record-based transmuted power Lomax distribution, which is a close relative rather than an outside comparator. The exponentiated Pareto and the record-based transmuted power Lomax attained their maxima along a flat ridge and on a boundary of the parameter space respectively, so that standard information-criterion comparisons are less straightforward for them and their entries should be interpreted cautiously. The Burr XII distribution fitted both data sets better than the proposed model; it carries no threshold parameter, however, and therefore it does not serve the purpose for which the present model is intended. The estimates obtained for the Floyd River data agree to three decimal places with those published by [14], which provides an external check on the computations.

The two applications span the settings in which Pareto-type models are most often used. The relief-time data are biomedical: they record the time to response following treatment, a quantity whose distribution determines the hazard of remaining without relief and is therefore of direct clinical relevance. The Floyd River series belongs to the hydrological setting in which the Pareto model originates and was chosen because it permits a direct comparison with the transmuted Pareto distribution on data used by its authors.

Goodness of fit was also assessed in absolute terms, by a bootstrap-calibrated Kolmogorov–Smirnov test, and the two applications gave different answers. The model is adequate for the relief-time data but is rejected for the Floyd River series, where the largest discrepancy is concentrated near the threshold, although the tail diagnostics also show some departure at the most extreme observations. This delimits the conditions under which the model is appropriate, and we state them explicitly. The RBTPD is suited to data possessing a genuine lower endpoint, whether physical, clinical or contractual, below which observations cannot fall and which is itself of interest: minimum times to response,

exceedances over a fixed threshold, losses above a deductible. It is not suited to series whose smallest observation is merely the least of the values happening to be observed, such as the annual maxima considered here, for which a model supported on the whole positive half-line will generally be preferable. The value of the threshold parameter and the cost of imposing it are two aspects of the same feature of the model, and the applications illustrate both.

Three limitations should be recorded. First, the moments of the model exist only up to order less than the shape parameter, a restriction inherited unchanged from the Pareto baseline, so that, for heavy-tailed data, the quantile-based measures of shape reported in Section 3.1 may be the more useful description. Second, the threshold parameter is a lower endpoint below which the density is zero, and the model is therefore best suited to data whose smallest observation is close to the bulk of the sample; where the threshold lies far below the mass of the data, the fit of every Pareto-type model is correspondingly constrained. Third, the model is non-regular in the threshold: the regular Fisher-information theory is not applicable to that parameter and, for $p < 1$, its estimator converges at rate n with an exponential rather than a normal limit, the boundary case $p = 1$ being slower still at rate $n^{1/2}$. Standard regular asymptotic inference is therefore available for α and for p in the interior of the unit interval, whereas the threshold requires non-standard inference and is handled here by resampling. The same feature that gives the parameter its interpretation removes it from the scope of the usual theory.

Several extensions suggest themselves. Bayesian estimation of the three parameters, with a prior on the unit interval for the transmutation parameter, would be a natural complement to the classical methods considered here. The model could be studied under censored sampling and under ranked set sampling, both of which are common in reliability applications. A regression formulation in which the transmutation parameter depends on covariates would allow the degree of departure from the baseline to vary across observations. Finally, the same map could be applied to other heavy-tailed baselines, and the dual construction based on lower record values remains to be developed for the Pareto family.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/axioms15110793/s1>, The Supplementary Material consists of three files: `rbtpd_revision.R`, the complete R code reproducing the simulation study, the applications, the bootstrap procedures and all tables and figures, with both data sets written directly into the script and the random seeds set at its head; `rbtpd_results.xlsx`, the full numerical results, including the Monte Carlo standard errors, the convergence rates, the boundary-solution frequencies and the starting-value diagnostics summarized in Section 5; and `session_info.txt`, recording the R version and the version of every package used.

Author Contributions: Conceptualization, Ö.A., İ.A., and Y.M.K.; methodology, Ö.A., İ.A., and Y.M.K.; software, Ö.A. and İ.A.; validation, Ö.A., İ.A., and Y.M.K.; formal analysis, Ö.A. and İ.A.; investigation, Ö.A.; data curation, Ö.A. and İ.A.; writing—original draft preparation, Ö.A. and İ.A.; writing—review and editing, İ.A. and Y.M.K.; visualization, Ö.A.; supervision, İ.A. and Y.M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The two data sets analyzed in this study are taken from the published sources cited in Section 6. The Supplementary Material accompanying this article contains three files. The first is the complete R code reproducing the simulation study, the applications, the bootstrap procedures, every table and every figure; both data sets are written directly into that script, so no

external data file has to be imported and no locale-dependent decimal separator is involved, the random seeds are set at the head of the script, and step-by-step instructions for running the analyses, with approximate running times, are given at the end of it. The second is a spreadsheet holding the full numerical results, including the Monte Carlo standard errors, the convergence rates, the boundary-solution frequencies and the starting-value diagnostics summarized in Section 5. The third is a session_info.txt file recording the R version and the version of every package used.

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Anthropic; Opus 5 and earlier versions, 2025–2026), ChatGPT (OpenAI, GPT-5.6 Thinking) and Gemini (Google, Gemini 2.5 Pro) for literature searching, for language editing of the manuscript text, for assistance in deriving and algebraically checking the analytical expressions reported in Sections 2 and 3, and for assistance in writing and debugging the R code used in the simulation study of Section 5 and the applications of Section 6. All analytical results were checked numerically against independent implementations, and the authors have reviewed and edited the output and take full responsibility for the content.

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

References

1. Pareto, V. *Cours d'Économie Politique*; F. Rouge: Lausanne, Switzerland, 1897.
2. Balkema, A.A.; de Haan, L. Residual life time at great age. *Ann. Probab.* **1974**, *2*, 792–804. [\[CrossRef\]](#)
3. Pickands, J. Statistical inference using extreme order statistics. *Ann. Stat.* **1975**, *3*, 119–131. [\[CrossRef\]](#)
4. Nadarajah, S. Exponentiated Pareto distributions. *Statistics* **2005**, *39*, 255–260. [\[CrossRef\]](#)
5. Akinsete, A.; Famoye, F.; Lee, C. The beta-Pareto distribution. *Statistics* **2008**, *42*, 547–563. [\[CrossRef\]](#)
6. Alzaatreh, A.; Famoye, F.; Lee, C. Gamma-Pareto distribution and its applications. *J. Mod. Appl. Stat. Methods* **2012**, *11*, 78–94. [\[CrossRef\]](#)
7. Bourguignon, M.; Silva, R.B.; Zea, L.M.; Cordeiro, G.M. The Kumaraswamy Pareto distribution. *J. Stat. Theory Appl.* **2013**, *12*, 129–144. [\[CrossRef\]](#)
8. Ihtisham, S.; Khalil, A.; Manzoor, S.; Khan, S.A.; Ali, A. Alpha-Power Pareto distribution: Its properties and applications. *PLoS ONE* **2019**, *14*, e0218027. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Alizadeh, M.; Cordeiro, G.M.; Nascimento, A.D.C.; Lima, M.C.S.; Ortega, E.M.M. Odd-Burr generalized family of distributions with some applications. *J. Stat. Comput. Simul.* **2017**, *87*, 367–389. [\[CrossRef\]](#)
10. Arik, I. New Distribution Families and Their Statistical Properties for Survival Analysis. Ph.D. Thesis, Anadolu University, Eskisehir, Türkiye, 2018.
11. Arik, I.; Kantar, Y.M. New odd Burr-Rayleigh distribution: Theory and applications. *Far East J. Theor. Stat.* **2019**, *55*, 53–82. [\[CrossRef\]](#)
12. Arik, I.; Kantar, Y.M. A new odd-Burr Pareto distribution: Statistical properties, estimation, and applications. *Pak. J. Stat. Oper. Res.* **2025**, *21*, 375–394. [\[CrossRef\]](#)
13. Shaw, W.T.; Buckley, I.R.C. The alchemy of probability distributions: Beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv* **2009**, arXiv:0901.0434.
14. Merovci, F.; Puka, L. Transmuted Pareto distribution. *ProbStat Forum* **2014**, *7*, 1–11.
15. Chhetri, S.B.; Long, H.; Aryal, G. The beta transmuted Pareto distribution: Theory and applications. *J. Stat. Appl. Probab.* **2017**, *6*, 243–258. [\[CrossRef\]](#)
16. Chhetri, S.B.; Akinsete, A.A.; Aryal, G.; Long, H. The Kumaraswamy transmuted Pareto distribution. *J. Stat. Distrib. Appl.* **2017**, *4*, 11. [\[CrossRef\]](#)
17. Rahman, M.M.; Al-Zahrani, B.; Shahbaz, M.Q. Cubic transmuted Pareto distribution. *Ann. Data Sci.* **2020**, *7*, 91–108. [\[CrossRef\]](#)
18. Balakrishnan, N.; He, M. A record-based transmuted family of distributions. In *Advances in Statistics—Theory and Applications*; Ghosh, I., Balakrishnan, N., Ng, H.K.T., Eds.; Springer: Cham, Switzerland, 2021; pp. 3–24.
19. Tanış, C.; Saraçoğlu, B. On the record-based transmuted model of Balakrishnan and He based on Weibull distribution. *Commun. Stat. Simul. Comput.* **2022**, *51*, 4204–4224. [\[CrossRef\]](#)
20. Arshad, M.; Khetan, M.; Kumar, V.; Pathak, A.K. Record-based transmuted generalized linear exponential distribution with increasing, decreasing and bathtub shaped failure rates. *Commun. Stat. Simul. Comput.* **2024**, *53*, 3489–3513. [\[CrossRef\]](#)
21. Tanış, C. A new Lindley distribution: Applications to COVID-19 patients data. *Soft Comput.* **2024**, *28*, 2863–2874. [\[CrossRef\]](#)
22. Pathak, A.K.; Arshad, M.; Pandey, A.K.; Ali, A. Record-based transmuted unit-omega distribution: Different methods of estimation and applications. *arXiv* **2024**, arXiv:2405.07958.

23. Alrweili, H. Statistical inference and data analysis of the record-based transmuted Burr X model. *Open Math.* **2025**, *23*, 20240121. [[CrossRef](#)]
24. Tanış, C. Record-based transmuted log-logistic distribution: Properties, simulation, and applications to petroleum rock and reactor pump data. *arXiv* **2025**, arXiv:2509.09757.
25. Sakthivel, K.M.; Nandhini, V. Record-based transmuted power Lomax distribution: Properties and its applications in reliability. *Reliab. Theory Appl.* **2022**, *17*, 574–592.
26. Arnold, B.C.; Balakrishnan, N.; Nagaraja, H.N. *Records*; Wiley: New York, NY, USA, 1998.
27. Bowley, A.L. *Elements of Statistics*, 4th ed.; Charles Scribner's Sons: New York, NY, USA, 1920.
28. Moors, J.J.A. A quantile alternative for kurtosis. *J. R. Stat. Soc. Ser. D* **1988**, *37*, 25–32. [[CrossRef](#)]
29. Renyi, A. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
30. Swain, J.J.; Venkatraman, S.; Wilson, J.R. Least-squares estimation of distribution functions in Johnson's translation system. *J. Stat. Comput. Simul.* **1988**, *29*, 271–297. [[CrossRef](#)]
31. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Stat.* **1952**, *23*, 193–212. [[CrossRef](#)]
32. Cheng, R.C.H.; Amin, N.A.K. Estimating parameters in continuous univariate distributions with a shifted origin. *J. R. Stat. Soc. Ser. B* **1983**, *45*, 394–403. [[CrossRef](#)]
33. Ranneby, B. The maximum spacing method: An estimation method related to the maximum likelihood method. *Scand. J. Stat.* **1984**, *11*, 93–112.
34. Gross, A.J.; Clark, V.A. *Survival Distributions: Reliability Applications in the Biomedical Sciences*; Wiley: New York, NY, USA, 1975.
35. Mudholkar, G.S.; Hutson, A.D. The exponentiated Weibull family: Some properties and a flood data application. *Commun. Stat. Theory Methods* **1996**, *25*, 3059–3083. [[CrossRef](#)]
36. Chernoff, H. On the distribution of the likelihood ratio. *Ann. Math. Stat.* **1954**, *25*, 573–578. [[CrossRef](#)]
37. Self, S.G.; Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **1987**, *82*, 605–610. [[CrossRef](#)]
38. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer: New York, NY, USA, 2002.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.