On discriminating between lognormal and Pareto tail: an unsupervised mixture-based approach

Marco Bee¹

Received: 13 November 2020 / Revised: 3 February 2022 / Accepted: 17 February 2022
© The Author(s) 2022, corrected publication 2022

Abstract

Many stochastic models in economics and finance are described by distributions with a lognormal body. Testing for a possible Pareto tail and estimating the parameters of the Pareto distribution in these models is an important topic. Although the problem has been extensively studied in the literature, most applications are characterized by some weaknesses. We propose a method that exploits all the available information by taking into account the data generating process of the whole population. After estimating a lognormal–Pareto mixture with a known threshold via the EM algorithm, we exploit this result to develop an unsupervised tail estimation approach based on the maximization of the profile likelihood function. Monte Carlo experiments and two empirical applications to the size of US metropolitan areas and of firms in an Italian district confirm that the proposed method works well and outperforms two commonly used techniques. Simulation results are available in an online supplementary appendix.

Keywords Mixture distributions · EM algorithm · Profile likelihood · Unsupervised tail estimation

Mathematics Subject Classification 62H30 · 91B70 · 62P20 · 91-10

1 Introduction

The lognormal and the Pareto distribution are common models for various phenomena in nature and in the social sciences. Besides empirical evidence, there are strong theoretical reasons supporting the use of these two distributions.

All the codes for simulating and estimating a lognormal–Pareto mixture using the methods developed in this paper are available in the `LNPar` R package, available at <https://github.com/marco-bee/LNPar>.

✉ Marco Bee
marco.bee@unitn.it

¹ Department of Economics and Management, University of Trento, Via Inama, 5, 38122 Trento, Italy

The lognormal distribution is often explained by the so-called Gibrat's law. Many growth processes result from the accumulation of a large number of small percentage changes. Under fairly mild conditions, the sum of the logarithms of the changes is closer to the normal than that of the summands, because of the central limit theorem. Hence, on the original scale, the distribution of sizes is approximately lognormal (Kleiber and Kotz 2003, Sect. 4.2).

On the other hand, the Pareto distribution, originally used as a model of income and wealth, is the prototypical heavy-tailed distribution, and has found important applications for the estimation of rare event probabilities in hydrology and finance (Kleiber and Kotz 2003, Chap. 3). Especially important is the fact that proportional random growth yields a power-law tail, so that there are strong theoretical motivations that support the existence of power-laws in fields as diverse as size of cities and firms, financial returns, volume of international trade and executive salaries; see Gabaix (2009). The estimation of the shape parameter is of paramount importance, in particular for economic applications (Reed 2001; Gabaix 2009). More generally, the Pareto distribution plays a key role in Extreme Value Theory: see, e.g., Gomes and Guillou (2015).

The two cases sketched above are encompassed by a lognormal–Pareto mixture, whose density is given by

$$f(x; \boldsymbol{\theta}) = \pi f_1(x; \mu, \sigma^2) + (1 - \pi) f_2(x; x_{\min}, \alpha), \quad (1)$$

where $x \in \mathbb{R}^+$, $\boldsymbol{\theta} = (\pi, \mu, \sigma^2, x_{\min}, \alpha)' \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta} = [0, 1] \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$, $f_1(x; \mu, \sigma^2)$ is the probability density function (pdf) of a $\text{Logn}(\mu, \sigma^2)$ random variable and $f_2(x; x_{\min}, \alpha)$ is the pdf of a $\text{Pareto}(x_{\min}, \alpha)$ random variable. In terms of sampling, each observation comes from a $\text{Logn}(\mu, \sigma^2)$ population (P_1 , say) with probability π or from a $\text{Pareto}(x_{\min}, \alpha)$ population (P_2) with probability $1 - \pi$. The pure lognormal and pure Pareto case are obtained when $\pi = 1$ and $\pi = 0$, respectively, however the latter is typically ruled out a priori. Hence, the distribution is always assumed to have a lognormal body. For statistical purposes, the observations available are a random sample from (1). Recently, Benzidia and Lubrano (2020) have fitted (1) in a Bayesian setup to study the distribution of American academic wages; a discussion of (1) and of possible alternative models can be found in Klugman et al. (2004, Sect. 4.4.7) and Abdul Majid and Ibrahim (2021).

Most investigations in the literature focus on two related issues: using hypothesis testing or goodness-of-fit techniques to discern whether the tail of the distribution is lognormal or Pareto, and estimating the parameters of the latter. In applications, Eeckhout (2004, 2009), Levy (2009), Malevergne et al. (2009), Rozenfeld et al. (2011), Berry and Okulicz-Kozaryn (2012), Hsu (2012), Ioannides and Skouras (2013), González-Val et al. (2015), Fazio and Modica (2015) and D'Acci (2019) deal with city size distribution, whereas Axtell (2001), Di Giovanni et al. (2011), Tang (2015), Bee et al. (2017) and Kondo et al. (2021) are concerned with firm size. In so doing, the majority of these approaches do not take into account the distribution of the entire population and employ only the observations in the tail, with a substantial loss of information. Hence, they are based on the implicit assumption that, if there is a Pareto tail above some threshold $x_{\min}$ representing the scale parameter of the Pareto distribution, all the observations above $x_{\min}$ are Pareto distributed. However, whereas all

the observations below x_{min} are lognormal, not all the observations above x_{min} are Pareto, since the supports of the lognormal and Pareto distributions overlap.

In this paper we develop a framework that explicitly takes into account the lognormal–Pareto mixture distribution of the population. In particular, our likelihood-based procedure starts by assuming that x_{min} is known, and estimates the parameters using the EM algorithm (Dempster et al. 1977). Then, this hypothesis is relaxed, and estimation is performed by maximizing the profile likelihood function based on all the observations.

The contribution of this work is twofold. From a statistical point of view, even though (1) is a special case of the dynamic mixture introduced by Frigessi et al. (2002), the properties of (1) are not studied in their paper. We carry out statistical analysis of (1) and develop an estimation procedure based on the EM algorithm and the maximization of the profile likelihood, which is different from the method used by Frigessi et al. (2002). In addition, we propose a log-likelihood ratio (llr) test for discriminating between lognormal and Pareto tail. Second, in practical applications, we point out that the commonly used methods for detection and estimation of a Pareto tail implicitly assume no overlap of the two distributions and do not use all the available information. Accordingly, simulation experiments and real-data analyses suggest that our method is preferable.

The rest of the paper is organized as follows. In Sect. 2 we detail the EM algorithm for estimating the parameters, the profile likelihood approach employed when x_{min} is unknown and the llr test for the null hypothesis of no Pareto tail. In Sect. 3 we discuss the results of simulation experiments aimed at assessing the small-sample properties of the estimators. In Sect. 4 we illustrate two applications to the size distribution of US metropolitan areas and of firms in an Italian district. Conclusions and open problems are discussed in Sect. 5. Simulation outcomes are available in an online supplementary appendix.

2 A mixture-based approach

Likelihood-based estimation is different according to whether or not x_{min} is known. We study the two cases in Sects. 2.2 and 2.3 respectively.

Let $\mathbf{x} = (x_1, \dots, x_n)'$ be a random sample from a random variable X with density (1). This distribution is related to the dynamic mixture of Frigessi et al. (2002), who propose a model defined by the following pdf:

$$f(x) = \frac{1 - p(x; \boldsymbol{\theta})f_1(x; \boldsymbol{\beta}) + p(x; \boldsymbol{\theta})f_2(x; \xi, \sigma)}{Z(\boldsymbol{\theta}, \boldsymbol{\beta}, \xi, \sigma)}, \quad (2)$$

where f_2 is the Generalized Pareto (GP) pdf and f_1 is some other pdf. They show that (2) can be rewritten as a pure two-population mixture model (Frigessi et al. 2002, p. 223). As for $p(x; \boldsymbol{\theta})$, they propose $p(x; \boldsymbol{\theta}) = 1/2 + (1/\pi) \arctan[(x - \mu)/\tau]$, but mention that it is also possible to use the Heaviside function (Frigessi et al. 2002, p. 222), which would make their model identical to (1), provided the component densities are the same.

In the present paper, the use of a fixed mixing weight π instead of a function depending on x is motivated by two main reasons. First, as outlined in Sect. 1, in economic applications one usually fits (1) in order to choose between Gibrat's law and the power-law theories, which imply a pure lognormal or a pure Pareto tail, respectively. In the latter case there is a threshold where the Pareto distribution starts, whereas the general version of Frigessi et al. (2002) model, with the mixing weight depending on x , does not have such a cut-off, and therefore never identifies a pure power-law. In other words, the extra flexibility of the Frigessi et al. (2002) approach does not match the theoretical properties of the models typically used in economic and financial applications.

Second, Frigessi et al. (2002) point out that a possible disadvantage of (1) is the discontinuity of the density at x_{min} . Indeed, most composite Pareto models in the literature are constructed such that the resulting pdf is continuous or even smooth (see, e.g., Scollnik 2007 or Abu Bakar et al. 2015, and the references therein). However, the number of free parameters in (1) is reduced by one if the continuity condition is enforced, and by two if a smooth pdf is desired, so that flexibility is reduced; see Scollnik (2007) for details. Furthermore, even though a smooth pdf may be preferable from a statistical point of view, in many applications, including the present setup, there is no theoretical reason to assume a population with continuous and smooth distribution. See Benzidia and Lubrano (2020) for a similar empirical application, and Klugman et al. (2004, Sect. 4.4.7) for an explicit treatment of a discontinuous composite model.

2.1 Background: the EM algorithm

MLE of finite mixture distributions is usually carried out by means of the EM algorithm (Dempster et al. 1977; McLachlan and Krishnan 2008). Let $f(\mathbf{x}; \boldsymbol{\theta})$ and $\ell(\boldsymbol{\theta})$ be the observed density and log-likelihood function respectively, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ a vector of parameters. Let $\mathbf{z}$ be the missing data vector, and $\mathbf{v} = (\mathbf{x}', \mathbf{z}')'$ be the complete data, with density and complete log-likelihood denoted by $g_c(\mathbf{v}; \boldsymbol{\theta})$ and $\ell_c(\boldsymbol{\theta})$ respectively. In a mixture setup, the missing data are the indicator variables of population membership, denoted by $z_{ij} = \mathbb{1}_{\{x_j \in P_j\}}$, where $\mathbb{1}_A$ is the indicator function of the set A and P_j is the j -th population. At the t -th iteration, the algorithm iterates two main steps. *E-step* Compute the conditional expectation of $\ell_c(\boldsymbol{\theta})$, given the current value of $\boldsymbol{\theta}$ and the observed sample $\mathbf{x}$:

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) \stackrel{\text{def}}{=} E_{\boldsymbol{\theta}^{(t)}} \{\ell_c(\boldsymbol{\theta}) | \mathbf{x}\}. \quad (3)$$

M-step Maximize, with respect to $\boldsymbol{\theta}$, the conditional expectation of $\ell_c(\boldsymbol{\theta})$ provided by the E-step:

$$\boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}). \quad (4)$$

The procedure is then repeated, i.e. (3) and (4) are recomputed using $\boldsymbol{\theta}^{(t+1)}$ in place of $\boldsymbol{\theta}^{(t)}$ until some convergence criterion is met.

The EM algorithm monotonically increases the observed likelihood at each iteration; moreover, under regularity conditions (Wu 1983), it converges to a stationary point and the estimators are asymptotically efficient.

The choice of starting values for the EM algorithm is a long-standing problem. Even in the most studied case, i.e. normal mixture estimation, there is no optimal technique (Biernacki et al. 2003; McLachlan and Krishnan 2008). As for our setup, in the known-threshold case there are “obvious” choices for all parameters: assuming that the observations are ordered in ascending order, let $n_{\log n} = \#\{x_i : x_i < x_{\min}\}$ and $n_{x_{\min}} = n - n_{\log n}$. Then we set

$$\begin{aligned}\pi^{(0)} &= n_{\log n}/n; \quad \mu^{(0)} = \frac{1}{n_{\log n}} \sum_{i=1}^{n_{\log n}} \log(x_i); \\ \psi^{(0)} &= \frac{1}{n_{\log n}} \sum_{i=1}^{n_{\log n}} \log(x_i)^2 - (\mu^{(0)})^2; \quad \alpha^{(0)} = \frac{n_{x_{\min}}}{\sum_{i=n_{\log n}+1}^n \log(x_i/x_{\min})},\end{aligned}$$

where $\psi \stackrel{\text{def}}{=} \sigma^2$. It is readily seen that these initial values are the MLEs under the assumption of complete separation of the two populations. Clearly, this is in general not true, but in most cases yields a good approximation of the true values. Since the algorithm may converge to a global or local maximizer, or to a saddle point, in practice it is recommended to use different initializations and keep track of the maximized value of the likelihood function (McLachlan and Krishnan 2008, Chap. 3). In the present case, we know in advance that $\mu^{(0)} \leq \mu$ and $\psi^{(0)} \leq \psi$, with equality only for completely separated populations. Hence, in these two cases, different starting values are always computed using positive random perturbations.

The unknown-threshold setup is more complicated. One possibility is to first guess the approximate value of $x_{\min}$, and then proceed as in the known-threshold case. If the approximate value of $x_{\min}$ does not seem to be reliable, we use a large number of random perturbations of the starting values. It will be shown in Sect. 2.2 that in the present setup the likelihood function is well-behaved and has no singularities when $\sigma \rightarrow 0$. Hence, the main problem is likely to be the convergence to a local maximum.

The algorithm is usually stopped when either the difference of the log-likelihood function values or the maximum absolute value in the vector of differences of the parameter estimators in two consecutive iterations is smaller than some predefined small tolerance. In this paper we use the latter approach (see Flury 1997, p. 267).

2.2 Known threshold

If we assume that $x_{\min}$ in (1) is known, the parameter vector is $\phi = (\pi, \mu, \sigma^2, \alpha)' \in \Phi$, $\Phi = [0, 1] \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$. Given a random sample $x_1, \dots, x_n$ from (1), the observed and complete log-likelihood functions are respectively equal to

$$\ell(\phi; \mathbf{x}) = \sum_{i=1}^n \log\{\pi f_1(x_i; \mu, \sigma^2) + (1 - \pi) f_2(x_i; x_{\min}, \alpha)\},$$

$$\ell_c(\boldsymbol{\phi}; \mathbf{x}, \mathbf{Z}) = \sum_{i=1}^n \{z_{i1} \log[\pi f_1(x_i; \mu, \sigma^2)] + z_{i2} \log[(1 - \pi) f_2(x_i; x_{\min}, \alpha)]\}, \quad (5)$$

where the missing-data matrix $\mathbf{Z}$ is an $n \times 2$ matrix whose (i, j) -th element z_{ij} is equal to 1 if the i -th observation belongs to the j -th population and zero otherwise ($i = 1, \dots, n, j = 1, 2$).

Unlike the typical mixture setup, here some of the z_{ij} s are known, because the observations smaller than $x_{\min}$ belong to the lognormal population: assuming $x_1 \leq \dots \leq x_n$, one has $z_{i1} = 1$ for $i = 1, \dots, n_{\log n}$. In normal mixture analysis, this framework is known as normal theory discrimination with partially classified data (Flury 1997, Sect. 9.5).

In the current setup the E-step (3) amounts to computing the posterior probabilities of the observations:

$$\begin{aligned} \tau_{i1}^{(t)} &= \begin{cases} 1 & i = 1, \dots, n_{\log n}, \\ \mathbb{E}_{\phi^{(t)}}(z_{i1}) = \frac{\pi f_1(x_i; \mu^{(t)}, \psi^{(t)})}{\pi f_1(x_i; \mu^{(t)}, \psi^{(t)}) + (1 - \pi) f_2(x_i; x_{\min}, \alpha^{(t)})}, & i = n_{\log n} + 1, \dots, n, \end{cases} \\ \tau_{i2}^{(t)} &= 1 - \tau_{i1}^{(t)}. \end{aligned} \quad (6)$$

The M-step for π is equal to the mean of the posterior probabilities (see, e.g., Titterton et al. 1985, Sect. 4.3.2). As for the remaining parameters, we exploit the complete data MLEs of the lognormal and of the Pareto distribution:

$$\pi^{(t)} = \frac{1}{n} \sum_{i=1}^n \tau_{i1}^{(t)}; \quad (7)$$

$$\mu^{(t)} = \frac{1}{n\pi^{(t)}} \sum_{i=1}^n \tau_{i1}^{(t)} \log(x_i); \quad (8)$$

$$\psi^{(t)} = \frac{1}{n\pi^{(t)}} \sum_{i=1}^n \tau_{i1}^{(t)} \log(x_i)^2 - (\mu^{(t)})^2; \quad (9)$$

$$\alpha^{(t)} = \frac{n(1 - \pi^{(t)})}{\sum_{i=n_{\log n}+1}^n \tau_{i2}^{(t)} \log(x_i/x_{\min})}. \quad (10)$$

We obtain the MLEs by iterating (6) to (10) until the stopping rule $\max_{i=1, \dots, 4} |\hat{\phi}_i| < 10^{-10}$ is satisfied.

Starting values are computed as follows: $p_0 = n_{\log n}/n$, $\mu_0 = \sum_{i=1}^n \log x_i/n$, $\psi_0 = \text{var}(\log x_i)$ and $\alpha_0 = n_{x_{\min}}/(\sum_{i=n_{\log n}+1}^n \log x_i/x_{\min})$. Moreover, we double-check the results with small random perturbations of the initial values.

Three two-dimensional representations of the observed log-likelihood (5) are displayed in Fig. 1. The contour plots are based on 1000 observations simulated from a lognormal–Pareto mixture with $\pi = 0.5$, $\mu = 0$, $\sigma = 1$, $x_{\min} = 5$, $\alpha = 1.5$. From left to right, the likelihood is displayed as a function of (α, μ) , (α, σ) , (σ, μ) respectively, with the remaining parameters being kept fixed at their true values.

Fig. 1 Two-dimensional representations of the observed log-likelihood as a function of (α, μ) (left), (α, σ) (center), (σ, μ) (right)

Table 1 Log-density values and value of the log-likelihood in Example 1

σ	$\log f(y_1)$	$\log f(y_2)$	$\log f(y_3)$	$\log f(y_4)$	$\log f(y_5)$	$\ell(\mathbf{y})$
0.1	2.443	8.032×10^{-11}	9.514×10^{-2}	7.407×10^{-2}	9.660×10^{-3}	-31.947
0.05	4.887	4.453×10^{-41}	9.514×10^{-2}	7.407×10^{-2}	9.660×10^{-3}	-100.921
0.02	12.217	1.399×10^{-254}	9.514×10^{-2}	7.407×10^{-2}	9.660×10^{-3}	-591.613
0.01	24.434	0	9.514×10^{-2}	7.407×10^{-2}	9.660×10^{-3}	-Inf

It is well known that the likelihood function of mixtures of normal distributions with different variances is ridged with singularities, typically observed when the mean of one component is equal to one observation and the variance of the component goes to 0 (Flury 1997, p. 652). However, this is not the case for lognormal–Pareto mixtures, as the following example shows.

Example 1 Five observations are sampled from the lognormal–Pareto mixture with parameters $\pi = 0.5$ with $\pi = 0.5$, $\mu = 0$, $\sigma = 1$, $x_{\min} = 5$, $\alpha = 1.5$. The log-values are $\mathbf{y} = (-0.203, 0.482, 1.792, 1.892, 2.707)'$. Table 1 displays the values of the log-density at each data point when $\mu = -0.203$, for various values of σ and the remaining three parameters equal to their true values. The value of the log-likelihood is shown as well.

Similarly to the normal mixture example in Flury (1997, p. 652), $\log f(y_1) \xrightarrow{\sigma \rightarrow 0} \infty$, but this effect is more than counterbalanced by the fact that $\log f(y_i) \xrightarrow{\sigma \rightarrow 0} -\infty$, $\forall y_i \neq$

$y_1, y_i < x_{min}$ (only y_2 , in this case). Hence, except for pathological cases when only one observation is smaller than the threshold, the log-likelihood function has no singularities when $\sigma \rightarrow 0$.

2.3 Unknown threshold

Relaxing the hypothesis of known threshold makes the estimation procedure more complex: complete-data MLE can only be carried out if class membership is known, but now class membership depends on x_{min} . In other words, splitting the observations according to the amount of information they contain about the parameters is impossible when x_{min} is unknown (Frigessi et al. 2002, p. 223). Hence, it is unfeasible to incorporate the estimation of the additional parameter x_{min} in the EM algorithm.

However, one can still exploit the methodology of Sect. 2.2 and develop a conditioning argument. This is the main motivation for the profile likelihood approach presented in this section: since, given a fixed value of x_{min} , the MLEs of the remaining parameters can be found as in Sect. 2.2, we use the EM algorithm to estimate (1) n times, conditionally on $x_{min} = x_n, x_{min} = x_{n-1}, \dots, x_{min} = x_1$. The function

$$g(x_{min}) : x_{min} \rightarrow \max_{\phi} \ell(\phi; \mathbf{x}), \quad (11)$$

which maps x_{min} to the maximized log-likelihood (5), is called profile log-likelihood function. Accordingly, the profile MLE of the threshold is the maximizer of (11):

$$\hat{x}_{min} = \arg \max_{x_{min}=x_1, \dots, x_n} g(x_{min}).$$

Having estimated the parameters, there is a Pareto tail if and only if $\hat{\pi}_2 \stackrel{\text{def}}{=} 1 - \hat{\pi}$ is larger than zero. Since $\hat{\pi}_2$ is the fitted fraction of Pareto observations, the estimated number of Pareto observations is $n\hat{\pi}_2$. Note that, in general, $n\hat{\pi}_2 < \hat{n}_{x_{min}}$, and $\hat{n}_{x_{min}} - n\hat{\pi}_2$ is equal to $\#\{x_i : x_i > \hat{x}_{min}, x_i \in P_1\}$, the number of lognormal observations larger than the estimated threshold. The estimate $\hat{x}_{min}$ only allows us to identify the observations that, according to the fitted distribution, are certainly not Pareto, since the observations $x_i : x_i \leq x_{min}$ belong to the lognormal population with probability 1.

The EM-based estimation method outlined above can in principle be extended to the case of $K > 2$ populations, possibly with different component densities. The process is likely to become computationally more demanding, because of the increased number of parameters. Even though this issue may be theoretically and numerically interesting, we leave it open to further investigations, since mixtures with more than two components are not relevant for most practical applications in economics and finance.

2.4 Testing for a Pareto tail

Since we use maximum likelihood for estimation, a test of the hypothesis $H_0 : \pi = 1$, which corresponds to the “no Pareto tail” case, can be constructed by means of the llr method. In a mixture setup, the classical llr asymptotic theory breaks down (Self and Liang 1987); hence it is necessary to find the null distribution via simulation. This is not difficult, but computationally expensive for large sample size. A pseudo-code is as follows:

1. Given data $\mathbf{x}$, compute $\hat{\ell}_0 \stackrel{\text{def}}{=} \max \ell_0(\mu, \sigma^2; \mathbf{x})$ and $\hat{\ell} \stackrel{\text{def}}{=} \max \ell(\boldsymbol{\theta}; \mathbf{x})$, where $\ell_0(\mu, \sigma^2; \mathbf{x})$ is the lognormal log-likelihood (i.e., the log-likelihood under the null hypothesis), and $\ell(\boldsymbol{\theta}; \mathbf{x})$ is the log-likelihood corresponding to (1). Compute $t_{obs} = -2 \log(\hat{\ell}_0/\hat{\ell})$.
2. Sample $(x_1^s, \dots, x_n^s)'$ under the null hypothesis, i.e. from the $\text{Logn}(\hat{\mu}_{obs}, \hat{\sigma}_{obs}^2)$ distribution, where $\hat{\mu}_{obs}$ and $\hat{\sigma}_{obs}^2$ are the lognormal MLEs based on the observed data. Use $(x_1^s, \dots, x_n^s)'$ to compute $t_{sim} = -2 \log(\hat{\ell}_0^s/\hat{\ell}^s)$.
3. Repeat steps 1-2 a large number of times B and let $\mathbf{t}_{sim} = (t_{sim,1}, \dots, t_{sim,B})'$. The p -value is $p = \#\{t_{sim,i} : t_{sim,i} > t_{obs}\}/B, i = 1, \dots, B$.

3 Simulation experiments

3.1 Known threshold

We first perform Monte Carlo experiments aimed at assessing the performance of the EM algorithm for various sample sizes when x_{min} is assumed to be known. Hence, we calculate the bias and mean-squared-error (MSE) of the MLEs of π, α, μ and σ^2 . A description of the procedure is given below.

1. For $i = 1 \dots, B$, simulate n observations from (1) and estimate the parameters π, α, μ and σ^2 by means of the EM algorithm (see Sect. 2.2);
2. Compute the bias and MSE of the B estimates of π, α, μ and σ^2 obtained at Step 1.

We use $B = 500$ and sample sizes $n = 100, 200, 500, 1000, 10\,000$. We repeat the preceding steps for 6 different parameter configurations, with $\mu = 0, \sigma^2 = 1$ and the remaining parameters reported below:

1. $\pi = 0.9, \alpha = 1$; 2. $\pi = 0.95, \alpha = 1$; 3. $\pi = 0.9, \alpha = 1.5$;
4. $\pi = 0.95, \alpha = 1.5$; 5. $\pi = 0.9, \alpha = 2$; 6. $\pi = 0.95, \alpha = 2$.

In all cases the starting values are computed as described at the end of Sect. 2.2. Small changes of the initial values have produced the same results. All the plots are reported in the online supplementary material.

The absolute bias and the MSE of the parameters are displayed in Figs. A.1 and A.2 respectively, with axes on logarithmic scale. Both measures are larger for $\hat{\alpha}$ than for the remaining parameters, especially when the sample size is small. This is not surprising, since the true value of π is close to 1, and hence only few observations are sampled from the Pareto distribution: for example, when $n = 50$ and $\pi = 0.95$, the average

number of observations sampled from the Pareto distribution is $n(1 - \pi) = 2.5$. As an additional information, not seen in Fig. A.1 since it shows the absolute value, we point out that the bias of $\hat{\alpha}$ is always positive. The MSE of each parameter in Fig. A.2 decreases approximately linearly on logarithmic scale. Similarly to the bias, the MSE is larger for $\hat{\alpha}$ than for $\hat{\pi}$, $\hat{\mu}$ and $\hat{\sigma}^2$.

3.2 Unknown threshold

In this setup, we assess the presence of a Pareto tail using two other approaches, namely the Uniformly Most Powerful Unbiased (UMPU) test of Malevergne et al. (2009) and the Clauset et al. (2009) method.

The UMPU test for the null hypothesis of exponentiality against the alternative of truncated normality is given by the clipped sample coefficient of variation $\bar{c} = \min\{1, \hat{\sigma}/\hat{\mu}\}$ (Del Castillo and Puig 1999), where $\hat{\mu}$ and $\hat{\sigma}$ are the estimated parameters of the truncated normal. Assuming $x_1 \leq \dots \leq x_n$, the test is recursively performed on sets $A_1 = \{x_n\}$, $A_2 = \{x_n, x_{n-1}\}$, ..., $A_r = \{x_n, x_{n-1}, \dots, x_{n-r+1}\}$, ... The estimate is $\hat{x}_{min} = x_{\hat{r}}$, where $\hat{r}$, the rank of the estimated threshold, is such that H_0 is accepted for $A_{\hat{r}}$ and rejected for $A_{\hat{r}+1}$. The estimate of α is the Pareto MLE based on the observations larger than $\hat{x}_{min}$:

$$\hat{\alpha} = \frac{\hat{n}_{x_{min}}}{\sum_{i=n_{logn}+1}^n \log(x_i/\hat{x}_{min})}, \quad (12)$$

where $\hat{n}_{x_{min}} = n - \hat{n}_{logn}$. This test is computationally simple and theoretically appealing, but is designed for truncated distributions. Hence, when an untruncated sample is available and the data-generating process is (1), it does not exploit all the available information.

Clauset et al. (2009) (CSN) propose a method based on the Kolmogorov-Smirnov (KS) distance. The estimated x_{min} is the value that minimizes the KS distance $D = \max_{x \geq x_{min}} |F_n(x) - F_P(x)|$ between the empirical (F_n) and the Pareto (F_P) cumulative distribution functions. Analogously to the UMPU method, the scaling parameter α is estimated via maximum likelihood using (12) and all the observations larger than the estimated threshold.

All the results are based on the following simulation experiment.

- Simulate $x_1, \dots, x_n$ from (1).
- Set $x_{min} = x_1$, estimate ϕ via the EM algorithm and compute the corresponding maximized log-likelihood $\ell(\hat{\phi}; \mathbf{x})$.
- Repeat step (b) with $x_{min} = x_2, \dots, x_{min} = x_n$.
- Set $\hat{x}_{min} = \arg \max_{x_{min}=x_1, \dots, x_n} g(x_{min})$, where g is given by (11); estimate π, μ, σ^2 and α via the EM algorithm with $x_{min} = \hat{x}_{min}$.
- Estimate x_{min} and α via the UMPU and CSN methods.

In addition to the six setups considered in the known-threshold case in Sect. 3.1, we include four frameworks, numbered 7–10, with true values of parameters motivated by the estimates in the empirical application presented in Sect. 4 below:

7–10. $\pi = 0.5$, $\mu = 12$, $\sigma^2 = 0.12$, $\alpha \in \{0.8, 1, 1.5, 2\}$, $x_{min} = 260\,000$.

As for UMPU, we use a 1% significance level. Unreported simulations at larger significance levels yield worse outcomes.

For setups 1–6, Figs. A.3 and A.4 respectively show the absolute value of the bias and the MSE of $\hat{x}_{min}$, whereas Figs. A.5 and A.6 display the same quantities for $\hat{\alpha}$. All the graphs are on doubly logarithmic scale.

As concerns the estimation of x_{min} , Figs. A.3 and A.4 clearly show that the performance of the mixture-based method proposed in this paper (MIX from now on) is considerably better than UMPU and CSN both in terms of bias and of MSE. This is true for all sample sizes, but more evident when $n \geq 5000$. As for the estimation of α (see Figs. A.5 and A.6), the outcomes are similar: both the absolute bias and the MSE of $\hat{\alpha}$ estimated via MIX are better when the sample size increases, but MIX is the best method also for small sample sizes.

The graphs for setups 7–10, displayed in Figs. A.7–A.10, are similar to Figs. A.3–A.6: MIX is always better for x_{min} , and in most cases for α , with the exception of the smallest sample sizes when $\alpha \leq 1$.

In comparative terms, UMPU and CSN have a better performance when the sample size is small. Why is this the case? They use all the observations larger than $\hat{x}_{min}$ for estimating α : although many of these observations are lognormal, they may contribute to decreasing the bias and MSE of the estimates with respect to MIX, whose estimate of α is mostly based on a much smaller number of Pareto observations with τ_{i2} close to 1. Of course, the effect of the lognormal observations on the UMPU and CSN estimators cannot be rigorously assessed and is only advantageous in the special case with large π and small n , so that it disappears as soon as the sample size becomes moderately large. This remark is supported by our Monte Carlo results: as the sample size increases, the MSE of $\hat{\alpha}$ decreases more quickly when π is smaller: compare, e.g., the two panels in the bottom line of Fig. A.6 and the last panel of Fig. A.10.

Finally, Figs. A.11–A.18 display the bias and MSE of the estimators of π , μ , σ^2 and α when x_{min} is known ($\hat{\pi}_K$, $\hat{\mu}_K$, $\hat{\sigma}_K^2$ and $\hat{\alpha}_K$) and when it is unknown ($\hat{\pi}_U$, $\hat{\mu}_U$, $\hat{\sigma}_U^2$ and $\hat{\alpha}_U$). In terms of bias, the outcomes are similar: only the absolute value of the bias of $\hat{\alpha}$ is slightly smaller in the known-threshold setup. On the other hand, the difference of the MSEs is related to the sample size: when $n < 1000$, the MSE is smaller for all parameters in the known-threshold case, when $n \geq 5000$ the MSEs are essentially identical. The difference between the two cases when $n \leq 1000$ is larger for $\hat{\alpha}$ than for the remaining estimators.

4 Empirical analysis

4.1 US metropolitan areas

In this application we consider the 2019 population estimate of the 415 US metropolitan areas computed by the US Census Bureau.¹ As mentioned in Sect. 1, for data of

¹ www.census.gov/data/datasets/time-series/demo/popest/2010s-total-metro-and-micro-statistical-areas.html.

Table 2 Estimated parameter values and number of Pareto observations using the MIX, UMPU and CSN approach

	MIX	UMPU 1%	CSN
$\hat{\pi}$	0.5167 (0.0787) [0.3810, 0.6846]	— — —	— — —
$\hat{\mu}$	13.2602 (0.1080) [13.0783, 13.4735]	— — —	— — —
$\hat{\sigma}^2$	1.5743 (0.1741) [1.2205, 1.9034]	— — —	— — —
$\hat{\alpha}$	1.1207 (0.2950) [0.8814, 1.9234]	1.1869 (0.4172) [0.9628, 2.6631]	0.7036 (0.2307) [0.6590, 1.7145]
$\hat{x}_{min}$	96 765 (986) [95 153, 99 525]	1 346 045 (752 639) [428 854, 4 691 217]	99 386 (388 340) [92 039, 1 934 317]
$\hat{n}_P$	214 (32.67) [131, 257]	64 (25.18) [22.5, 163.5]	389 (90.42) [54, 397]
$\hat{r}$	395 (4.10) [390, 406]	64 (25.18) [22.5, 163.5]	389 (90.42) [54, 397]

Standard errors are in parentheses, 95% confidence intervals are in brackets

this kind there is a large body of literature trying to identify whether the lognormal distribution is the true model for all observations or there is a Pareto tail. In particular, the size distribution of the US metropolitan areas has been investigated by Gabaix and Ibragimov (2011). Hence, there are valid theoretical reasons supporting (1) as a model for the data. On the empirical side, we have carried out an Anderson-Darling test of the null hypothesis that the estimated mixture (see the values of the parameters in Table 2 below) and the observed data are not significantly different. The test yields a value $AD = 0.484$, with p value equal to 0.763. The data are displayed in Fig. 2 along with the fitted lognormal–Pareto mixture and an estimated kernel density.

The results obtained with the MIX, UMPU and CSN approaches are shown in Table 2. For each parameter, the three rows display the point estimate, the standard errors (in parentheses) and the 95% confidence intervals (in brackets); standard errors and confidence intervals have been computed via non-parametric bootstrap with 500 replications. Recall that the estimated number of Pareto observations $\hat{n}_P$ is given by $n\hat{\pi}_2$ for MIX and by $\hat{x}_{min}$ for UMPU and CSN.

The estimates related to the Pareto distribution, i.e. $\hat{\alpha}$, $\hat{x}_{min}$ and $\hat{n}_P$, are quite different in the three cases. $\hat{\alpha}^{MIX}$ is similar to $\hat{\alpha}^{UMPU}$, but the confidence interval is

Fig. 2 Histogram of the 2019 population estimate of the 415 US metropolitan areas, fitted lognormal-Pareto mixture (MIX) and estimated kernel density (KD). The numerical values of the mixture parameters are displayed in Table 2

shorter; $\hat{x}_{min}^{MIX}$ is comparable to $\hat{x}_{min}^{CS}$, but it is worth noting that $\hat{x}_{min}^{MIX}$ is much more stable than both $\hat{x}_{min}^{UMPU}$ and $\hat{x}_{min}^{CS}$. As a consequence, $\hat{r}^{MIX}$ is very close to $\hat{r}^{CS}$, but the latter has a larger variance. Moreover, $\hat{n}_p^{CS}$ is much larger, which is implausible since the two distributions considerably overlap: note that $\hat{r}^{MIX} - \hat{n}_p^{MIX} = 181$, i.e., according to the MIX approach, there are 181 lognormal observations above the estimated threshold.

The test of $H_0 : \pi = 1$ with $B = 500$ bootstrap replications² clearly rejects the null hypothesis: the test statistics is equal to 102.11, with p value equal to 0.

For comparison purposes, we have also employed four models based on classical size distributions such as the lognormal, the gamma, the Weibull and the Generalized Beta distribution of the second kind (GB2). For estimation, we have used the `fitdistr` command of the R `MASS` package for the first two distributions, the `dglm` command of the R `dglm` package for the gamma and the `mlfit.gb2` command of the R `GB2` package for the GB2. The QQ-plots of the observed data versus data simulated from the estimated distributions are shown in Figs. B.1–B.4 in the online supplementary appendix. None of the distributions yields a good fit: the first three models underestimate the tail of the distribution and the last one overestimates it.

Moreover, we have fitted a two-population lognormal mixture X_{MLN} and a two-population Pareto mixture X_{MP} . The densities are respectively given by

$$f_{X_{MLN}}(x) = \pi f_{LN1}(x; \mu_1, \sigma_1^2) + (1 - \pi) f_{LN2}(x; \mu_2, \sigma_2^2),$$

² The test has been carried out with $B \in \{300, 400, 500, 600, 700\}$: according to basic descriptive measures, for $B \geq 400$ the null distribution shows very little changes. In applications, it is advisable to choose a sufficiently large B by performing such an analysis on a case-by-case basis.

$$f_{X_{MP}} = \pi f_{P1}(x; x_{min}, \alpha_1) + (1 - \pi) f_{P2}(x; x_{min}, \alpha_2). \quad (13)$$

Estimation of the lognormal mixture is straightforward, since the logarithm of a sample from (13) follows a normal mixture with the same parameters, which can be estimated by a standard application of the EM algorithm.

The Pareto mixture can also be fitted via the EM algorithm, provided x_{min} is known. To save space, we omit the details here. However, our estimation experiments with the data at hand, with x_{min} set to different values, have always returned estimates of α_1 and α_2 identical up to the 5-th decimal digit, i.e. essentially a single Pareto distribution. Hence, empirical evidence suggests that a Pareto mixture is not an appropriate distribution for these data.

Figure 3 shows the QQ-plots of the observed data versus data simulated from the estimated lognormal–Pareto mixture (top) and the lognormal mixture (bottom). According to the plots, the former yields a better fit.

4.2 Firm size

The dataset studied in this section contains the number of employees in year 2016 in all the firms of the Trento district in Northern Italy.³ An histogram representation along with the fitted lognormal–Pareto mixture and an estimated kernel density is shown in Fig. 4. The total number of observations is $n = 183$. It is worth noting that the histogram suggests a discontinuity between 4500 and 5000, which is correctly estimated the MIX method ($\hat{x}_{min} = 4717$, see Table 3).

By means of the same analysis of the previous section we obtain the estimated parameters displayed in Table 3.

The fit of the lognormal–Pareto mixture is very good, since the Anderson-Darling test that the estimated mixture and the observed data are not significantly different is equal to 0.164, with p -value equal to 0.997. On the other hand, standard deviations and confidence intervals in Table 3 falsely suggest that MIX is the worst approach. However, the correct interpretation is that it identifies a shorter Pareto tail and a smaller number of Pareto observations with respect to UMPU and CSN. Consequently, the standard errors related to the estimators of the Pareto parameters are larger. Moreover, the standard error of $\hat{x}_{min}^{MIX}$ is large since $\hat{x}_{min}^{MIX}$ is much larger than $\hat{x}_{min}^{UMPU}$ and $\hat{x}_{min}^{CS}$, and the largest observations are scattered over a wide range.

The llr test of $H_0 : \pi = 1$ is equal to 13.98, with corresponding p value, obtained with $B = 500$ bootstrap replications, equal to 0.176. Hence, we cannot reject H_0 at the usual significance levels.

Since, according to the test, a Pareto component should not be included in the model, we have found another explanation for the instability of the Pareto estimators in Table 3.

³ <https://dati.trentino.it/dataset/addetti-in-imprese-attive>.

Fig. 3 QQ-plots of the observed data versus data simulated from the lognormal–Pareto mixture (top) and the lognormal mixture (bottom)

5 Conclusions

Unsupervised maximum likelihood estimation of a lognormal–Pareto mixture provides the investigator with a reliable estimate of the tail of the distribution. Simulation evidence and real-data analyses suggest that this method outperforms two commonly used approaches to the problem.

Fig. 4 Number of employees in all the firms in the Trento district in 2016, fitted lognormal–Pareto mixture (MIX) and estimated kernel density (KD)

Table 3 Estimated parameter values and number of Pareto observations using the MIX, UMPU and CSN approach

	MIX	UMPU 1%	CSN
$\hat{\pi}$	0.9200 (0.0787) [0.7273, 0.9845]	— — —	— — —
$\hat{\mu}$	5.6704 (0.1080) [5.4013, 5.9969]	— — —	— — —
$\hat{\sigma}^2$	1.8449 (0.1741) [1.5465 2.5003]	— — —	— — —
$\hat{\alpha}$	0.9478 (0.3668) [0.3350 2.3985]	0.7229 (0.2896) [0.5791, 1.6688]	0.6500 (0.1179) [0.5400, 1.0050]
$\hat{x}_{min}$	4717 (9300.22) [19, 47 703]	778 (1810.3) [136, 6076]	250 (216.74) [130, 1104]
$\hat{n}_P$	15 (32.67) [2, 50]	53 (40.83) [14, 142]	105 (22.46) [44, 139]
$\hat{r}$	17 (59.2092) [2, 178]	53 (40.83) [14, 142]	105 (22.46) [44, 139]

Standard errors are in parentheses, 95% confidence intervals are in brackets

In most setups, the lognormal–Pareto mixture has a substantial theoretical support, but it is important to check that empirical evidence also supports this model as an appropriate distribution for the data. If goodness-of-fit tests reveal that the true data-generating process is different from (1), the model for all data is likely to be a mixture with the same structure as (1), with the lognormal replaced by some other density. In this case, it is not difficult to modify the EM algorithm to estimate the parameters of such a distribution, as long as one knows how to compute complete-data MLEs. Unless the available sample is truncated, a mixture-based model, possibly with a different distribution for the body, is likely to be a more sensible approach with respect to the commonly used techniques.

Various topics remain open to future research. On the theoretical side, the model can be extended by considering a K -population mixture with $K > 2$. On the computational side, the llr test of the hypothesis $H_0 : \pi = 1$ requires Monte Carlo simulation to find the null distribution. Since this task gets computationally heavier as the sample size increases, easing this burden is an issue that requires further investigation.

Two more issues are worth noting. First, it would be interesting to compare the performance of our likelihood-based method and of the Bayesian approach developed by Benzidia and Lubrano (2020). Second, our testing procedure in Sect. 2.4 is based on parametric bootstrap. As pointed out by Hall and Horowitz (2013), non-parametric bootstrap is more robust with respect to model misspecification, albeit at the price of increased computational complexity. Further work on the use of non-parametric bootstrap is needed to shed light on the strengths and weaknesses of this choice.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11634-022-00497-4>.

Acknowledgements An earlier version of this paper has been presented at the workshop “Models and Learning for Clustering and Classification” (University of Catania, September 2020). I would like to thank the participants for their valuable suggestions. I would also like to thank the Associate Editor and three anonymous reviewers for valuable comments that have considerably improved a preliminary version of the paper.

Funding Open access funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Abdul Majid M, Ibrahim K (2021) On Bayesian approach to composite Pareto models. *PLoS ONE* 16:e0257762

- Abu Bakar S, Hamzah N, Maghsoudi M, Nadarajah S (2015) Modeling loss data using composite models. *Insur Math Econom* 61:146–154
- Axtell RL (2001) Zipf distribution of U.S. firm sizes. *Science* 293(5536):1818–1820
- Bee M, Riccaboni M, Schiavo S (2017) Where Gibrat meets Zipf: scale and scope of French firms. *Physica A* 481:265–275
- Benzidia M, Lubrano M (2020) A Bayesian look at American academic wages: from wage dispersion to wage compression. *J Econ Inequal* 18:213–238
- Berry BJ, Okulicz-Kozaryn A (2012) The city size distribution debate: resolution for US urban regions and megalopolitan areas. *Cities* 29(Supplement 1):S17–S23
- Biernacki C, Celeux G, Govaert G (2003) Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Comput Stat Data Anal* 41(3):561–575
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–673
- D'Acci L (2019) *The mathematics of urban morphology*. Birkhäuser, Boston
- Del Castillo J, Puig P (1999) The best test of exponentiality against singly truncated normal alternatives. *J Am Stat Assoc* 94:529–532
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39(1):1–38
- Di Giovanni J, Levchenko AA, Rancière R (2011) Power laws in firm size and openness to trade: measurement and implications. *J Int Econ* 85(1):42–52
- Eeckhout J (2004) Gibrat's law for (all) cities. *Am Econ Rev* 94(5):1429–51
- Eeckhout J (2009) Gibrat's law for (all) cities: reply. *Am Econ Rev* 99(4):1676–83
- Fazio G, Modica M (2015) Pareto or log-normal? best fit and truncation in the distribution of all cities. *J Reg Sci* 55(5):736–756
- Flury B (1997) *A first course in multivariate statistics*. Springer, Berlin
- Frigessi A, Haug O, Rue H (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* 3(5):219–235
- Gabaix X (2009) Power laws in economics and finance. *Annu Rev Econ* 1:255–93
- Gabaix X, Ibragimov R (2011) Rank-1/2: a simple way to improve the OLS estimation of tail exponents. *J Bus Econ Stat* 29(1):24–39
- Gomes M, Guillou A (2015) Extreme value theory and statistics of univariate extremes: a review. *Int Stat Rev* 83(2):263–292
- González-Val R, Ramos A, Sanz-Gracia F, Vera-Cabello M (2015) Size distributions for all cities: Which one is best? *Pap Reg Sci* 94(1):177–196
- Hall P, Horowitz J (2013) A simple bootstrap method for constructing nonparametric confidence bands for functions. *Ann Stat* 41:1892–1921
- Hsu W-T (2012) Central place theory and city size distribution. *Econ J* 122(563):903–932
- Ioannides Y, Skouras S (2013) US city size distribution: robustly Pareto, but only in the tail. *J Urban Econ* 73(1):18–29
- Kleiber C, Kotz S (2003) *Statistical size distributions in economics and actuarial sciences*. Wiley, New York
- Klugman SA, Panjer HH, Willmot GE (2004) *Loss models: from data to decisions*, 2nd edn. Wiley, New York
- Kondo I, Lewis L, Stella A (2021) Heavy tailed, but not Zipf: firm and establishment size in the U.S. U.S. Census working paper number CES-21-15
- Levy M (2009) Gibrat's law for (all) cities: comment. *Am Econ Rev* 99(4):1672–75
- Malevergne Y, Pisarenko V, Sornette D (2009) Gibrat's law for cities: uniformly most powerful unbiased test of the Pareto against the lognormal. *Swiss Finance Institute Research Paper Series*, pp 09–40
- McLachlan G, Krishnan T (2008) *The EM algorithm and extensions*, 2nd edn. Wiley, New York
- Reed W (2001) The Pareto, Zipf and other power laws. *Econ Lett* 74(1):15–19
- Rozenfeld H, Rybski D, Gabaix X, Makse H (2011) The area and population of cities: new insights from a different perspective on cities. *Am Econ Rev* 101(5):2205–25
- Scollnik DPM (2007) On composite lognormal–Pareto models. *Scand Actuar J* 1:20–33
- Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82(398):605–610
- Tang A (2015) Does Gibrat's law hold for Swedish energy firms? *Empir Econ* 49:659–674

- Titterton D, Smith A, Makov U (1985) Statistical analysis of finite mixture distributions. Wiley, New York
- Wu CFJ (1983) On the convergence properties of the EM algorithm. *Ann Stat* 11(1):95–103

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.