Estimating the Parameters of a Tapered Pareto Distribution

M. Vaičiulis\textsuperscript{1}\textsuperscript{*} and N. M. Markovich\textsuperscript{2}\textsuperscript{**}

\textsuperscript{1}Institute of Data Science and Digital Technologies, Vilnius University, Vilnius, 2600 Lithuania
\textsuperscript{2}Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, 117997 Russia

\textsuperscript{*}marijus.vaiciulis@mif.vu.lt, \textsuperscript{**}nat.markovich@gmail.com

Received November 19, 2020; revised February 26, 2021; accepted March 16, 2021

Abstract—The article deals with the problem of estimating the parameters of a tapered Pareto distribution. Using the moment method, we obtain new estimates depending on an additional parameter. We prove that the joint asymptotic distribution of these estimates is Gaussian. A procedure is proposed that permits one to choose the additional parameter in an optimal way. The new estimates are compared with the corresponding maximum likelihood estimates. By way of example, an application of the new estimates to the COVID-19 incidence data is given. A new algorithm for a random variable generator with a tapered Pareto distribution is proposed.

Keywords: tapered Pareto distribution, parameter estimation, moment method estimates, maximum likelihood estimates, COVID-19 incidence

DOI: 10.1134/S000511792108004X

1. INTRODUCTION

A random variable $X$ is said to be distributed according to the (exponentially) tapered Pareto distribution if it has the cumulative distribution function

$$F_{a,\beta,\theta}(x) = 1 - \left(\frac{a}{x}\right)^\beta \exp\left\{\frac{a - x}{\theta}\right\}, \quad x \geq a,$$

(1)

where $a > 0$ is the minimum value of $X$, $\beta > 0$ is the shape parameter, and $\theta \geq a$ is the upper cutoff parameter. The probability distribution density of this distribution has the form

$$f_{a,\beta,\theta}(x) = \left(\frac{\beta}{x} + \frac{1}{\theta}\right) \left(\frac{a}{x}\right)^\beta \exp\left\{\frac{a - x}{\theta}\right\}, \quad x \geq a.$$

The tapered Pareto distribution, also known as the modified Gutenberg– Richter law, was first proposed by Pareto [1]. Vere-Jones et al. [2] noticed the following property of the tapered Pareto distribution. Let a random variable $\chi_1$ have a Pareto distribution with the distribution function $F_{a,\beta}(x) = 1 - (a/x)^\beta, \ x \geq a$, and let $\chi_2$ be a random variable having an exponential distribution with the distribution function $F_0(x) = 1 - \exp\{x/\theta\}, \ x \geq 0$, with $\chi_1$ and $\chi_2$ being independent random variables. Then the random variable

$$X = \min\{\chi_1, \chi_2 + a\}$$

(2)

has the distribution function $F_{a,\beta,\theta}(x), \ x \geq a$. This fact is used when modeling the tapered Pareto distribution; see [2–4].
The following definition is based on the violation of the Cramér condition for heavy-tailed distributions.

**Definition 1** [5]. A distribution function $F$ is said to have a heavy tail (on the right) if

$$
\int_{-\infty}^{+\infty} e^{tx} F(dx) = \infty
$$

for all $t > 0$. If $F$ does not satisfy condition (3), then it is said to have a light tail.

For example, the Pareto distribution function $F_{a,\beta}(x), x \geq a$, has a heavy tail, while the distribution function of the exponential law has a light tail. The distribution function $F_{a,\beta,\theta}(x), x \geq a$, has a light tail. Indeed, in this case for $0 < t < 1/\theta$ we have

$$
\mathbb{E}e^{tx} = \frac{e^{at}}{a(1-t\theta)} + \beta \left(1 - \frac{1}{a(1-t\theta)}\right) \left(\frac{a(1-t\theta)}{\theta}\right)^\beta \exp\left\{at + \frac{a(1-t\theta)}{\theta}\right\} \Gamma\left(-\beta, \frac{a(1-t\theta)}{\theta}\right),
$$

where $\Gamma(-\beta, x) = \int_x^{\infty} t^{-\beta-1} e^{-t} dt$ is the incomplete gamma function. Here and in the following, $\mathbb{E}$ stands for the expectation. Recall that for $x > 0$ the last integral converges for all real $\beta$.

The tapered Pareto distribution is used to model the magnitude of earthquakes in seismology [3, 4] and wildfire sizes [6, 7]. It was established in [8] that the tapered Pareto distribution well describes the sizes of trees. Some authors support the idea that financial data follows this distribution; see [9]. The present paper is apparently the first to apply the tapered Pareto distribution to COVID-19 incidence data for various countries.

By $(X_1, \ldots, X_n)$ we denote a random sample, where $X_1, \ldots, X_n$ are independent instances of a random variable $X$ with distribution function $F_{a,\beta,\theta}(x), x \geq a$.

The maximum likelihood estimates $\hat{\beta}_n^{(1)}$ and $\hat{\lambda}_n^{(1)}$ of the parameters $\beta$ and $\lambda = a/\theta$ are a solution of the equations

$$
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 - \hat{\lambda}_n^{(1)}} \left(\frac{X_i}{\hat{a}_n} - 1 - \hat{b}_n \left(\frac{X_i}{\hat{a}_n}\right)\right) = 1,
$$

(4)

$$
\hat{b}_n \hat{\beta}_n^{(1)} + \left(\frac{X_n}{\hat{a}_n} - 1\right) \hat{\lambda}_n^{(1)} = 1,
$$

where

$$
\hat{a}_n = \min\{X_1, \ldots, X_n\},
$$

$$
\hat{b}_n = \frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{X_i}{\hat{a}_n}\right),
$$

(5)

$$
\hat{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.
$$

The nonlinear equation (4) can only be solved numerically. This can be done, for example, by the Newton–Raphson method with the initial point $\hat{X}_n - \hat{a}_n$; see [2, 4].

Set

$$
S_{11}(\beta, \lambda) = \mathbb{E}\left(\frac{X^2}{(\beta + \lambda X)^2}\right),
$$

$$
S_{22}(\beta, \lambda) = \mathbb{E}\left(\frac{1}{(\beta + \lambda X)^2}\right),
$$

$$
S_{12}(\beta, \lambda) = \mathbb{E}\left(\frac{X}{(\beta + \lambda X)^2}\right),
$$
where the random variable $\chi$ has the distribution function $F_{1,\beta,1/\lambda}$. The following theorem puts together the known results on the joint asymptotic normality of $\hat{\beta}_n^{(1)}$ and $\hat{\lambda}_n^{(1)}$. In particular, the case where $a$ is known was discussed in [3]. In the case of unknown $a$, the result below is a straightforward consequence of Theorem 2.4.(b) and Lemma 2.6.(a) in [9].

**Theorem 1.** Let $(X_1, \ldots, X_n)$ be a sample with distribution function $F_{a,\beta,a/\lambda}$. Assume that the parameter $a$ is known and estimated by formula (5). Then

$$\sqrt{n} \left( \hat{\beta}_n^{(1)} - \beta, \hat{\lambda}_n^{(1)} - \lambda \right) \xrightarrow{d} (\xi_1, \xi_2), \quad n \to \infty,$$

where $\xrightarrow{d}$ means the convergence in distribution. Here $(\xi_1, \xi_2)$ is a Gaussian vector with $\mathbb{E}(\xi_1) = \mathbb{E}(\xi_2) = 0$, $\mathbb{E}(\xi_1^2) = S_1^2(\beta, \lambda)$, $\mathbb{E}(\xi_2^2) = S_2^2(\beta, \lambda)$, and $\mathbb{E}(\xi_1 \xi_2) = S_{12}(\beta, \lambda)$, where

$$S_1^2(\beta, \lambda) = \frac{s_{11}(\beta, \lambda)}{s_{11}(\beta, \lambda)s_{22}(\beta, \lambda) - s_{12}^2(\beta, \lambda)},$$

$$S_2^2(\beta, \lambda) = \frac{s_{22}(\beta, \lambda)}{s_{11}(\beta, \lambda)s_{22}(\beta, \lambda) - s_{12}^2(\beta, \lambda)},$$

$$S_{12}(\beta, \lambda) = - \frac{s_{12}(\beta, \lambda)}{s_{11}(\beta, \lambda)s_{22}(\beta, \lambda) - s_{12}^2(\beta, \lambda)}.$$

The maximal likelihood estimates $\hat{\beta}_n^{(1)}$ and $\hat{\lambda}_n^{(1)}$ have several drawbacks. First, as was mentioned above, the solution of Eq. (4) can only be obtained numerically. Moreover, it mainly depends only on several largest observations in the sample, see [4]. Second, left-truncated samples are often used in statistical practice, and the maximal likelihood estimate for the parameter $\theta = a/\lambda$ is significantly biased for such samples; see [4] for more details. Therefore, we need estimates constructed by other methods for the parameters $\beta$ and $\lambda$.

The aim of the present paper is to propose new estimates for the parameters of the tapered Pareto distribution, give a theoretical study of these estimates, and apply this distribution to COVID-19 incidence data modeling.

The article is organized as follows. Section 2 introduces estimates for the parameters $\beta$ and $\lambda$ of the tapered Pareto distribution using the moment method. The main result is presented, which asserts that these estimates for $\beta$ and $\lambda$ have a joint asymptotically normal distribution. Section 3 discusses the modeling of a random variable with a tapered Pareto distribution. In Sec. 4, the newly introduced estimates are compared with the estimates obtained by the maximal likelihood method. Section 5 contains application of the results to the COVID-19 incidence data. Conclusions are stated in Sec. 6. The proofs are placed in the Appendix.

2. ESTIMATION OF THE PARAMETERS OF THE TAPERED PARETO DISTRIBUTION BY THE MOMENT METHOD

Recall that the one-parameter Box–Cox transform with parameter $r \in \mathbb{R}$ is defined as

$$h_r(x) = \begin{cases} 
\ln(x), & r = 0 \\
(x^r - 1)/r, & r \neq 0.
\end{cases}$$

Set

$$\nu_r = \mathbb{E}(h_r(X/a)), \quad (6)$$

where $X$ is a random variable with distribution function $F_{a,\beta,a/\lambda}$. It can readily be verified that

$$\nu_r = \lambda^{\beta-r}e^{\lambda \Gamma(r - \beta, \lambda)}, \quad r \in \mathbb{R}. \quad (7)$$
The incomplete gamma function satisfies the property
\[ \Gamma (r - \beta + 1, \lambda) = \lambda^{-\beta}e^{-\lambda} + (r - \beta)\Gamma (r - \beta, \lambda); \]  
(8)
see, e.g., [10]. Combining (7) and (8), we obtain
\[ \Lambda_r(\beta, \lambda) = 0, \]  
(9)
where \( \Lambda_r(\beta, \lambda) = \lambda\nu_{r+1} - (r - \beta)\nu_r - 1 \). The system of equations \( \Lambda_r(\beta, \lambda) = 0, \Lambda_{r+1}(\beta, \lambda) = 0 \) can be represented as
\[
\begin{align*}
\lambda\nu_{r+1} + \beta\nu_r &= 1 + r\nu_r, \\
\lambda\nu_{r+2} + \beta\nu_{r+1} &= 1 + (1 + r)\nu_{r+1}.
\end{align*}
\]
By Theorem 4 in [11], the inequality \( \nu_{r+1}^2 - \nu_r\nu_{r+2} < 0 \) is satisfied for all \( r \in \mathbb{R}, \beta > 0 \), and \( \lambda > 0 \). Therefore, the last system of equations for \( \beta \) and \( \lambda \) has the unique solution
\[
\beta = r - \frac{g^{(1)}(\nu_{r+1}, \nu_{r+2})}{g^{(3)}(\nu_r, \nu_{r+1}, \nu_{r+2})}, \quad \lambda = \frac{g^{(2)}(\nu_r, \nu_{r+1})}{g^{(3)}(\nu_r, \nu_{r+1}, \nu_{r+2})},
\]  
(10)
where
\[
g^{(1)}(x_2, x_3) = x_2 - x_3 + x_2^2, \quad g^{(2)}(x_1, x_2) = x_1 - x_2 + x_1x_2, \quad g^{(3)}(x_1, x_2, x_3) = x_1x_3 - x_2^2, \quad (x_1, x_2, x_3) \in \mathbb{R}^3.
\]
Replacing \( \nu_r, \nu_{r+1}, \) and \( \nu_{r+2} \) in (10) by their sample counterparts \( \hat{\nu}_{n,r}, \hat{\nu}_{n,r+1}, \) and \( \hat{\nu}_{n,r+2}, \) where
\[
\hat{\nu}_{n,r} = \frac{1}{n} \sum_{i=1}^{n} h_r \left( \frac{X_i}{a_n} \right)
\]
and \( (X_1, \ldots, X_n) \) is a random sample with distribution function \( F_{a, \beta, a/\lambda} \), we introduce a new family of estimates for the pair of parameters \( (\beta, \lambda) \),
\[
\hat{\beta}^{(2)}_{n,r} = r - \frac{g^{(1)}(\hat{\nu}_{n,r+1}, \hat{\nu}_{n,r+2})}{g^{(3)}(\hat{\nu}_{n,r}, \hat{\nu}_{n,r+1}, \hat{\nu}_{n,r+2})}, \quad \hat{\lambda}^{(2)}_{n,r} = \frac{g^{(2)}(\hat{\nu}_{n,r}, \hat{\nu}_{n,r+1})}{g^{(3)}(\hat{\nu}_{n,r}, \hat{\nu}_{n,r+1}, \hat{\nu}_{n,r+2})}.
\]  
(14)
(15)
The proposed estimates \( \hat{\beta}^{(2)}_{n,r} \) and \( \hat{\lambda}^{(2)}_{n,r} \) depend on the Box–Cox parameter \( r \). The optimal choice of this parameter is obtained in Sec. 4.

Before we state the main result, let us introduce the notation
\[
c_{0,r} = \lambda\nu_{r+3} - 2\nu_{r+2} - 1, \quad c_{1,r} = 1 - 2\lambda\nu_{r+2}, \quad c_{2,r} := \lambda\nu_{r+1},
\]
\[
c_{3,r} = -\lambda\nu_{r+2} + \nu_{r+1} + 1, \quad c_{4,r} = 2\lambda\nu_{r+1} + \nu_r - 1, \quad c_{5,r} = -\lambda\nu_r.
\]
Let \( Y_t, t \in \mathbb{R} \), be a Gaussian process with zero mean and covariance function
\[
\rho_{s,t} = \mathbb{E} (Y_sY_t) = \text{Cov}(h_s(\chi), h_t(\chi)), \]  
(16)
where the random variable \( \chi \) is the same as in Theorem 1. In the case of \( st \neq 0 \), the covariances \( \rho_{s,t} \) can be represented as follows:
\[
\rho_{s,t} = \frac{\nu_{s+t} - \nu_t}{s} + \frac{\nu_{s+t} - \nu_s}{t} - \nu_s\nu_t.
\]  
(17)
The covariances \( \rho_{0,t}, t \in \mathbb{R} \), can be expressed in terms of the Meijer G-function (see the definition in [12]).

**Theorem 2.** Let \( X_1, \ldots, X_n \) be a random sample with distribution function \( F_{a,\beta,a/\lambda} \), and let \( r \in \mathbb{R} \). Then

\[
\sqrt{n} \left( \hat{\beta}_{n,r}^{(2)} - \beta, \hat{\lambda}_{n,r}^{(2)} - \lambda \right) \xrightarrow{d} (\eta_1, r, \eta_2, r), \quad n \to \infty,
\]

where \((\eta_1, r, \eta_2, r)\) is the Gaussian vector with zero means, the variances \( \mathbb{E}(\eta_1^2, r) = \sigma_{1,r}^2(\beta, \lambda) \) and \( \mathbb{E}(\eta_2^2, r) = \sigma_{2,r}^2(\beta, \lambda) \), and the covariance \( \mathbb{E}(\eta_1, r, \eta_2, r) = \sigma_{12,r}(\beta, \lambda) \). Here

\[
\sigma_{1,r}^2(\beta, \lambda) = \frac{1}{(\nu_r \nu_{r+1} - \nu_{r+2}^2)^2} \sum_{i,j=0}^2 c_{i,r} c_{j,r} \rho_{r+1,r+j}, \quad (19)
\]

\[
\sigma_{2,r}^2(\beta, \lambda) = \frac{1}{(\nu_r \nu_{r+1} - \nu_{r+2}^2)^2} \sum_{i,j=0}^2 c_{i+3,r} c_{j+3,r} \rho_{r+1,r+j}, \quad (20)
\]

\[
\sigma_{12,r}(\beta, \lambda) = \frac{1}{(\nu_r \nu_{r+1} - \nu_{r+2}^2)^2} \sum_{i,j=0}^2 c_{i,r} c_{j+3,r} \rho_{r+1,r+j}. \quad (21)
\]

If the parameter \( a \) is known, then we can apply an obvious modification of the estimates \( \hat{\beta}_{n,r}^{(2)} \) and \( \hat{\lambda}_{n,r}^{(2)} \), for which relation (18) remains true.

Consider the estimation of the cutoff parameter \( \theta \) in the case where the other parameters \( a \) and \( \beta \) are known. Combining \( \theta = a/\lambda \) and (9), we obtain

\[
\theta = \frac{a \nu_{r+1}}{1 + (r - \beta) \nu_r}.
\]

Replacing \( \nu_r \) and \( \nu_{r+1} \) by \( \tilde{\nu}_n, r \) and \( \tilde{\nu}_n, r+1 \), respectively, we obtain the estimate

\[
\hat{\theta}_{n,r} = \frac{a \tilde{\nu}_n, r+1}{1 + (r - \beta) \tilde{\nu}_n, r}.
\]

Note that the estimate \( \hat{\theta}_{n,1} \) coincides with the estimate of the parameter \( \theta \) introduced in [4]. Let us generalize Theorem 1 in [4].

**Theorem 3.** Let \( X_1, \ldots, X_n \) be a random sample with distribution function \( F_{a,\beta,\theta} \), where the parameters \( a \) and \( \beta \) are known, and let \( r \in \mathbb{R} \). Then

\[
\sqrt{n} \left( \hat{\theta}_{n,r} - \theta \right) \xrightarrow{d} \mathcal{N}(0, \sigma_r^2), \quad n \to \infty,
\]

where

\[
\sigma_r^2 = \frac{\theta^2}{\nu_{r+1}^2} \left( \frac{\theta^2(r - \beta)}{a^2} \rho_{r,r} + \frac{2\theta(r - \beta)}{a} \rho_{r,r+1} + \rho_{r+1,r+1} \right), \quad (23)
\]

and \( \rho_{s,t} \) is defined according to (16).

3. MODELING RANDOM VARIABLES WITH TAPERED PARETO DISTRIBUTION

To model a random variable with continuous distribution function in a number of cases, one can use the inverse function method: if \( F \) is a continuous strictly monotone increasing distribution function and \( F^{-1} \) is its inverse, then the random variable \( F^{-1}(U) \), where \( U \) is a random variable uniformly distributed on \((0,1)\), has the distribution function \( F \); see Theorem 2.1 in [13]. For
example, the random variable $F_{a,\beta}^-(U)$, where $F_{a,\beta}^-(y) = a(1 - y)^{-1/\beta}$, $0 < y < 1$, is distributed by the Pareto law $F_{a,\beta}(x)$, $x \geq a$, and the random variable $\tilde{F}_{a,\theta}^+(U)$, where $\tilde{F}_{a,\theta}^+(y) = a - \theta \ln(1 - y)$, $0 < y < 1$, has the distribution function $\tilde{F}_{a,\theta}(x) = F_\theta(x - a)$, $x \geq a$.

It is noted in [2] that a random variable with tapered Pareto distribution and given parameters $a$, $\beta$, and $\theta$ can be modeled based on (2). Namely, the random variable

$$
\min \left\{ F_{a,\beta}^-(1 - U_1), \tilde{F}_{a,\theta}^+(1 - U_2) \right\}
$$

has the distribution function $F_{a,\beta,\theta}(x)$, where $U_1$ and $U_2$ are independent random variables uniformly distributed on $(0,1)$. A misapplication of this method occurs in the paper [14], which states that the random variable

$$
V = \min \left\{ F_{a,\beta}^-(1 - U), \tilde{F}_{a,\theta}^+(1 - U) \right\},
$$

where $U$ is a random variable uniformly distributed on $(0,1)$, has the distribution function $F_{a,\beta,\theta}(x)$.

Prior to obtaining the distribution of the random variable $V$, recall that the Lambert function $W(x)$ is defined for $x \geq -1/e$ as the solution of the equation

$$
W(x) \exp \{W(x)\} = x.
$$

For $x - 1/e \leq x < 0$, there exist two possible real values of the function $W(x)$ (see Fig. 1 in [15]). The branch satisfying $W(x) \geq -1$ is usually denoted by $W_0(x)$; and the branch satisfying $W(x) \leq -1$, by $W_{-1}(x)$.

**Proposition 1.** Let $c = a/(\beta \theta)$. If $c \geq 1$, then the random variable $V$ has the distribution function $\tilde{F}_{a,\theta}(x)$, $x \geq a$. If $0 < c < 1$, then for $x \geq a$ one has

$$
P (V \leq x) = \begin{cases} 
F_{a,\beta}(x), & x < -(a/c)\exp\{-c\} \left\{-c \exp\{-c\}\right\} \\
\tilde{F}_{a,\theta}(x), & x \geq -(a/c)\exp\{-c\} \left\{-c \exp\{-c\}\right\}.
\end{cases}
$$

(24)

Note that the distribution function $P (V \leq x)$, $x \geq a$, can be rewritten as follows. For each $c > 0$, one has

$$
P (V \leq x) = \max \left\{ F_{a,\beta}(x), \tilde{F}_{a,\theta}(x) \right\}, \quad x \geq a.
$$

Obviously, $P (V > x) / \left(1 - \tilde{F}_{a,\theta}(x)\right) \to 1$ as $x \to \infty$. It follows that max $\left\{ F_{a,\beta}(x), \tilde{F}_{a,\theta}(x) \right\}$, $x \geq a$, can be viewed as one more way of tapering the Pareto distribution by an exponential law. In the present paper, we do not consider the estimation of parameters of the distribution function max $\left\{ F_{a,\beta}(x), \tilde{F}_{a,\theta}(x) \right\}$, $x \geq a$.

A random variable with tapered Pareto distribution is modeled in [9] using the elimination method. This method is set forth in detail in [13]. Note that the inverse function of the distribution function $F_{a,\beta,\theta}(x)$, $x \geq a$, is expressed as

$$
F_{a,\beta,\theta}^+(y) = a \frac{W_0(c \exp(1 - y)^{-1/\beta})}{c}, \quad 0 < y < 1,
$$

(25)

where $c$ is the same as in Proposition 1. Therefore, an application of Theorem 2.1 in [13] allows one to model a random variable with tapered Pareto distribution using the inverse function method.

**Proposition 2.** If $U$ is a random variable uniformly distributed on the interval $(0,1)$, then $F_{a,\beta,\theta}^-(1 - U)$ has the distribution function $F_{a,\beta,\theta}$.
4. COMPARISON OF THE PAIRS \((\hat{\beta}_n^{(1)}, \hat{\lambda}_n^{(1)})\) AND \((\hat{\beta}_{n,r}^{(2)}, \hat{\lambda}_{n,r}^{(2)})\) OF PARAMETER ESTIMATES BY THE MAXIMUM LIKELIHOOD METHOD AND THE MOMENT METHOD

Recall that the generalized variance of the Gaussian vector \((\eta_{1,r}, \eta_{2,r})\) defined in Theorem 3 is the determinant of its covariance matrix,

\[
GV_r(\beta, \lambda) = \sigma^2_{1,r}(\beta, \lambda)\sigma^2_{2,r}(\beta, \lambda) - \sigma^2_{12,r}(\beta, \lambda).
\]

(26)

The generalized variance \(GV_r(\beta, \lambda)\) characterizes the degree of random spread. Let us take \(GV_r(\beta, \lambda)\) as the asymptotic measure of the efficiency of the estimates \((\hat{\beta}_n^{(2)}, \hat{\lambda}_{n,r}^{(2)})\). The generalized variance \(GV_r(\beta, \lambda)\) depends on the Box–Cox parameter \(r\). It is natural to choose an \(r\) so as to minimize \(GV_r(\beta, \lambda)\). However, the expression for \(GV_r(\beta, \lambda)\) is complicated; namely, if \(r + i \neq 0, i = 0, 1, \ldots, 4\), then

\[
GV_r(\beta, \lambda) = \frac{\lambda^2(d_{1,r}(\beta, \lambda) + d_{2,r}(\beta, \lambda)\Gamma(2r - \beta, \lambda) + d_{3,r}(\beta, \lambda)\Gamma^2(2r - \beta, \lambda))}{(\lambda^{2r} + d_{4,r}(\beta, \lambda)\Gamma(r - \beta, \lambda) + d_{5,r}(\beta, \lambda)\Gamma^2(r - \beta, \lambda))^2},
\]

(27)

where

\[
d_{1,r}(\beta, \lambda) = -\lambda^4r(2\beta + 23r + 2\beta r - 2r - 1),
\]

\[
d_{2,r}(\beta, \lambda) = -2e^{\lambda}\lambda^{\beta+2r}(-\beta^2 - \beta\lambda + 2\beta r^2 + 2\lambda r^2 - 4r^2 - \beta^2 r - 2\beta\lambda r + 3\beta r - \lambda^2 r - 2r),
\]

\[
d_{3,r}(\beta, \lambda) = e^{2\lambda\lambda^{2\beta}(2r - \beta)}(\beta + 4r^2 + 2r),
\]

\[
d_{4,r}(\beta, \lambda) = -e^{\lambda}\lambda^{\beta+r}(\beta + \lambda - r + 1),
\]

\[
d_{5,r}(\beta, \lambda) = -e^{2\lambda\lambda^{2\beta}(r - \beta)}.
\]

Therefore, the minimization of \(GV_r(\beta, \lambda)\) with respect to \(r\) in the present paper is performed numerically by the principal axis method. The point of global minimum of \(r^*\) on \(\mathbb{R}\) will be called the optimal choice of the parameter \(r\). The contour plot (with the values indicated on isolines) of \(r^*\) as a function of \(\beta\) and \(\lambda\) is presented in Fig. 1.

We define the asymptotic relative efficiency (ARE) of \((\hat{\beta}_{n,r}^{(2)}, \hat{\lambda}_{n,r}^{(2)})\) with respect to \((\hat{\beta}_n^{(1)}, \hat{\lambda}_n^{(1)})\) as the generalized variance ratio for the Gaussian vectors \((\eta_{1,r}, \eta_{2,r})\) and \((\xi_1, \xi_2)\) defined in Theorems 2

Fig. 1. Contour plot of the function \(r^*(\beta, \lambda)\).
ESTIMATING THE PARAMETERS OF A TAPERED PARETO DISTRIBUTION

Fig. 2. Contour plot of \( \text{ARE}(\beta, \lambda) \) (left) and the graph of \( \tilde{\text{ARE}}(\beta_{i,j}, \lambda_{i,j}), \ 0 \leq i \leq 29, 0 \leq j \leq 19 \) (right).

Fig. 3. Contour plot of \( \text{ARE}_1(\beta, \lambda) \) (left) and the graph of \( \tilde{\text{ARE}}_1(\beta_{i,j}, \lambda_{i,j}), \ 0 \leq i \leq 29, 0 \leq j \leq 19 \) (right).

and 1, respectively,

\[
\text{ARE}(\beta, \lambda) = \frac{\sigma_{1,r}^2(\beta, \lambda) \sigma_{2,r}^2(\beta, \lambda) - \sigma_{12,r}^2(\beta, \lambda)}{S_1^2(\beta, \lambda) S_2^2(\beta, \lambda) - S_{12}^2(\beta, \lambda)}, \quad \beta > 0, \quad 0 < \lambda < 1.
\]

In a similar way, we define the ARE for \( \hat{\beta}_{(2)}^{(2)} \) with respect to \( \hat{\beta}_{(1)}^{(1)} \) and for \( \hat{\lambda}_{(2)}^{(2)} \) with respect to \( \hat{\lambda}_{(1)}^{(1)} \),

\[
\text{ARE}_1(\beta, \lambda) = \frac{\sigma_{1,r}^2(\beta, \lambda)}{S_1^2(\beta, \lambda)}, \quad \text{ARE}_2(\beta, \lambda) = \frac{\sigma_{2,r}^2(\beta, \lambda)}{S_2^2(\beta, \lambda)}, \quad \beta > 0, \quad 0 < \lambda < 1.
\]

The contour plot in Fig. 2 (left) shows that \( \text{ARE}(\beta, \lambda) \) is close to 1 almost in the entire rectangle \((\beta, \lambda) \in (0, 1, 5) \times (0, 1)\). The theoretical prevalence of the estimates \( \left( \hat{\beta}_{(1)}^{(1)}, \hat{\lambda}_{(1)}^{(1)} \right) \) is more significant in the domain where the variables \( \lambda \) are small. This remarks also holds for \( \text{ARE}_i(\beta, \lambda), i \in \{1, 2\}; \) see Fig. 3 (left) and Fig. 4 (left), respectively.

In what follows, we use numerical simulation to demonstrate the performance of our estimates. In the simulation, we compare the estimates \( \left( \hat{\beta}_{(1)}^{(1)}, \hat{\lambda}_{(1)}^{(1)} \right) \) and \( \left( \hat{\beta}_{(2)}^{(2)}, \hat{\lambda}_{(2)}^{(2)} \right) \) using samples of medium and small volume \( n \) from the distribution (1). In the entire simulation, it is assumed that the parameter \( a = 1 \) is unknown and is estimated using \( \hat{a}_n \).

For practical purposes, to choose the parameter \( r \) in an optimal way, it suffices to minimize \( \text{GV}_r(\beta, \lambda) \) (see (27)) with respect to the parameter \( r \in (-1, 0) \). Since the optimal choice of \( r^* \) depends on the unknown parameters \( \beta \) and \( \lambda \), we propose the following iteration procedure.

AUTOMATION AND REMOTE CONTROL Vol. 82 No. 8 2021
If the \( k \)th iteration in the estimation of the optimal choice of \( r^* \) is performed, in which the approximation \( r^*_{k+1} \) is calculated, then we perform the \((k+1)\)st iteration as follows:

1. Calculate the estimates \( \hat{\beta}_{n,r^*_{k}} \) and \( \hat{\lambda}_{n,r^*_{k}} \) by formulas (14) and (15), respectively.

2. Calculate \( r^*_{k+1} = \arg\min \{ \epsilon \mid \epsilon = \frac{1}{n} \} : \text{GV}_r \left( \hat{\beta}_{n,r^*_{k}}, \hat{\lambda}_{n,r^*_{k}} \right) \} \) by formula (27).

3. Check the stop criterion \( |r^*_{k+1} - r^*_{k}| \leq \epsilon \) and if it is not satisfied, then go to step 1.

After the stop criterion is satisfied, the value \( r^*_{k+1} \) is taken for \( r^* \), and accordingly, one sets \( \hat{\beta}_{n,r^*} = \hat{\beta}_{n,r^*_{k}} \) and \( \hat{\lambda}_{n,r^*} = \hat{\lambda}_{n,r^*_{k}} \). One more constraint is used to avoid a large number of iterations. Namely, if \( |r^*_{k+1} - r^*_{k}| \leq \epsilon \) for all \( 0 \leq k \leq 9 \), then \( r^* = r^*_{10} \), \( \hat{\beta}_{n,r^*} = \hat{\beta}_{n,r^*_{10}} \), and \( \hat{\lambda}_{n,r^*} = \hat{\lambda}_{n,r^*_{10}} \). In the modeling, we have used the initial approximation \( r^*_0 = -1/2 \) and the accuracy \( \epsilon = 0.001 \). Note that step 2 in our procedure can be performed by combining the functions “Last” and “FindMinimum” in “Wolfram Mathematica 10.”

In the first part of the modeling, the rectangle \((0, 3/2) \times (0, 1)\) is divided into the squares

\[
s_{i,j} = \left\{(i/20, (i+1)/20) \times (j/20, (j+1)/20), \quad 0 \leq i \leq 29, \quad 0 \leq j \leq 19\right\}.
\]

For the true values of the parameters \( \beta \) and \( \lambda \) we take the coordinates

\[
\beta_{i,j} = \frac{2i + 1}{40}, \quad \lambda_{i,j} = \frac{2j + 1}{40}
\]

of the center of the square \( s_{i,j} \) and generate \( N = 750 \) realizations of the random sample from the distribution (1) with a size of \( n = 2500 \) observations.

Let \( \left( \hat{\beta}_n^{(1,k)}, \hat{\lambda}_n^{(1,k)} \right) \) and \( \left( \hat{\beta}_n^{(2,k)}, \hat{\lambda}_n^{(2,k)} \right) = \left( \hat{\beta}_{n,r^*}, \hat{\lambda}_{n,r^*} \right) \) denote the estimates of the parameters \( (\beta, \lambda) \) calculated based on the \( k \)th realization. Consider the empirical moments

\[
m_{i,j} = \frac{1}{N} \sum_{k=1}^{N} \left( \hat{\beta}_n^{(i,k)} \right)^j, \quad i = 1, 2, \quad j = 1, 2,
\]

\[
M_{i,j} = \frac{1}{N} \sum_{k=1}^{N} \left( \hat{\lambda}_n^{(i,k)} \right)^j, \quad i = 1, 2, \quad j = 1, 2,
\]

\[
\mu_i = \frac{1}{N} \sum_{k=1}^{N} \hat{\beta}_n^{(i,k)} \hat{\lambda}_n^{(i,k)}, \quad i = 1, 2.
\]

Fig. 4. Contour plot of \( \text{ARE}_2(\beta, \lambda) \) (left) and the graph of \( \text{ARE}_2(\beta_{i,j}, \lambda_{i,j}) \), \( 0 \leq i \leq 29, \quad 0 \leq j \leq 19 \) (right).
The sample analogs of the characteristics \( \text{ARE}(\beta, \lambda) \), \( \text{ARE}_1(\beta, \lambda) \), and \( \text{ARE}_2(\beta, \lambda) \) have the form

\[
\text{ARE}(\beta_{i,j}, \lambda_{i,j}) = \frac{(m_{2,2} - (m_{2,1})^2)}{(m_{1,2} - (m_{1,1})^2)} \left( M_{2,2} - (M_{2,1})^2 \right) - \frac{(\mu_2 - m_{2,1}M_{2,1})^2}{(\mu_1 - m_{1,1}M_{1,1})^2},
\]

\[
\text{ARE}_1(\beta_{i,j}, \lambda_{i,j}) = \frac{m_{2,2} - (m_{2,1})^2}{m_{1,2} - (m_{1,1})^2}, \quad \text{ARE}_2(\beta_{i,j}, \lambda_{i,j}) = \frac{M_{2,2} - (M_{2,1})^2}{M_{1,2} - (M_{1,1})^2}.
\]

The graph of \( \text{ARE}(\beta_{i,j}, \lambda_{i,j}) \), \( 0 \leq i \leq 29 \), \( 0 \leq j \leq 19 \), is presented in Fig. 2 (right), where the square \( s_{i,j} \) is colored as follows:

\[
\begin{align*}
\text{black} & \quad \text{if} \ \text{ARE}(\beta_{i,j}, \lambda_{i,j}) \geq 1.03 \\
\text{dark gray} & \quad \text{if} \ 1.01 \leq \text{ARE}(\beta_{i,j}, \lambda_{i,j}) < 1.03 \\
\text{gray} & \quad \text{if} \ 1 \leq \text{ARE}(\beta_{i,j}, \lambda_{i,j}) < 1.01 \\
\text{white} & \quad \text{if} \ \text{ARE}(\beta_{i,j}, \lambda_{i,j}) < 1.
\end{align*}
\]

The same color palette is used to produce the graphs of \( \text{ARE}_1(\beta_{i,j}, \lambda_{i,j}) \) and \( \text{ARE}_2(\beta_{i,j}, \lambda_{i,j}) \); see Fig. 3 (right) and Fig. 4 (right), respectively.

The second part of modeling deals with small-volume samples. The true values of \((\beta, \lambda)\) were chosen to be \(\beta_i = i/20\), \(\lambda_i = c\beta_i\), and \(1 \leq i \leq 29\), where \(c = 1/96, 1/24, 1/6\). Samples from the distribution \(F_{1,\beta_i,1/\lambda_i}\) of volume \(n = 110\) were modeled \(N = 1500\) times. From each sample \((X_1^{(k)}, \ldots, X_n^{(k)})\), \(1 \leq k \leq N\), we formed three samples of volume \(n = 100\):

(a) The nontruncated sample \((X_1^{(k)}, \ldots, X_{100}^{(k)})\).

(b) The sample with truncated five least observations \((X_6^{(k)}, \ldots, X_{105}^{(k)}, X_{105,105}^{(k)})\), where \(X_1^{(k)} \leq \ldots \leq X_n^{(k)}\).

(c) The sample with truncated ten least observations \((X_{11}^{(k)}, \ldots, X_{110}^{(k)}, X_{110,110}^{(k)})\).

For the efficiency of parameter estimation for nontruncated samples, we considered the ratios of the mean squares of deviation from the true parameter values,

\[
\psi_1(c, \beta_i) = \frac{m_{2,2} - 2\beta_i m_{2,1} + \beta_i^2}{m_{1,2} - 2\beta_i m_{1,1} + \beta_i^2},
\]

\[
\Psi_1(c, \beta_i) = \frac{M_{2,2} - 2c\beta_i M_{2,1} + c^2\beta_i^2}{M_{1,2} - 2c\beta_i M_{1,1} + c^2\beta_i^2}.
\]

Let \(\psi_2(c, \beta_i)\), \(\psi_3(c, \beta_i)\) and \(\Psi_2(c, \beta_i)\), \(\Psi_3(c, \beta_i)\) be the counterparts of the empirical characteristics \(\psi_1(c, \beta_i)\) and \(\Psi_1(c, \beta_i)\) calculated based on the left-truncated samples. The numerical results are presented in Figs. 5–7.

Based on the modeling results, we can draw the following conclusions:

1. It follows from comparing the contour graph \(\text{ARE}(\beta, \lambda)\) and the graph \(\text{ARE}(s_{i,j}, \lambda_{i,j})\), \(0 \leq i \leq 29\), \(0 \leq j \leq 19\) in Fig. 2 that the results of modeling in Fig. 2 (right) correspond on the whole to the theoretical results in Fig. 2 (left). First, the values of \(\text{ARE}(\beta, \lambda)\) do not exceed 1.01 in the rectangle \((\beta, \lambda) \in (0, 1.5] \times (0, 2, 1)\), as reflected in the graph of \(\text{ARE}(s_{i,j}, \lambda_{i,j})\). Moreover, the sufficient number of white squares in the rectangle in Fig. 2 (right) \((\beta, \lambda) \in (0, 1.5] \times (0, 2, 1)\) indicates that our estimates \((\hat{\beta}_{n,n}^{(2)}, \hat{\lambda}_{n,n}^{(2)})\) can be useful for practical applications. Second, the results of modeling in Fig. 2 (right) rather well reflect the
prevalence of the maximal likelihood estimates for small values of $\lambda$. Essentially, the graphs in Figs. 3 and 4 are no different from the graphs in Fig. 2.

2. Since $\psi_1(c, \beta_i) > 1$ for all $\beta_i$, $1 \leq i \leq 29$, and $c = 1/96, 1/24, 1/6$, the estimate $\hat{\beta}_n^{(1)}$ prevails over the estimate $\hat{\beta}_n^{(2)}$ at all studied points $(\beta_i, \lambda_i)$, $1 \leq i \leq 29$. It should be noted that the prevalence of the estimate $\hat{\beta}_n^{(1)}$ persists when estimating the parameter $\beta$ based on the left-truncated samples; see the graphs of broken lines $\psi_j(c, \beta_i)$, $j = 2, 3$, in Figs. 5–7 (left). The truncation of the samples does not affect the relative statistics of $\psi_j(c, \beta_i)$, $j = 2, 3$; i.e., the values of $\psi_1$, $\psi_2$, and $\psi_3$ for fixed $c$ and $\beta_i$ are no different in practical terms.

3. Comparing the estimates $\hat{\lambda}_n^{(1)}$ and $\hat{\lambda}_n^{(2)}$ with nontruncated samples depends on the value of the parameter $c$. For $c = 1/6$, none of the estimates $\hat{\lambda}_n^{(1)}$ and $\hat{\lambda}_n^{(2)}$ exceeds the other; see the
Table 1. Empirical characteristics

|  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|
| $\beta$ | 0.05 | 0.75 | 1.45 | 0.05 | 0.75 | 1.45 | 0.05 | 0.75 | 1.45 |
| $100 \cdot \lambda$ | $5/6$ | $75/6$ | $145/6$ | $5/24$ | $75/24$ | $145/24$ | $5/96$ | $75/96$ | $145/96$ |
| $100 \cdot BI_1$ | 0.01 | 0.55 | 5.25 | 0.04 | 0.70 | 5.12 | 0.18 | 1.50 | 7.79 |
| $100 \cdot bi_1$ | 0.01 | 0.57 | 5.34 | 0.04 | 0.73 | 5.13 | 0.18 | 1.51 | 7.72 |
| $100 \cdot BI_2$ | 0.13 | 0.63 | 5.32 | 0.04 | 0.91 | 5.26 | 1.40 | 2.40 | 8.58 |
| $100 \cdot bi_2$ | 0.13 | 0.63 | 5.35 | 0.03 | 0.94 | 5.27 | 1.36 | 2.28 | 8.46 |
| $100 \cdot BI_3$ | 0.42 | 0.70 | 5.30 | 1.32 | 1.17 | 5.44 | 3.21 | 3.26 | 8.41 |
| $100 \cdot bi_3$ | 0.41 | 0.68 | 5.30 | 1.27 | 1.16 | 0.56 | 3.11 | 2.98 | 9.04 |
| $100 \cdot RD_1$ | 0.01 | 1.16 | 8.87 | 0.06 | 2.23 | 11.45 | 0.24 | 6.15 | 23.23 |
| $100 \cdot rd_1$ | 0.01 | 1.17 | 8.81 | 0.06 | 2.24 | 11.40 | 0.24 | 6.15 | 23.73 |
| $100 \cdot RD_2$ | 0.11 | 1.32 | 9.30 | 0.39 | 2.45 | 11.98 | 0.97 | 6.76 | 24.30 |
| $100 \cdot rd_2$ | 0.10 | 1.19 | 8.88 | 0.35 | 2.34 | 11.30 | 0.87 | 6.58 | 23.62 |
| $100 \cdot RD_3$ | 0.34 | 1.43 | 9.31 | 0.94 | 2.71 | 12.50 | 1.73 | 7.28 | 25.04 |
| $100 \cdot rd_3$ | 0.30 | 1.23 | 8.35 | 0.85 | 2.46 | 11.60 | 1.54 | 6.88 | 23.87 |

Graph of the broken line $\Psi_1(1/6, \beta_i), 1 \leq i \leq 29$ in Fig. 7 (right). For $\beta_i \in [1, 3/2)$, the values of the broken line $\Psi_1(1/96, \beta_i)$ oscillate around unity, but for $\beta_i \in (0, 1)$ the estimate $\hat{\lambda}_n^{(1)}$ takes precedence; see Fig. 5 (right). A similar conclusion can also be drawn with regard to the estimates $\hat{\lambda}_n^{(1)}$ and $\hat{\lambda}_{n,r}^{(2)}$ for $c = 1/24$. When truncating the observation on the left, the estimate $\hat{\lambda}_{n,r}^{(2)}$ exceeds the estimate $\hat{\lambda}_n^{(1)}$; see the graphs of the broken lines $\Psi_2$ and $\Psi_3$ in Figs. 5–7 (right). Proceeding from the results of modeling presented in Table 1, we can conclude that the precedence of the estimate $\hat{\lambda}_{n,r}^{(2)}$ is mainly created owing to a better empirical standard deviation. In this table, by $BI_1 = M_{1,1} - \lambda$, $bi_1 = M_{2,1} - \lambda$, $RD_1 = M_{1,2} - (M_{1,1})^2$, $rd_1 = M_{2,2} - (M_{2,1})^2$, we have denoted the empirical biases and standard deviation of the estimates $\hat{\lambda}_n^{(1)}$ and $\hat{\lambda}_{n,r}^{(2)}$, for nontruncated samples. With truncated samples, we use the same notation but with subscripts 2 and 3, respectively.

4. The values of $BI_1$ given in Table 1 make it possible to judge the bias of the estimate $\hat{\lambda}_n^{(1)}$ for nontruncated samples. The results of modeling do not contradict the conclusion made in [4] that the maximal likelihood estimate (just as the estimate $\hat{\lambda}_{n,r}^{(2)}$) proves to be considerably biased to the right. Left truncation leads to both estimates of the parameter $\lambda$ being biased to the right with the increase in the standard deviations depending on the parameter $c$.

5. APPLICATION TO COVID-19 DATA

In the present paper, data related to the incidence of the coronavirus infection COVID-19 by day (see https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge) were investigated. The number of individuals recovered by day was analyzed by preliminary truncation on the left with different levels. The analysis was based on the data collected up to June 22, 2020, i.e., during the first wave of the pandemic. Let us apply the tapered Pareto distribution as a model for the distribution of the number of recovered individuals for two neighboring
countries of East Asia—Japan and South Korea. The average population density in Japan is 338 people, and in South Korea it is 496 people per 1 sq. km. These countries are comparable in terms of the human development index, ranking 19 and 22 in the world list of countries, respectively.

The estimates of parameters (by the maximal likelihood and moment methods) of the tapered Pareto distribution are presented in Tables 2 and 3, where, for the analysis to be complete, we also provide estimates for the upper truncation parameter $\theta$. The parameter $a$ is assumed to be known and equal to the selected left truncation level. In Tables 2 and 3, $n(a)$ denotes the volume of the left-truncated sample.

If the maximum likelihood and moment methods produce differing estimates for $\beta$ and $\lambda$ (or $\theta$) for one particular country, it is difficult to decide whether this difference is statistically significant. Therefore, we apply several goodness-of-fit tests. Let $F_n$ be the empirical distribution function of the sample $X_1, \ldots, X_n$, and let $X_{1,n} < X_{2,n} < \cdots < X_{m,n}$ be its order statistics. Let $\hat{\beta}$ and $\hat{\lambda}$ denote the estimates of the parameters $\beta$ and $\lambda$.

The Kolmogorov–Smirnov (KS) test determines the largest distance between $F_n$ and the distribution function of the tapered Pareto distribution $F_{a,\hat{\beta},\hat{\lambda}},$

$$ K_n = \sqrt{n} \sup_{x \geq a} \left| F_n(x) - F_{a,\hat{\beta},\hat{\lambda}}(x) \right|. $$

The KS statistics has the form $K_n = \sqrt{m} \max \{K_m^-, K_m^+\}$, where

$$ K_m^+ = \max_{1 \leq i \leq m} \left\{ F_n(X_{i,n}) - F_{a,\hat{\beta},\hat{\lambda}}(X_{i,n}) \right\}, $$

$$ K_m^- = \max_{2 \leq i \leq m} \left\{ F_{a,\hat{\beta},\hat{\lambda}}(X_{i,n}) - F_n(X_{i-1,n}) \right\}. $$

The statistics $\tilde{K}_n$ for Kuiper’s test (K) has a representation in the form $\tilde{K}_n = \sqrt{m} (K_m^- + K_m^+)$. The statistics for the Cramér–von Mises–Smirnov test (CvMS) uses a quadratic measure of closeness between $F_n(x)$ and $F_{a,\hat{\beta},\hat{\lambda}},$

$$ \omega_n^2 = n \int_a^\infty \left( F_n(x) - F_{a,\hat{\beta},\hat{\lambda}}(x) \right)^2 dF_{a,\hat{\beta},\hat{\lambda}}(x). $$
Table 4. p-values for KS, K, CvMS, and AD tests

|      | Japan |         |          |         |          |          |          |          |
|------|-------|---------|----------|---------|----------|----------|----------|----------|
|      | KS    | K       | CvMS     | AD      |          |          |          |          |
| a    | $F_1$ | $F_2$  | $F_1$   | $F_2$  | $F_1$   | $F_2$   | $F_1$   | $F_2$   |
| 5    | 0.040 | 0.129   | 0.015   | 0.042   | 0.070   | 0.161   | 0.001   | 0.002   |
| 10   | 0.130 | 0.281   | 0.030   | 0.049   | 0.152   | 0.235   | 0.002   | 0.003   |
| 15   | 0.220 | 0.406   | 0.039   | 0.051   | 0.214   | 0.278   | 0.080   | 0.094   |
| 20   | 0.294 | 0.408   | 0.048   | 0.059   | 0.254   | 0.305   | 0.090   | 0.098   |

|      | South Korea |         |          |         |          |          |          |          |
|      | KS          | K       | CvMS     | AD      |          |          |          |          |
| a    | $F_1$       | $F_2$  | $F_1$   | $F_2$  | $F_1$   | $F_2$   | $F_1$   | $F_2$   |
| 5    | 0.284       | 0.531  | 0.062   | 0.128  | 0.136   | 0.278   | 0.042   | 0.072   |
| 10   | 0.533       | 0.625  | 0.165   | 0.239  | 0.577   | 0.701   | 0.512   | 0.631   |
| 15   | 0.690       | 0.782  | 0.324   | 0.422  | 0.726   | 0.802   | 0.170   | 0.187   |
| 20   | 0.728       | 0.792  | 0.585   | 0.634  | 0.829   | 0.860   | 0.232   | 0.242   |

It is convenient to perform practical calculations for the CvMS test using the formula

$$\omega_n^2 = \frac{1}{12m} + \sum_{i=1}^{m} \left( F_{a,\hat{\beta},a/\hat{\lambda}}(X_{i,n}) - \frac{2i - 1}{2m} \right)^2. $$

The statistics of the Anderson–Darling test (AD) uses a weighted quadratic measure of closeness between $F_n(x)$ and $F_{a,\hat{\beta},a/\hat{\lambda}}$,

$$A_n^2 = n \int_{a}^{\infty} \frac{\left( F_n(x) - F_{a,\hat{\beta},a/\hat{\lambda}}(x) \right)^2}{\left( 1 - F_{a,\hat{\beta},a/\hat{\lambda}}(x) \right) F_{a,\hat{\beta},a/\hat{\lambda}}(x)} dF_{a,\hat{\beta},a/\hat{\lambda}}(x). $$

Practical calculation of statistics for the AD test is performed using the formula

$$A_n^2 = -m - \sum_{i=1}^{m} \frac{2i - 1}{m} \left( \ln \left( F_{a,\hat{\beta},a/\hat{\lambda}}(X_{i,n}) \right) + \ln \left( 1 - F_{a,\hat{\beta},a/\hat{\lambda}}(X_{m+1-i,n}) \right) \right). $$

With the null hypothesis that the sample corresponds to the distribution $F_{a,\beta,a/\lambda}$, the statistics $K_n$, $\tilde{K}_n$, $\omega_n^2$, and $A_n^2$ decrease with the number of observations increasing to infinity. Using the procedures provided in the “Wolfram Mathematica 10” package, it is easy to calculate the p-values for the above-listed statistical tests. Table 4 contains the p-values for the KS, K, CvMS, and AD tests, where the p-values that are above the level of significance $\alpha = 0.05$ are printed in boldface. In this case, the null hypothesis is not rejected. In Table 4, $F_1$ and $F_2$ denote the distribution functions $F_{a,\hat{\beta}^{(1)}(1),a/\hat{\lambda}^{(1)}(1)}$ and $F_{a,\hat{\beta}^{(2)}(2),a/\hat{\lambda}^{(2)}(2)}$, respectively. Recall that $F_{a,\hat{\beta}^{(1)}(1),a/\hat{\lambda}^{(1)}(1)}$ and $F_{a,\hat{\beta}^{(2)}(2),a/\hat{\lambda}^{(2)}(2)}$ denote models with parameters estimated by the maximum likelihood method and by the moment method proposed in the paper, respectively.

We can draw the following conclusions based on the results of analysis.

1. It is difficult to identify a parametric model based on a small sample of approximately 100 observations (see the $N(a)$ in Tables 2, 3), because one can select a set of models that equally well describe the sample data from the standpoint of several goodness-of-fit criteria.
2. Comparing the corresponding \(p\)-values for the \(F_1\) and \(F_2\) models in Table 4, note that all the tests used higher \(p\)-values for the \(F_2\) model than for the \(F_1\) model. This indicates that the model with the new parameter estimates \((\beta, \lambda)\) describes the analyzed data better than the \(F_1\) model.

4. The \(p\)-values given in Table 4 grow with the truncation level \(a\). Hence it follows that the observations (for both Japan and South Korea) from the lower part of the variation series are worse described by the tapered Pareto distribution.

5. Table 4 shows that the tapered Pareto distribution better describes the number of individuals recovered from COVID-19 in South Korea than it does in Japan. This is reflected in the estimates of the parameters \(\beta\) and \(\lambda\) (at the corresponding level of truncation), which differ by 1.5–2 times; see Tables 2 and 3.

6. CONCLUSIONS

The paper uses the moment method to produce a family of new estimates for the parameters of the tapered Pareto distribution. It is proved that the joint asymptotic distribution of the new estimates is Gaussian. A procedure for the optimal choice of the additional parameter is proposed. Theoretical comparison has shown that for nontruncated samples, the pair of the newly introduced estimates is slightly inferior in performance to the corresponding pair of maximum likelihood estimates. Computer simulations have shown that the maximum likelihood estimates lose their precedence for some true values of the parameters \(\beta\) and \(\lambda\); see Figs. 2–4.

The novelty of the present paper is also a proposal for a random number generator distributed according to the tapered Pareto distribution.

In statistical data analysis problems, one often uses truncation of observations on the left. Empirical studies of the statistical properties of estimates for samples containing left-truncated observations have shown that of the estimates \(\hat{\lambda}_{(1)}\) and \(\hat{\lambda}_{(2)}\), it is preferable to use the newly introduced estimates \(\hat{\lambda}_{n,r}^{(2)}\). As for the estimates \(\hat{\beta}_{n}^{(1)}\) and \(\hat{\beta}_{n,r}^{(2)}\), data truncation affects their efficiency in approximately the same way.

Computer simulation has revealed that with both left-truncated and nontruncated data, the bias of the estimates \(\hat{\lambda}_{(1)}\) and \(\hat{\lambda}_{n,r}^{(2)}\) contributes considerably to their standard deviations. Elimination (or reduction) of the bias of the discussed estimates is closely related to the study of the asymptotics of the expectations \(\mathbb{E}(\hat{\lambda}_{n}^{(1)} - \lambda)\) and \(\mathbb{E}(\hat{\lambda}_{n,r}^{(2)} - \lambda)\) as \(n \to \infty\). The present authors intend to solve this problem in the future.

APPENDIX

**Lemma 1.** Let \((X_1, \ldots, X_n)\) be a sample of a random variable with distribution \(F_{a,\beta, a/\lambda}\), and let \(r \in \mathbb{R}\). Then
\[
\sqrt{n} h_r (a/\hat{a}_n) \xrightarrow{\mathcal{D}} 0, \quad n \to \infty.
\] (A.1)

**Proof.** The assertion in Lemma 1 follows from the relation
\[
nh_r (a/\hat{a}_n) \xrightarrow{\mathcal{D}} - \frac{Z}{\lambda + \beta}, \quad n \to \infty,
\] (A.2)
where \(Z\) is an exponentially distributed random variable with distribution function \(P(Z \leq x) = 1 - \exp\{-x\}, x \geq 0\).

For each \(r \in \mathbb{R}\), the function \(h_r (x)\) strictly increases on \([1, \infty)\). Using this property, we can verify the identity
\[
h_r (a/\hat{a}_n) = - \min \left\{ h_{-r} \left( \frac{X_1}{a} \right), \ldots, h_{-r} \left( \frac{X_n}{a} \right) \right\}.
\]
Set \[ L_r = \begin{cases} +\infty, & r \leq 0 \\ 1/r, & r > 0. \end{cases} \]

We have \( P(h_{-r}(X_1/a) \leq x) = F_{1,\beta,1/\lambda(h_{-r}^{-1}(x))}, x \in [1, L_r). \) Using (25), one can readily verify that the function \( h_{-r}(F_{1,\beta,1/\lambda}^{-1}(y)), 0 \leq y \leq 1, \) is a generalized inverse of the distribution function \( F_{1,\beta,1/\lambda(h_{-r}^{-1}(x))}, x \in [1, L_r). \) According to Theorem 8.3.6(ii) in [16], the relation

\[
\lim_{n \to \infty} \frac{F_{1,\beta,1/\lambda}(h_{-r}(1/n) - h_{-r}(F_{1,\beta,1/\lambda}(0)))}{h_{-r}(F_{1,\beta,1/\lambda}(1/n) - h_{-r}(F_{1,\beta,1/\lambda}(0)))} \to Z, \quad n \to \infty,
\]

will be proved if we show that \( h_{-r}(F_{1,\beta,1/\lambda}(0)) \) is finite and one has

\[
\lim_{\epsilon \to 0^+} \frac{r^\epsilon}{f_{1,\beta,1/\lambda}(\exp(\epsilon x))} \exp\{\epsilon x - 1\} = x
\]

for all \( x > 0. \)

Combining (25) and the identity \( W_0(xe^x) = x, \) which holds true for \( x \geq -1, \) we obtain \( h_{-r}(F_{1,\beta,1/\lambda}(0)) = h_{-r}(1) = 0. \) Applying l’Hôpital’s rule, we conclude that the left-hand side of (A.3) is

\[
x \lim_{\epsilon \to 0^+} \frac{f_{1,\beta,1/\lambda}(\exp(\epsilon x))}{f_{1,\beta,1/\lambda}(\exp(x))} \exp\{\epsilon(x - 1)\} = x
\]

for \( r = 0 \) and

\[
x \lim_{\epsilon \to 0^+} \frac{f_{1,\beta,1/\lambda}(1 - r\epsilon x)^{-1/r}}{f_{1,\beta,1/\lambda}(1 - r\epsilon)^{-1/r}} \left(1 + \frac{r\epsilon(1 - x)}{1 - r\epsilon}\right)^{-1 - 1/r} = x
\]

for \( r \neq 0. \)

To complete the proof of relation (A.2), it remains to prove that \( h_{-r}(F_{1,\beta,1/\lambda}(n^{-1})) \sim (\lambda + \beta)^{-1}n^{-1}, n \to \infty. \) Recall that

\[
\frac{dW_0(t)}{dt} = \frac{W_0(t)}{t(1 + W_0(t))}, \quad t \neq 0;
\]

see, e.g., (3.2) in [15]. Applying l’Hôpital’s rule one more time, we obtain

\[
\lim_{n \to \infty} \frac{h_{-r}(F_{1,\beta,1/\lambda}(n^{-1}))}{n^{-1}} = \lim_{t \to 0^+} \frac{\left(F_{1,\beta,1/\lambda}^{-1}(t)\right)^r}{\beta(1 - t) \left(1 + (\lambda/\beta)F_{1,\beta,1/\lambda}^{-1}(t)\right)^r} = \frac{1}{\lambda + \beta}.
\]

This completes the proof of the lemma. \( \blacksquare \)

**Lemma 2.** Let the assumptions of Lemma 1 be satisfied. Then

\[
\sqrt{n}\left((\hat{\nu}_{n,r}, \hat{\nu}_{n,r+1}, \hat{\nu}_{n,r+2}) - (\nu_r, \nu_{r+1}, \nu_{r+2})\right) \to_d (Y_r, Y_{r+1}, Y_{r+2}) \quad (A.4)
\]

as \( n \to \infty. \)
**Proof.** By the Cramér–Wold theorem (see Theorem 29.4 in [17]), it suffices to show that for each constant vector \((c_0, c_1, c_2) \in \mathbb{R}^3\) one has

\[
\sqrt{n} \sum_{k=0}^{2} c_k (\hat{\nu}_{n,r+k} - \nu_{r+k}) \overset{d}{\to} \sum_{k=0}^{2} c_k Y_{r+k}, \quad n \to \infty.
\]

Note that

\[
\zeta_i = \sum_{k=0}^{2} c_k \left( h_{r+k} \left( \frac{X_i}{a} \right) - \nu_{r+k} \right), \quad i = 1, 2, \ldots
\]

are independent identically distributed random variables with \(\mathbb{E}(\zeta_i) = 0\) and

\[
\text{Var}(\zeta_i) = \sum_{i,j=0}^{2} c_i c_j \rho_{r+i,r+j}.
\]

Using the Lindeberg–Lévy central limit theorem (see, e.g., Theorem 27.1 in [17]), we conclude that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i \overset{d}{\to} \sum_{k=0}^{2} c_k Y_{r+k}, \quad n \to \infty.
\]

(A.5)

In view of the decomposition

\[
\sum_{k=0}^{2} c_k (\hat{\nu}_{n,r+k} - \nu_{r+k}) = T_n + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i,
\]

where

\[
T_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{2} c_k \left( h_{r+k} \left( \frac{X_i}{\hat{a}_n} \right) - h_{r+k} \left( \frac{X_i}{a} \right) \right),
\]

it remains to show that

\[
\sqrt{n} T_n \overset{p}{\to} 0, \quad n \to \infty,
\]

by noticing that the relation

\[
h_r \left( \frac{X_1}{\hat{a}_n} \right) - h_r \left( \frac{X_1}{a} \right) = \begin{cases} h_0 \left( a/\hat{a}_n \right), & r = 0 \\ rh_r \left( a/\hat{a}_n \right) \{ h_r \left( X_1/a \right) - \nu_r \} + (r\nu_r + 1) h_r \left( a/\hat{a}_n \right), & r \neq 0 \end{cases}
\]

follows from (A.1) and (A.5). □

**Proof of Theorem 2.** Applying the Cramér–Wold theorem one more time, it suffices to prove that for each constant vector \((c_0, c_1) \in \mathbb{R}^2\) one has

\[
\sqrt{n} \left( (c_0 \hat{\beta}^{(2)}_{n,r} + c_1 \hat{\lambda}^{(2)}_{n,r}) - (c_0 \beta + c_1 \lambda) \right) \overset{d}{\to} c_0 \eta_{1,r} + c_1 \eta_{2,r}, \quad n \to \infty.
\]

(A.6)

Set

\[
g(x_1, x_2, x_3) = c_0 r + \frac{c_1 g^{(2)}(x_1, x_2) - c_0 g^{(1)}(x_2, x_3)}{g^{(3)}(x_1, x_2, x_3)},
\]

where the functions \(g^{(1)}, g^{(2)}\) and \(g^{(3)}\) are the same as in (11)–(13). Using (10) and (14)–(15), we obtain

\[
(c_0 \hat{\beta}^{(2)}_{n,r} + c_1 \hat{\lambda}^{(2)}_{n,r}) - (c_0 \beta + c_1 \lambda) = g(\hat{\nu}_{n,r}, \hat{\nu}_{n,r+1}, \hat{\nu}_{n,r+2}) - g(\nu_r, \nu_{r+1}, \nu_{r+2}).
\]
Let \( g'(x_1, x_2, x_3) \) denote the partial derivative of \( g \) with respect to \( x_i \). One can readily verify that

\[
\begin{align*}
g'_1 (\nu_r, \nu_{r+1}, \nu_{r+2}) &= c_0 \frac{\nu_{r+2} (\nu_{r+1} + \nu^2_{r+1} - \nu_{r+2})}{(\nu_r \nu_{r+2} - \nu^2_{r+1})^2} + c_1 \frac{\nu_{r+1} (\nu_{r+2} - \nu_{r+1} - \nu^2_{r+1})}{(\nu_r \nu_{r+2} - \nu^2_{r+1})^2}, \\
g'_2 (\nu_r, \nu_{r+1}, \nu_{r+2}) &= c_0 \frac{2\nu_{r+1} \nu_{r+2} - \nu^2_{r+1} - \nu_r \nu_{r+2} - 2\nu_r \nu_{r+1} \nu_{r+2}}{(\nu_r \nu_{r+2} - \nu^2_{r+1})^2} + c_1 \frac{\nu_r \nu_{r+1} - \nu^2_{r+1} + \nu_r \nu_{r+2} + \nu^2_{r+2}}{(\nu_r \nu_{r+2} - \nu^2_{r+1})^2}, \\
g'_3 (\nu_r, \nu_{r+1}, \nu_{r+2}) &= c_0 \frac{\nu_{r+1} (\nu_r - \nu_{r+1} + \nu_r \nu_{r+1})}{(\nu_r \nu_{r+2} - \nu^2_{r+1})^2} + c_1 \frac{\nu_r (\nu_{r+1} - \nu_r - \nu_r \nu_{r+1})}{(\nu_r \nu_{r+2} - \nu^2_{r+1})^2}.
\end{align*}
\]

Using the delta method (see, e.g., Theorem 3.1 in [18]), we conclude that

\[
\sqrt{n}(g(\hat{\nu}_{n,r}, \hat{\nu}_{n,r+1}, \hat{\nu}_{n,r+2}) - g(\nu_r, \nu_{r+1}, \nu_{r+2}))
\]

converges in distribution to a random variable that has normal distribution with expectation 0 and variance

\[
G \begin{pmatrix} \rho_{r,r} & \rho_{r,r+1} & \rho_{r,r+2} \\ \rho_{r+r,1} & \rho_{r+1,r+1} & \rho_{r+1,r+2} \\ \rho_{r,r+2} & \rho_{r+1,r+2} & \rho_{r+2,r+2} \end{pmatrix} G^T = c_0^2 \sigma^2_{11,r}(\beta, \lambda) + 2c_0 c_1 \sigma_{12,r}(\beta, \lambda) + c_1^2 \sigma^2_{22,r}(\beta, \lambda),
\]

where \( \sigma^2_{11,r}(\beta, \lambda) \), \( \sigma_{12,r}(\beta, \lambda) \), and \( \sigma^2_{22,r}(\beta, \lambda) \) are given in (19)–(21) and

\[
G = (g'_1 (\nu_r, \nu_{r+1}, \nu_{r+2}) \ g'_2 (\nu_r, \nu_{r+1}, \nu_{r+2}) \ g'_3 (\nu_r, \nu_{r+1}, \nu_{r+2})).
\]

It remains to note that the random variable \( c_0 \eta_{1,r} + c_1 \eta_{2,r} \) is distributed by the normal law with zero mean and variance \( c_0^2 \sigma^2_{11,r}(\beta, \lambda) + 2c_0 c_1 \sigma_{12,r}(\beta, \lambda) + c_1^2 \sigma^2_{22,r}(\beta, \lambda) \). \( \blacksquare \)

We omit the proof of Theorem 3, because it is similar to that of Theorem 2.

**Proof of Proposition 1.** Set \( J(y) = F^-_{a,\beta}(1 - y) - \tilde{F}^-_{a,\beta}(1 - y), \ 0 < y < 1 \). We have

\[
\frac{dJ(y)}{dy} = \frac{\theta}{y} \left( 1 - \frac{c}{y^{1/\beta}} \right).
\]

This readily implies that for \( c \geq 1 \) the function \( J(y) \) strictly increases on \((0,1)\). Together with \( J(1) = 0 \), this yields \( J(y) > 0 \) for \( 0 < y < 1 \). Thus, in the case of \( c \geq 1 \) we have \( V \overset{d}{=} \tilde{F}^-_{a,\beta}(1 - U) \), or, equivalently, the random variable \( V \) has the biased exponential distribution function \( F_{\beta}(x - a), \ x \geq a \).

Let \( 0 < c < 1 \). The equation \( J(y) = 0 \) can be rewritten as

\[
-cy^{-1/\beta} \exp \{-cy^{-1/\beta}\} = -c \exp \{-c\}. \tag{A.7}
\]

Since \(-1/e < -c \exp \{-c\} < 0 \) for each \( 0 < c < 1 \), we see that Eq. (A.7) has two roots

\[
y_0 = \left( -\frac{1}{c} W_{-1}(c \exp \{-c\}) \right)^{-\beta} \quad \text{and} \quad y_1 = \left( -\frac{1}{c} W_0(-c \exp \{-c\}) \right)^{-\beta}.
\]

**AUTOMATION AND REMOTE CONTROL**  Vol. 82  No. 8  2021
Applying the identity $W_0(-c \exp(-c)) = -c$, we obtain $y_1 = 1$. Using the fact that $W_{-1}(x)$, $-1/e < x < 0$, is strictly nondecreasing, one can readily verify that $J(y) > 0$ for $0 < y < y_0$ and $J(y) < 0$ for $y_0 < y < 1$. Then for $x \geq a$ we have

$$P(V \leq x) = P\left(F_{a,\theta}^\rightarrow(1-U) \leq x, \ 0 < U < y_0\right) + P\left(F_{a,\beta}^\rightarrow(1-U) \leq x, \ y_0 \leq U < 1\right)$$

$$= P\left(0 < U < y_0, \ 1 - F_{a,\theta}(x) \leq U < 1\right) + P\left(\max \{y_0, 1 - F_{a,\beta}(x)\} \leq U < 1\right).$$

The proof of Proposition 1 can be completed by noticing that

$$F_{a,\theta}(1-y) = F_{a,\beta}(1-y) = -(a/c)W_{-1}(-c \exp(-c))$$

and using

$$P\left(0 < U < y_0, \ 1 - F_{a,\theta}(x) \leq U < 1\right) = \begin{cases} 0, & x < F_{a,\theta}(1-y) \\ y_0 - 1 + F_{a,\theta}(x), & x \geq F_{a,\theta}(1-y), \end{cases}$$

$$P\left(\max \{y_0, 1 - F_{a,\beta}(x)\} \leq U < 1\right) = \begin{cases} F_{a,\beta}(x), & x < F_{a,\beta}(1-y) \\ 1 - y_0, & x \geq F_{a,\beta}(1-y). \end{cases}$$

**FUNDING**

The research by the second author was partly supported by a grant from the Russian Foundation for Basic Research, project no. 19-01-00090.

**REFERENCES**

1. Pareto, V., The new theories of economics, *J. Political Econ.*, 1897, vol. 5, no. 4, pp. 485–502.
2. Vere-Jones, D., Robinson, R., and Yang, W.Z., Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation, *Geophys. J. Int.*, 2001, vol. 144, pp. 517–531.
3. Kagan, Y.Y., Seismic moment distribution revisited: I. Statistical results, *Geophys. J. Int.*, 2002, vol. 148, pp. 520–541.
4. Kagan, Y.Y. and Schoenberg, F., Estimation of the upper cutoff parameter for the tapered Pareto distribution, *J. Appl. Probab.*, 2001, vol. 38A, pp. 158–175.
5. Foss, S., Korshunov, D., and Zachary, S., *An Introduction to Heavy-Tailed and Subexponential Distributions*, New York: Springer, 2011.
6. Cumming, S.G., A parametric model of the fire-size distribution, *Can. J. Forest Res.*, 2001, vol. 31, no. 8, pp. 1297–1303.
7. Schoenberg, F.P., Peng, R., and Woods, J., On the distribution of wildfire sizes, *Environmetrics*, 2003, vol. 14, pp. 583–592.
8. Swetnam, T.L., Falk, D.A., Lynch, A.M., and Yool, S.R., Estimating individual tree mid- and understory rank-size distributions from airborne laser scanning in semi-arid forests, *Forest Ecol. Manage.*, 2014, vol. 330, pp. 271–282.
9. Meerschaert, M.M., Roy, P., and Shao, Q., Parameter estimation for tempered power law distributions, *Commun. Stat. Theory Methods*, 2012, vol. 41, pp. 1839–1856.
10. Jameson, G.J.O., The incomplete gamma functions, *Math. Gaz.*, 2016, vol. 100, pp. 298–306.
11. Baricz, Á. and Ismail, M.E.H., Turán type inequalities for Tricomi confluent hypergeometric functions, *Constr. Approximation*, 2013, vol. 37, pp. 195–221.
12. Beals, R. and Szmigielski, J., Meijer G-functions: a gentle introduction, Notices Am. Math. Soc., 2013, vol. 60, pp. 866–872.

13. Devroye, L., Non-Uniform Random Variate Generation, New York: Springer-Verlag, 1986.

14. Geist, E.L. and Parsons, T., Undersampling power-law size distributions: effect on the assessment of extreme natural hazards, Nat. Hazards, 2014, vol. 72, pp. 565–595.

15. Corless, R.M., Gonner, G.H., Hare, D.E., Jeffrey, D.J., and Knuth, D.E., On the Lambert W function, Adv. Comput. Math., 1996, vol. 5, pp. 329–359.

16. Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N., A First Course in Order Statistics, New York: Wiley, 1992.

17. Bilingsley, P., Probability and Measure, New York: Wiley, 1995.

18. Van der Vaart, A.W., Asymptotic Statistics, Cambridge: Cambridge Univ. Press, 2000.

This paper was recommended for publication by A.I. Mikhal’skii, a member of the Editorial Board