



Goodness-of-fit test for the one-sided Lévy distribution

Aditi Kumari and Deepesh Bhati

Department of Statistics, Central University of Rajasthan, Ajmer, India

ABSTRACT

The main aim of this work is to develop a new goodness-of-fit test for the one-sided Lévy distribution. The proposed test is based on the scale-ratio approach in which two estimators of the scale parameter of one-sided Lévy distribution are confronted. The asymptotic distribution of the test statistic is obtained under null hypotheses. The performance of the test is demonstrated using simulated observations from various known distributions. Finally, two real-world datasets are analyzed.

ARTICLE HISTORY

Received 8 September 2021

Accepted 14 August 2023

KEYWORDS

Asymptotic normality; one-sided Lévy distribution; gamma distribution; Monte Carlo simulation

2010 MATHEMATICS

SUBJECT

CLASSIFICATIONS

62E20; 62F03

1. Introduction

In 1920, French mathematician Paul Lévy with Aleksander Khinchine developed the theory of stable distribution [10]. Subsequently, it has been applied in different disciplines, like economics, physics, hydrology, biology, and signals processing to capture asymmetry, tail behavior, and high kurtosis in datasets [1]. The one-sided Lévy distribution is a special case of stable distribution with positive support and has a heavier tail than any exponential tail distributions. Dumé [5] showed that the Lévy distribution could well describe the sequence of polarity reversals. Rogers [9] used the one-sided Lévy distribution to model the length of paths that are followed by photons after reflection from a turbid media. Despite its use in statistical modeling of physical phenomenon, the construction of goodness-of-fit test for the one-sided Lévy distribution has not attracted much attention from the researchers. Recently, Bhati and Kattumannil [3] proposed Jackknife empirical likelihood (JEL) ratio and adjusted Jackknife empirical likelihood (AJEL) ratio tests for testing the one-sided Lévy distribution. Motivated with the fact that very few tests are available on the goodness-of-fit test for the one-sided Lévy distribution in the statistical literature, we propose a new test for the one-sided Lévy distribution in this article.

This manuscript is structured as follows: In Section 2, we propose a new estimator of the scale parameter of the one-sided Lévy distribution. The proposed test and its asymptotic distribution under null and alternative hypothesis is introduced in Section 3. Simulated critical values at different significant levels and empirical power of the proposed test and their comparison with existing tests are discussed using Monte Carlo Simulation in

CONTACT Deepesh Bhati  deepesh.bhati@curaj.ac.in  Department of Statistics, Central University of Rajasthan, Ajmer, India

Section 4. Finally, we apply our test to two real-world datasets in Section 5.

2. A new estimator of σ

A positive random variable (rv) is said to follow a one-sided Lévy distribution with scale parameter $\sigma > 0$, denoted as $Lv(\sigma)$, if its probability density function (pdf) is of the form

$$f(x; \sigma) = \sqrt{\frac{\sigma}{2\pi}} x^{-3/2} e^{-\frac{\sigma}{2x}}, \quad x > 0, \tag{1}$$

and the cumulative distribution function (cdf) is given by

$$F(x) = 2 \left(1 - \Phi \left(\left(\frac{\sigma}{x} \right)^{1/2} \right) \right), \quad x > 0, \tag{2}$$

where $\Phi(\cdot)$ is the cdf of a standard normal random variable. The one-sided Lévy distribution can also be obtained as a particular case of the inverse gamma distribution with the shape parameter $1/2$ and the scale parameter $\sigma/2$, which implies that the one-sided Lévy distribution possesses finite inverse moments. Lemma 2.1 will be used later to construct our proposed test statistic.

Lemma 2.1: *Let $X \sim Lv(\sigma)$ then for $l = 0, 1, 2, \dots$, and $k = 1, 2, \dots$, the rv $Y = 1/X$ follows the relation*

$$\sigma \mathbb{E} \left(Y^{l+1} (\log Y)^k \right) = (2l + 1) \mathbb{E} \left(Y^l (\log Y)^k \right) + 2k \mathbb{E} \left(Y^l (\log Y)^{k-1} \right). \tag{3}$$

Proof: Let $X \sim Lv(\sigma)$, then the transformed rv $Y = 1/X \sim \text{Gamma} \left(\frac{1}{2}, \frac{\sigma}{2} \right)$ with density $f_Y(y) = \sqrt{\frac{\sigma}{2\pi}} y^{-1/2} e^{-\frac{\sigma y}{2}}, y > 0$ and hence

$$\mathbb{E} \left(Y^l (\log Y)^k \right) = \sqrt{\frac{\sigma}{2\pi}} \int_0^\infty y^{l-1/2} (\log y)^k e^{-\sigma y/2} dy,$$

noting the fact that, $\frac{d}{dy} (y^{l+1/2} (\log y)^k) = (l + 1/2) y^{l-1/2} (\log y)^k + k y^{l-1/2} (\log y)^{k-1}$, the above integral can be re-written as

$$\begin{aligned} \mathbb{E} \left(Y^l (\log Y)^k \right) &= \sqrt{\frac{\sigma}{2\pi}} \frac{2}{2l + 1} \int_0^\infty \frac{d}{dy} \left(y^{l+1/2} (\log y)^k \right) e^{-\sigma y/2} dy \\ &\quad - \sqrt{\frac{\sigma}{2\pi}} \frac{2k}{2l + 1} \int_0^\infty y^{l-1/2} (\log y)^{k-1} e^{-\sigma y/2} dy \end{aligned}$$

solving the first integral by parts, we obtain

$$\mathbb{E} \left(Y^l (\log Y)^k \right) = \frac{\sigma}{2l + 1} \mathbb{E} \left(Y^{l+1} (\log Y)^k \right) - \frac{2k}{2l + 1} \mathbb{E} \left(Y^l (\log Y)^{k-1} \right).$$

Hence the lemma follows. ■

Let $\mathbb{X} := \{X : X = \sigma \tilde{X}, \sigma > 0\}$ be a generic collection of non-negative rvs with one (scale) parameter σ , where \tilde{X} is a unit scale rv. Then the covariance between X and $\log X$ allows easy scale extraction of scale parameter σ , that means,

$$\text{cov}(X, \log X) = \text{cov}(\sigma \tilde{X}, \log(\sigma \tilde{X})) = \sigma \text{cov}(\tilde{X}, \log \tilde{X})$$

which gives

$$\sigma = \frac{\text{cov}(X, \log X)}{\text{cov}(\tilde{X}, \log \tilde{X})},$$

where $\text{cov}(\tilde{X}, \log \tilde{X})$ is some known constant depending on the distribution of \tilde{X} . We utilize the above idea to estimate the scale parameter of one-sided Lévy distribution. Let rv $X \sim Lv(\sigma)$, then rv \tilde{X} follows $Lv(1)$. Define another collection of rvs using the inverse transform of X , say $\mathbb{Y} := \{Y : Y = 1/X, X \in \mathbb{X}\}$ and $\tilde{Y} = 1/\tilde{X}$. Hence rv Y and \tilde{Y} follows $Gamma(\frac{1}{2}, \frac{\sigma}{2})$ and $Gamma(\frac{1}{2}, \frac{1}{2})$ respectively. Hereinafter, we consider rv $Y = 1/X$ and $Z = \log Y$ with mean $\mu_Y = \frac{1}{\sigma}$ and variance $\sigma_Y^2 = \frac{2}{\sigma^2}$ and with simple computation the mean and variance of rv Z are $\mu_Z = -\gamma - \log(2\sigma)$, and $\sigma_Z^2 = \frac{\pi^2}{2}$ respectively. Then $\text{cov}(Y, \log Y) = \text{cov}(\frac{\tilde{Y}}{\sigma}, \log \frac{\tilde{Y}}{\sigma}) = \frac{1}{\sigma} \text{Cov}(\tilde{Y}, \log \tilde{Y})$ gives $\sigma = \frac{\text{Cov}(\tilde{Y}, \log \tilde{Y})}{\text{Cov}(Y, \log Y)}$. Using $\mathbb{E}(\tilde{Y} \log \tilde{Y})$, $\mathbb{E}(\tilde{Y})$ and $\mathbb{E}(\log \tilde{Y})$ respectively in $\text{Cov}(\tilde{Y}, \log \tilde{Y})$ we get $\text{Cov}(\tilde{Y}, \log \tilde{Y}) = 2 - \gamma - \log 2 + \gamma + \log 2 = 2$, where γ is Euler's constant, we get

$$\sigma = \frac{2}{\text{Cov}(Y, \log Y)} \tag{4}$$

Remark 2.1: The above approach can be applied to any generic family of distributions that have finite moments/log-moments for estimating the scale parameter.

For a random sample X_1, X_2, \dots, X_n of size n from $Lv(\sigma)$, consider the transformed observations $Y_i = 1/X_i$ and $Z_i = \log Y_i$, $i = 1, 2, \dots, n$ with $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ and $\bar{Z}_n = \sum_{i=1}^n Z_i/n$, then the proposed new estimator ($\tilde{\sigma}_n$) of σ based on sample covariance $\widehat{Cov}_n(Y, Z)$ is given as

$$\tilde{\sigma}_n = \frac{2}{\widehat{Cov}_n(Y, Z)} = \frac{2(n-1)}{\sum_{i=1}^n (Y_i - \bar{Y}_n)(Z_i - \bar{Z}_n)}. \tag{5}$$

The sample covariance $\widehat{Cov}_n(Y, Z)$ can also be written as $\widehat{Cov}_n(Y, Z) = \frac{1}{(n-1)} \sum_{i=1}^n (Y_i - \mu_Y)(Z_i - \mu_Z) - (\bar{Y}_n - \mu_Y)(\bar{Z}_n - \mu_Z) = (n-1)^{-1} \sum_{i=1}^n (Y_i - \mu_Y)(Z_i - \mu_Z) + O_p(n^{-1})$ with $\mathbb{E}(\widehat{Cov}_n(Y, Z)) = \frac{2}{\sigma} + O_p(n^{-1})$ and $\mathbb{V}\text{ar}(\widehat{Cov}_n(Y, Z)) = \frac{1}{n} \mathbb{V}\text{ar}((Y_i - \mu_Y)(Z_i - \mu_Z))$. This shows that $\frac{1}{\tilde{\sigma}_n}$ is a consistent estimator of $\frac{1}{\sigma}$. Hence by invariance property of consistency, which means that, if T_n is any consistent estimator of some parameter θ and $\xi(\cdot)$ is a continuous function then $\xi(T_n)$ is a consistent estimator of $\xi(\theta)$. Therefore using this result with ξ defined as $\xi(t) = 1/t$, one can say that $\tilde{\sigma}_n$ is an consistent estimator of σ_n .

For benchmark purposes, we use the maximum likelihood estimator of scale parameter σ , which is the solution to the likelihood equation $\frac{n}{2\sigma} - \sum_{i=1}^n \frac{Y_i}{2} = 0$ and is given by

$$\hat{\sigma}_n = (\bar{Y}_n)^{-1}, \tag{6}$$

and this estimator will be used later for the construction of our ratio test statistic.

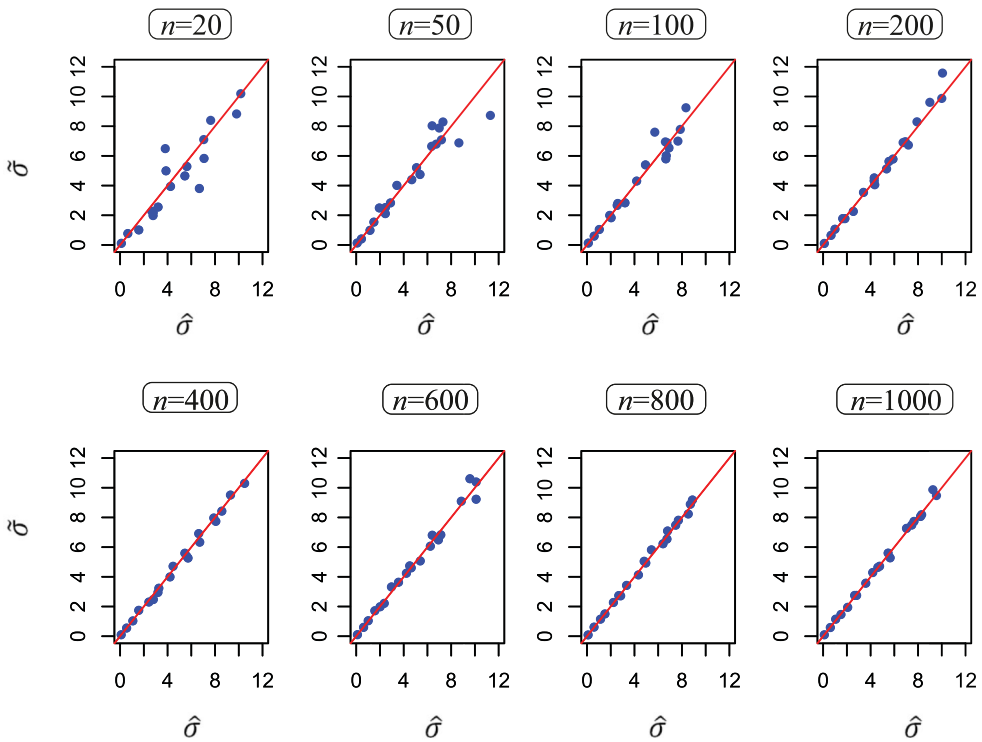


Figure 1. Scatter plot of estimator of σ obtained, by using maximum likelihood estimator (MLE) and Method of covariance (MCoV) for different sample sizes.

In Figure 1, we give the results of a comparative study that we conducted to compare the estimated value of σ obtained using the method of maximum likelihood estimate (MLE) and method of covariance (MCoV) for different sample sizes. We generate a sample of size $n = 20, 50, 100, 200, 400, 600, 800$ and 1000 . For each n , samples from $L\nu(\sigma)$, σ varies from 0 to 10, are generated and the estimated value of σ from ML and MCoV method are plotted. In the figure, we observe that the estimated value of σ by both the methods falls closer to the diagonal line as the sample size increases, which is an important aspect because the test introduced later depends on these being similar. Further to complement this, we compare the performance of the new estimator $\tilde{\sigma}_n$ with $\hat{\sigma}$ by generating 100,000 samples each of size 10, 20, 50, 100 and 200 from $L\nu(5)$. For each sample with size n , we compute $\hat{\sigma}_n$ and $\tilde{\sigma}_n$ and depict the box plot of it. Figure 2 gives the box plot for both the estimators for different sample sizes. From this figure, we observe that the inter-quartile range of the box plots decreases as sample size increases. And the median line of these box plots are very close to the actual value of σ (i.e. $\sigma = 5$). In addition to this, MCoV estimate shows more dispersion as compared to ML estimate for all sample sizes, nevertheless, this spread decreases as sample sizes increase. The MCoV estimate is more right-skewed as compared to MLE estimate for small samples. As the exact expressions of the variance and bias of the MCoV estimator is not in a closed form, we use simulation to get a visual look on the behavior $n \rightarrow \infty$. The Bias($\hat{\sigma}_n$), Bias($\tilde{\sigma}_n$), MSE($\hat{\sigma}_n$) and MSE($\tilde{\sigma}_n$) for different sample sizes varies from $n = 10, 50, 100, 150, 200, \dots, 1500$ is obtained from 100,000

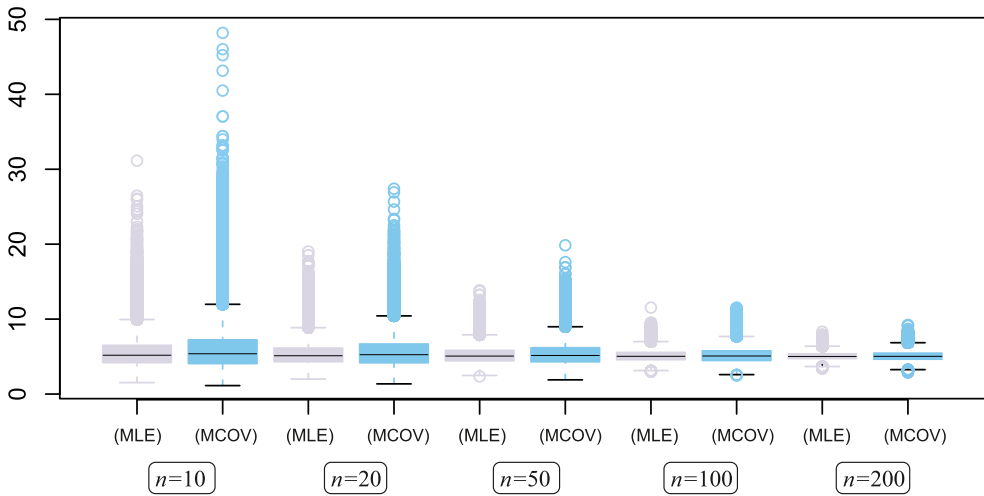


Figure 2. Box plots of estimator of σ obtained, by using maximum likelihood estimator (MLE) and Method of covariance (MCOV), from 10,000 samples of size $n = 10, 20, 50, 100,$ and 200 from L_v with $\sigma = 5$.

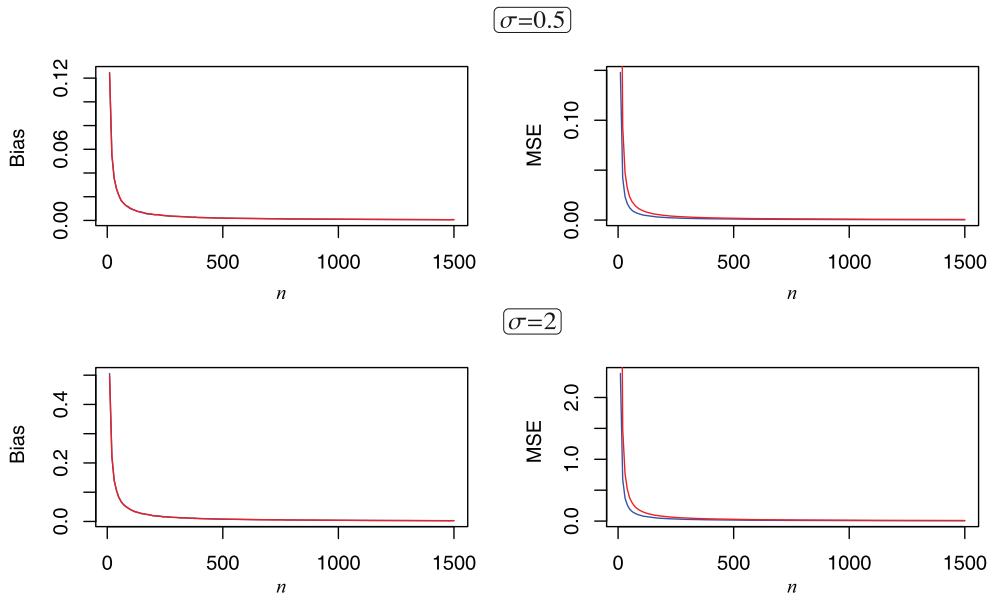


Figure 3. Bias and Means square error (MSE) of Maximum Likelihood Estimator (MLE) (in Blue) and Method of covariance (MCOV) estimator (in Red) for $\sigma = 0.5$ and $\sigma = 2$.

samples each of size n and is drawn in Figure 3. Here, the Bias and MSE is decreasing as the sample size increases, which suggests that for both the ML and MCOV estimates the MLE and MSE approaches 0. It can be further observed that, irrespective of sample size, the ML estimator has small bias and MSE as compared to MCOV estimator. Thus, rather than parameter fitting, our main purpose is to introduce a new goodness-of-fit testing using this new estimator.

3. Scale-ratio test

Let X be a rv with continuous cdf F with support \mathbb{R}^+ and let \mathcal{L} denote the one-sided Lévy family of distributions having density (1). We propose a goodness-of-fit test for the composite null hypothesis $\mathcal{H}_0 : F \in \mathcal{L}$ versus the alternative hypothesis $\mathcal{H}_1 : F \notin \mathcal{L}$, based on a random sample X_1, \dots, X_n of size n from F . Similar to the well-known Shapiro–Wilk test (see [11]) for testing the normality, we will construct our test statistic by taking the ratio of the estimators $\tilde{\sigma}_n$ and $\hat{\sigma}_n$, that is $V_n = \tilde{\sigma}_n/\hat{\sigma}_n = \frac{2\bar{Y}_n}{\widehat{Cov}_n(Y,Z)}$ for testing \mathcal{H}_0 . Note that, V_n is pivotal quantity with respect to σ . Hence, in order to test the \mathcal{H}_0 hypothesis, the value of the ratio V_n is expected to be close to one. In following section, we discuss the asymptotic distribution of V_n and propose the test statistic.

3.1. Asymptotic distribution of V_n

Theorem 3.1: Let $X_i, i = 1, 2, \dots, n$ be iid positive rv's having finite second inverse moment and finite second log moment, set $Y_i = \frac{1}{X_i}$ and $Z_i = \log Y_i$. Then, for $\mathbf{T}_n = (\bar{Y}_n, \widehat{Cov}_n(Y, Z))^T$, we have

$$\sqrt{n}(\mathbf{T}_n - \Psi) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

where \xrightarrow{d} denotes convergence in distribution, $\mathbf{0} = (0, 0)^T$, $\Psi = (\mu_Y, \sigma_{YZ})^T$ and $\Sigma = \begin{pmatrix} cc\text{Var}(U_1) & \text{Cov}(U_1, U_2) \\ \text{Cov}(U_2, U_1) & \text{Var}(U_2) \end{pmatrix}$, with $U_1 = (Y_1 - \mu_Y)$ and $U_2 = (Y_1 - \mu_Y)(Z_1 - \mu_Z) - \sigma_{YZ}$. Moreover, for $V_n = \frac{2\bar{Y}_n}{\widehat{Cov}_n(Y,Z)}$, we have

$$\sqrt{n} \left(V_n - \frac{2\mu_Y}{\sigma_{YZ}} \right) \xrightarrow{d} N(0, \eta^2),$$

where $\eta^2 = [h'(\Psi)]^T \Sigma h'(\Psi)$ with $h(x_1, x_2) = \frac{2x_1}{x_2}$ and hence $h'(\Psi) = \left(\frac{2}{\sigma_{YZ}}, -\frac{2\mu_Y}{\sigma_{YZ}^2} \right)^T$.

Proof: Notice that, $\widehat{Cov}_n(Y, Z) = \frac{1}{n} \sum_{i=1}^{(n-1)} (Y_i - \mu_Y)(Z_i - \mu_Z) - (\bar{Y}_n - \mu_Y)(\bar{Z}_n - \mu_Z) = (n-1)^{-1} \sum_{i=1}^n (Y_i - \mu_Y)(Z_i - \mu_Z) + O_p(n^{-1})$. Also, by the Central Limit theorem, we have $\bar{Y}_n - \mu_Y = O_p(n^{-1/2})$. Therefore, for any values t_1 and t_2 , we consider

$$\begin{aligned} I_n &= \sqrt{n} \{ t_1(\bar{Y}_n - \mu_Y) + t_2(\widehat{Cov}_n(Y, Z) - \sigma_{YZ}) \} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [t_1(Y_i - \mu_Y) + t_2((Y_i - \mu_Y)(Z_i - \mu_Z) - \sigma_{YZ})] + O_p(n^{-1}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n R_i + O_p(n^{-1}), \end{aligned}$$

where $R_i = t_1(Y_i - \mu_Y) + t_2((Y_i - \mu_Y)(Z_i - \mu_Z) - \sigma_{YZ}), i = 1, 2, \dots, n$ are iid r.v. with $\mathbb{E}(R_i) = 0$ and $\mathbb{V}\text{ar}(R_i) = \mathbf{t}^T \Sigma \mathbf{t}$ with $\mathbf{t}^T = (t_1, t_2)$. By the Central Limit Theorem along with Slutsky's theorem, we conclude that I_n is asymptotically normally distributed. Since

\mathbf{t}^\top is arbitrary, by Cramér Wold theorem [4, p. 9], we get

$$\sqrt{n} \left\{ (\bar{Y}_n, \widehat{\text{Cov}}_n(Y, Z))^\top - (\mu_Y, \sigma_{YZ})^\top \right\} \xrightarrow{d} N(0, \Sigma).$$

Define $h : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ such that $h(x_1, x_2) = 2x_1/x_2$. Notice that $V_n = h(\bar{Y}_n, \widehat{\text{Cov}}_n(Y, Z))$, we apply the multivariate delta theorem to obtain the desired result. That is, let $\Psi = (\mu_Y, \sigma_{YZ})$ and notice that $h'_1(x_1, x_2) = 2/x_2$ and $h'_2(x_1, x_2) = -2x_1/x_2^2$ where $h'_i(x_1, x_2) = \frac{\partial}{\partial x_i} h(x_1, x_2)$. Hence, $h'(\Psi) = \left(\frac{2}{\sigma_{YZ}}, -\frac{2\mu_Y}{\sigma_{YZ}^2} \right)^\top$. ■

Theorem 3.1 gives rise to the following corollary, which tells us what happens under \mathcal{H}_0 .

Corollary 3.1: Under \mathcal{H}_0 , if $X \sim Lv(\sigma)$, then $\sqrt{n}(V_n - 1) \xrightarrow{d} N(0, \eta^2)$ where $\eta^2 = \frac{1}{2}(\sigma_Z^2 - 2) = \frac{1}{4}(\pi^2 - 4)$.

Proof: When $Y \sim \text{Gamma}\left(\frac{1}{2}, \frac{\sigma}{2}\right)$, by using Lemma 2.1 and by proceeding iteratively, we have the following equations.

$$\mathbb{E} \left(\left(Y - \frac{1}{\sigma} \right) \log Y \right) = \frac{2}{\sigma}, \tag{7}$$

$$\mathbb{E} \left(\left(Y - \frac{1}{\sigma} \right)^2 \log Y \right) = \frac{2}{\sigma} \mathbb{E}(Y \log Y). \tag{8}$$

By using (3), (7) and (8), it is possible to obtain the covariance expression as

$$\begin{aligned} & \text{cov}(Y - \mu_Y, (Y - \mu_Y)(Z - \mu_Z)) \\ &= \mathbb{E} \left((Y - \mu_Y) \left((Y - \mu_Y)(Z - \mu_Z) - \frac{2}{\sigma} \right) \right), \\ &= \mathbb{E} \left(\left(Y - \frac{1}{\sigma} \right) \left(\left(Y - \frac{1}{\sigma} \right) (\log Y - \mu_Z) - \frac{2}{\sigma} \right) \right), \\ &= \frac{2}{\sigma} \mathbb{E}(Y \log Y) - \frac{2}{\sigma^2} \mu_Z, \\ &= \frac{2}{\sigma^2} \mathbb{E}(\log Y) + \frac{4}{\sigma^2} - \frac{2}{\sigma^2} \mu_Z = \frac{4}{\sigma^2}. \end{aligned}$$

Now, from Lemma 2.1, we obtain the following relation

$$\mathbb{E} \left(\left(Y - \frac{1}{\sigma} \right)^2 (\log Y)^2 \right) = \frac{2}{\sigma^2} \left(\mathbb{E}((\log Y)^2) + 4\mathbb{E}(\log Y) + 4 \right),$$

$$\begin{aligned} \text{and } \text{Var}((Y - \mu_Y)(Z - \mu_Z)) &= \mathbb{E}((Y - \mu_Y)^2(Z - \mu_Z)^2) \\ &\quad - (\mathbb{E}((Y - \mu_Y)(Z - \mu_Z)))^2, \\ &= \mathbb{E} \left(\left(Y - \frac{1}{\sigma} \right)^2 (\log Y - \mu_Z)^2 \right) - \frac{4}{\sigma^2}, \\ &= \frac{2}{\sigma^2} (\sigma_Z^2 + 2). \end{aligned}$$

Therefore, if we denote by $\Sigma_{ij}, i, j = 1, 2$ the elements of the covariance matrix Σ , then $\Sigma_{1,1} = \frac{2}{\sigma^2}, \Sigma_{2,2} = \frac{2}{\sigma^2}(\sigma_Z^2 + 2), \Sigma_{1,2} = \Sigma_{2,1} = \frac{4}{\sigma^2}$. Hence, by Theorem 3.1, for $h'(\Psi) = (\sigma, -\frac{\sigma}{2})^\top, \eta^2 = [h'(\Psi)]^\top \Sigma h'(\Psi) = \frac{1}{2}(\sigma_Z^2 - 2)$. ■

In view of Corollary 3.1, for testing the one-sided Lévy distribution hypothesis we propose a test

$$V_n^* = \sqrt{\frac{4n}{\pi^2 - 4}} (V_n - 1), \tag{9}$$

which is asymptotically distributed as standard normal $N(0, 1)$ under \mathcal{H}_0 . Figure 4 shows the histogram of test statistic (V_n^*) values obtained from 100,000 samples drawn from $Lv(1)$ for different sample sizes (it is to be noted that the test statistics does not depend on σ , hence for brevity, we choose samples from Lévy distribution with $\sigma = 1$). We further superimpose the kernel density plot (in red) and standard normal density (in blue) on the histogram to visualize the test statistic behavior. For small sample size, the distribution of V_n^* is right-skewed, this occurs due to the fact that the estimator $\tilde{\sigma}_n$ have relatively more dispersion than estimator $\hat{\sigma}_n$, see Figure 2. However, the kernel density plot tends to symmetric and overlap with the density of standard normal as the sample size is above 250. There are criteria in the literature which can be used to assess the closeness of the simulated type I error rates to the nominal size, see for example [2]. Therefore, for testing \mathcal{H}_0 based on the sample X_1, X_2, \dots, X_n of size n from a continuous cdf F with support on the positive real numbers, the null hypothesis \mathcal{H}_0 is rejected at the significance level (SL) α , if V_n^* deviates away from 0. That is for a relatively small sample size, reject \mathcal{H}_0 at α if $V_n^* < c_{n,\alpha/2}^V$ or $V_n^* > c_{n,1-\alpha/2}^V$

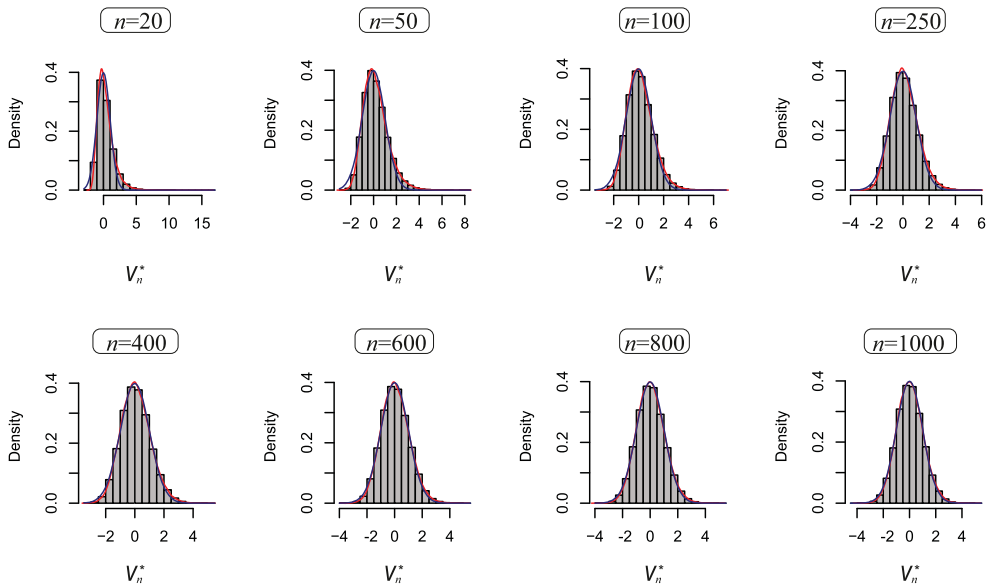


Figure 4. Histogram, kernel density plot(in red color) of the test statistic V_n^* and density of standard normal distribution (in blue color) obtained from 100,000 realizations from $Lv(1)$ of different size n .

where $c_{n,\alpha/2}^v$ and $c_{n,1-\alpha/2}^v$ are such that

$$1 - \alpha = P\left(c_{n,\alpha/2}^v < V_n^* < c_{n,1-\alpha/2}^v \mid \mathcal{H}_0\right).$$

Note that, the statistic V_n^* does not depend on unknown parameter σ and hence is scale invariant. Thus, we use $\sigma = 1$ for computation of empirical size and empirical power. On the other hand, for large sample, we reject the null hypothesis at SL α , if $|V_n^*| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the 100(1 - $\alpha/2$)% quantile of the standard normal distribution. This relies on the fact for large sample size the upper quantiles of the null distribution of $|V_n^*|$ obtained by MC simulation are close to the upper quantiles of the standard half-normal distribution, which is asymptotic null distribution of $|V_n^*|$ (see [12]).

4. Simulation study

To investigate the performance of V_n^* test given in (9), we first evaluate the critical values at SL α by Monte Carlo simulation for different sample sizes. As the proposed test is asymptotically normally distributed, we also examine whether the critical values from the asymptotic distribution can be used in place of simulated critical values. These critical values are then being used to obtain the empirical size and power of the test. Finally, we compare the empirical power of V_n^* test with Jackknife Empirical Likelihood (JEL) and the Adjusted Jackknife Empirical Likelihood (AJEL) tests introduced in Ref. [3]. To begin the investigation, we use the following stepwise procedure to obtain the simulated critical values of V_n^* .

- Step 1 Fix sample size n and $\sigma = 1$.
- Step 2 Generate a sample (x_i 's) of size n from $Lv(\sigma)$.
- Step 3 Compute $y_i = 1/x_i$ and $z_i = \log y_i, i = 1, \dots, n$.
- Step 4 Calculate $V_n = \frac{2(n-1) \sum_{i=1}^n (y_i)}{n \sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z})}$ and then obtain

$$V_n^* = \sqrt{\frac{4n}{\pi^2 - 4}} (V_n - 1)$$

- Step 5 Repeat steps 1–4 B times to get the realization of V_n^* say $V_{n,b}^*, b = 1, \dots, B$.
- Step 6 Then, for sample size $n < 250$, the critical value at SL α , ($c_{n,\alpha}^v$), is the α -th quantile of $V_{n,1}^*, V_{n,2}^*, \dots, V_{n,B}^*$ and for sample of size $n \geq 250$, the critical value at SL α , is the 2α -th upper quantile of $|V_{n,1}^*|, |V_{n,2}^*|, \dots, |V_{n,B}^*|$.

The above procedure is used to obtain the simulated critical values at different SL α for V_n^* tests for small and moderate sample size and for large sample size. The critical value obtained for different sample size $n > 250$ seems close to the critical value of half-normal value. Hence, as a thumb rule, we suggest the readers to use the simulated critical value for $n < 250$ and for $n \geq 250$, critical values, as quantile, of Half-normal distribution can be used. These results are shown in Tables 1 and 2, respectively. Table 1 provides the critical constants of the test corresponding to SL $\alpha = 0.025, 0.05, 0.90$, and 0.95 for samples of size $n = 10, 15, 20, 25, 30, \dots, 250$. These values were obtained as the average of three runs of

Table 1. Simulated critical value of the V_n^* test for different SL α .

n	$c_{n,0.025}^V$	$c_{n,0.05}^V$	$c_{n,0.090}^V$	$c_{n,0.095}^V$	n	$c_{n,0.025}^V$	$c_{n,0.05}^V$	$c_{n,0.090}^V$	$c_{n,0.095}^V$
10	-1.225	-1.085	3.3590	4.584	130	-1.671	-1.437	1.904	2.313
15	-1.315	-1.154	2.8075	3.664	135	-1.685	-1.441	1.908	2.321
20	-1.378	-1.204	2.5551	3.291	140	-1.684	-1.439	1.906	2.334
25	-1.432	-1.243	2.4329	3.072	145	-1.685	-1.439	1.893	2.313
30	-1.459	-1.270	2.3180	2.916	150	-1.695	-1.445	1.894	2.318
35	-1.494	-1.292	2.2659	2.849	155	-1.702	-1.444	1.891	2.303
40	-1.513	-1.306	2.2034	2.758	160	-1.704	-1.453	1.894	2.305
45	-1.526	-1.320	2.1599	2.673	165	-1.708	-1.449	1.883	2.287
50	-1.539	-1.328	2.1346	2.652	170	-1.714	-1.457	1.887	2.298
55	-1.557	-1.336	2.0933	2.601	175	-1.710	-1.457	1.879	2.285
60	-1.572	-1.350	2.0757	2.566	180	-1.710	-1.458	1.876	2.274
65	-1.578	-1.364	2.0541	2.514	185	-1.720	-1.469	1.857	2.258
70	-1.596	-1.371	2.0333	2.516	190	-1.716	-1.458	1.880	2.275
75	-1.600	-1.376	2.0280	2.488	195	-1.722	-1.467	1.859	2.260
80	-1.623	-1.389	2.0158	2.462	200	-1.733	-1.476	1.857	2.250
85	-1.624	-1.395	1.9824	2.441	205	-1.732	-1.473	1.850	2.247
90	-1.633	-1.396	1.9768	2.399	210	-1.728	-1.471	1.860	2.258
95	-1.632	-1.402	1.9668	2.414	215	-1.742	-1.478	1.855	2.238
100	-1.645	-1.410	1.9437	2.400	220	-1.737	-1.476	1.845	2.235
105	-1.654	-1.412	1.9513	2.377	225	-1.747	-1.485	1.859	2.236
110	-1.651	-1.417	1.9357	2.375	230	-1.744	-1.474	1.846	2.243
115	-1.665	-1.420	1.9407	2.365	235	-1.742	-1.484	1.851	2.243
120	-1.672	-1.424	1.9254	2.343	240	-1.743	-1.487	1.839	2.232
125	-1.665	-1.426	1.9284	2.346	245	-1.749	-1.488	1.850	2.230
					250	-1.753	-1.491	1.836	2.227

Table 2. Upper quantiles of $|V_n^*|$ under H_0 for different sample sizes, $n \geq 250$.

n	$c_{n,0.90}^V$	$c_{n,0.95}^V$
250	1.653	1.994
500	1.655	1.985
1000	1.653	1.981
1500	1.645	1.960
2000	1.648	1.964
2500	1.644	1.964
3000	1.650	1.964
3500	1.647	1.961
4000	1.647	1.959
4500	1.641	1.963
5000	1.646	1.959

100,000 MC samples each from the one-sided Lévy distribution with $\sigma = 1$. Table 2 contains the 90% and 95% quantiles of $|V_n^*|$ under \mathcal{H}_0 for $n = 250, 500, 1000, 1500, \dots, 5000$ which were obtained from 100,000 MC samples each. These values are close to the 95% and 97.5% quantiles of the asymptotic standard normal distribution (1.645 and 1.959). The empirical size of the test is then obtained by generating 100, 000 samples each of size (n) varies sample sizes from null hypothesis i.e. $Lv(1)$ and proportion of samples for which the test statistic that falls in the critical region are computed. In Figure 5, we compute the empirical size at 5% SL using the simulated critical values (red) and critical values from asymptotic normal distribution (blue). We can observe in figure that the empirical size obtained from simulated critical value falls close to nominal level i.e. 0.05 for small and

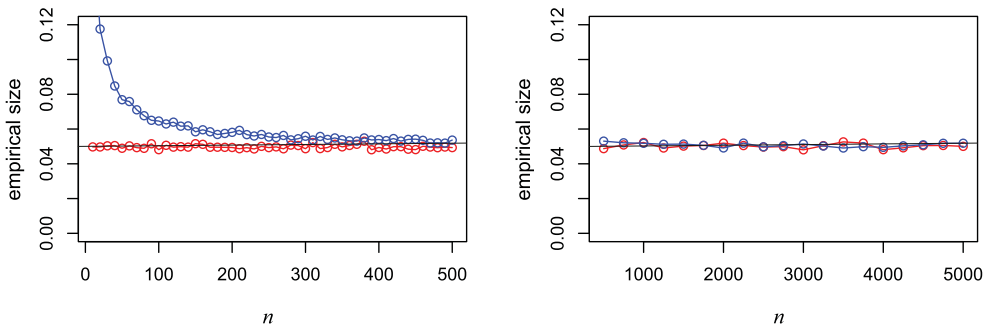


Figure 5. Empirical size obtained using simulated critical value (in red) and critical value of asymptotic normal distribution (in blue) for different sample size n at SL $\alpha = 0.05$.

moderate sample size. Whereas it falls away from the nominal level (0.05) for critical values from asymptotic normal distribution. However, for large sample size ($n \geq 250$), the empirical size obtained by considering simulated critical value or critical values from asymptotic normal distribution fall close to nominal level. Therefore we use the simulated critical value for computation of empirical power.

In order to compute the empirical power of V_n^* , JEL and AJEL tests, we considered several families of distributions from alternative hypothesis, with support $[0, \infty)$, such as Lognormal (0, 1), Chi-square (4), Gamma (2, 3), Fréchet (0.5, 0, 1), Pareto (1.5, 1), Log-Gamma (3, 2), Pareto (0.75, 1), Inverse Gaussian (1, 1.5), Burr (1.5, 0.5, 0.5, 1), Rayleigh (1), Weibull (1.75, 1) and Half-Normal (1). These results are shown in Table 3. For sample size n from each alternative family, we generate 100, 000 independent samples of size n and obtained the value of each test statistic. Then the proportion of samples for which the test statistic falls in critical region are recorded as empirical power. The critical region for JEL and AJEL are obtained as suggested in Ref. [3]. For the family of distributions from alternative class having the right tail heavier than the normal distribution, say Lognormal, chi-square, Gamma, Fréchet, Pareto, Log-Gamma, Burr and Inverse Gaussian, the V_n^* test shows better empirical power than JEL and AJEL, whereas for the distributions having tail equivalent to normal distribution such as Rayleigh, Weibull and Half-Normal the V_n^* test have relatively less power as compare to JEL and AJEL for large sample size. In addition to this, we generate random variables from a stable distribution with skew parameter $\beta = 1$, and stability parameter values were varied from 0.05 to 0.99 using the R-package ‘stabledist’. In Figure 6, the empirical power curve for V_n^* test is presented. The empirical power was computed using samples of sizes $n = 200$ and for 10,000 repetitions. As observed from Figure 6, the empirical powers of the tests decrease as stability parameter approaches 0.5 and increase as the values of deviate from 0.5. It is noteworthy that when stability parameter is 0.5, the empirical power is close to 0.05, consistent with the case of the one-sided Lévy distribution, as anticipated.

5. Application

In this section, we illustrate the use of test proposed in Section 4 with the help of two real datasets discussed in Ref. [3]. The first dataset is from Ref. [8], which represents

Table 3. Proportion of samples falls in the critical region for different family of distributions from alternative class.

<i>n</i>	Lognormal (0, 1)			Chi-square (4)			Gamma (2, 3)		
	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL
10	0.401	0.116	0.119	0.649	0.248	0.255	0.652	0.250	0.257
20	0.704	0.212	0.217	0.818	0.437	0.445	0.814	0.423	0.432
30	0.831	0.297	0.304	0.869	0.638	0.643	0.869	0.636	0.644
50	0.938	0.524	0.482	0.917	0.903	0.905	0.913	0.895	0.899
100	0.988	0.868	0.856	0.949	0.934	0.942	0.959	0.948	0.948
<i>n</i>	Fréchet			Pareto (0.75, 1)			Log-Gamma (3, 2)		
	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL
10	0.165	0.056	0.054	0.483	0.087	0.092	0.827	0.163	0.171
20	0.313	0.066	0.066	0.845	0.109	0.112	0.998	0.270	0.279
30	0.429	0.072	0.072	0.964	0.158	0.161	1.000	0.433	0.439
50	0.636	0.076	0.077	0.999	0.302	0.308	1.000	0.725	0.731
100	0.895	0.094	0.091	1.000	0.631	0.612	1.000	0.973	0.938
<i>n</i>	Pareto (1.5, 1)			Burr (1.5, 0.5, 0.5)			Inverse-Gaussian (1, 1.5)		
	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL
10	0.981	0.347	0.357	0.921	0.086	0.082	0.481	0.112	0.116
20	1.000	0.526	0.539	0.997	0.138	0.136	0.872	0.173	0.179
30	1.000	0.777	0.785	1.000	0.184	0.184	0.977	0.285	0.289
50	1.000	0.977	0.977	1.000	0.292	0.294	0.999	0.514	0.520
100	1.000	1.000	1.000	1.000	0.527	0.506	1.000	0.866	0.853
<i>n</i>	Rayleigh (1)			Weibull (1.75, 1)			Half-Normal (1)		
	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL	V_n^*	JEL	AJEL
10	0.832	0.413	0.423	0.723	0.448	0.371	0.722	0.326	0.332
20	0.913	0.669	0.678	0.854	0.583	0.532	0.844	0.543	0.553
30	0.952	0.853	0.863	0.886	0.761	0.713	0.886	0.772	0.779
50	0.969	0.993	0.991	0.912	0.968	0.958	0.919	0.965	0.967
100	0.973	1.000	1.000	0.976	1.000	1.000	0.978	0.992	1.000

the lifetime of the pressure of $n = 20$ constructed vessels subjected to a certain constant pressure. While fitting the gamma distribution to this data, the ML estimator of shape parameter is close to 0.5 and as the one-sided Lévy distribution is a special case of inverse of Gamma distribution, we consider the inverse of these observations for our purpose. The second dataset comprises a $n = 31$ weighted average of rainfall (in mm) data in January for the whole country starting from 1981 to 2011 released by Meteorological Department, Ministry of Earth Sciences, Government of India. This data is based on more than 2000 rain gauge readings spread over the entire country, and it is available at www.data.gov.in. To get the preliminary idea of model fitting to these data. For parameter estimation, maximum likelihood estimation method is used. The qq -plot between the sample quantiles and fitted one-sided Lévy quantiles are presented in Figure 7 for both datasets. The Kolmogorov–Smirnov test values $K_n = \max\{D^+, D^-\}$, where $D^+ = \max_{j=1, \dots, n} (j/n - \hat{F}(X_{(j)}))$, $D^- = \max_{j=1, \dots, n} (\hat{F}(X_{(j)}) - (j-1)/n)$ and $\hat{F}(X_{(j)})$ is estimated cdf of one sided lévy rv, with bootstrap p -value ($p_{\text{Bootstrap}}^1$) for datasets 1 and 2 are 0.154 (0.498) and 0.461 (< 0.0001), respectively.

For the first dataset, the KS - $p_{\text{Bootstrap}}$ -value is larger than the 5% significance level hence confirming that the one-sided Lévy distribution could be a plausible choice for modeling.

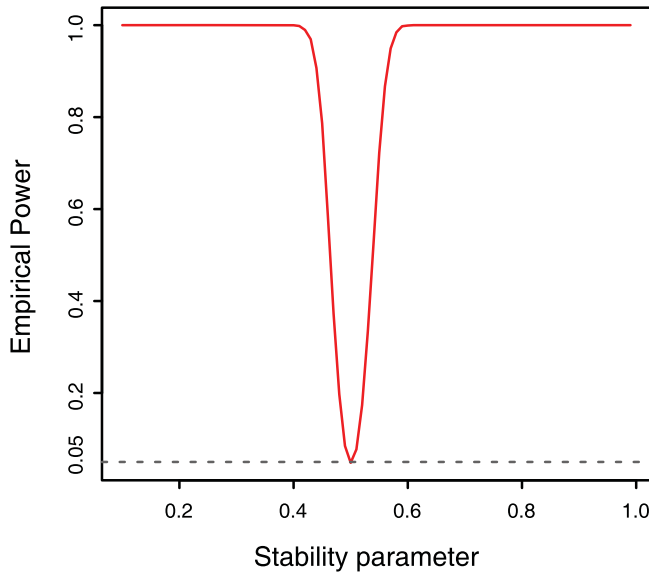


Figure 6. Empirical power curve of the V_n^* test for sample size 200.

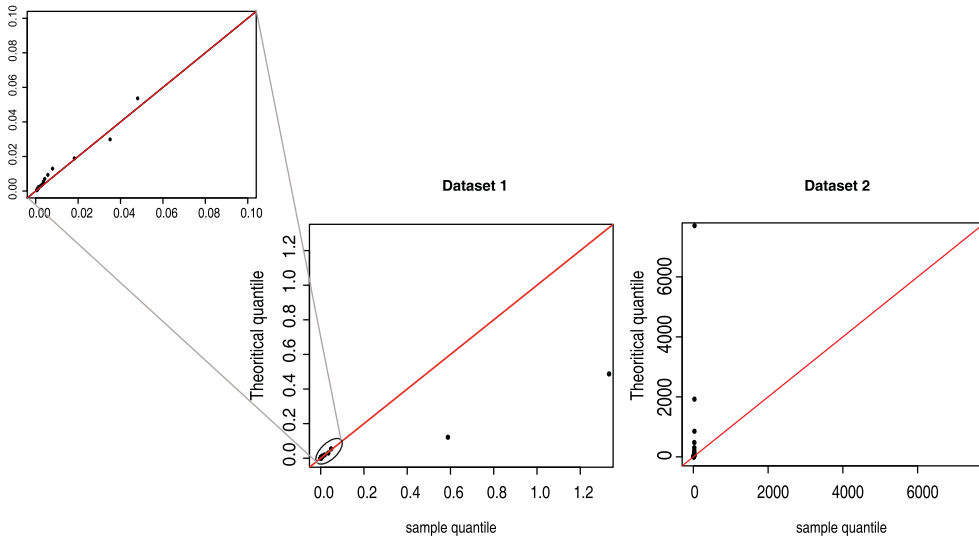


Figure 7. qq-plot between the empirical cdf and fitted one-sided Lévy CDF.

Whereas for dataset 2 the $KS-p_{\text{Bootstrap}}$ -value is very low as compared to the 5% significance level, hence this rejects the null hypothesis. The same conclusion can also be drawn from qq-plots, for dataset 1 the empirical and theoretical cdf values falls near to diagonal line whereas they depart from the diagonal line for dataset 2. We now apply the proposed V_n^* -test to these two datasets, and for comparison purposes, we use JEL and AJEL tests given by Ref. [3]. These tests will be compared with reference to the size of the tests. Nevertheless, the size of the tests based on asymptotic critical value may not be close to the nominal significance level, this behavior depends on the parent family of distribution. Hence in such

cases Bootstrap method, introduced by Ref. [6], can be used effectively. By construction, the bootstrap method is efficient in approximating the cut-off points of the test’s critical region. Despite the fact that the bootstrap method is computationally time consuming, it gives more accurate value of critical levels, as compared to obtaining asymptotic critical values. In our framework, to compute the bootstrap p -value, we present following brief outline of the bootstrap procedure. Denote the data by x_1, \dots, x_n . We fit the model to the data. Then,

- Step 1 Compute the test statistic values, say $t_{V_n^*}, t_{JEL}, t_{AJEL}$.
- Step 2 Use the fitted model to perform parametric bootstrapping.
 - (a) Generate B set of resampled data, denote it as $\hat{x}_1^{(i)}, \dots, \hat{x}_n^{(i)}, i = 1, 2, \dots, B$.
 - (b) For each set of the resampled data, compute the test statistics values $t_{V_n^*}^i, t_{JEL}^i, t_{AJEL}^i$ statistics, for $i = 1, 2, \dots, B$.
- Step 3 For the two-tailed hypothesis, the equal-tail $p_{\text{Bootstrap}}$ -value of the V_n^* test statistic is computed by following relation (see Ref. [7])

$$2 \min \left(\frac{\sum_{i=1}^B \mathbb{I}\{t_{V_n^*}^i < t_{V_n^*}\}}{B}, \frac{\sum_{i=1}^B \mathbb{I}\{t_{V_n^*}^i > t_{V_n^*}\}}{B} \right),$$

where $\mathbb{I}(\cdot)$ is an indicator function, and for JEL and $AJEL$ test, the $p_{\text{Bootstrap}}$ -values are, respectively, be given as

$$\frac{\sum_{i=1}^B \mathbb{I}\{t_{JEL}^i \geq t_{JEL}\}}{B} \quad \text{and} \quad \frac{\sum_{i=1}^B \mathbb{I}\{t_{AJEL}^i \geq t_{AJEL}\}}{B}.$$

The test statistic values and their $p_{\text{Bootstrap}}$ -value are presented in Table 4. We observe from Table 4 that the $p_{\text{Bootstrap}}$ -value for Dataset 1 for all the three tests is higher than the significance value 0.05, hence confirming that the one-sided Lévy distribution cannot be rejected for modeling. However, for Dataset-2, $p_{\text{Bootstrap}}$ -value for V_n^* test is less than the significance value 0.05 and hence the null hypothesis is rejected, which means the one-sided Lévy distribution is not a good choice to model this dataset. In addition to this, the simulated critical points, discussed in Section 4, at 5% SL obtained by V_n^* -test for Datasets 1 and 2 are $(-1.204, 2.555)$ and $(-1.273, 2.309)$, respectively, and V_n^* -test value for Dataset 1 lies within these critical points whereas for Dataset 2, it lies outside the critical points which gives evidence of accepting and rejecting the null hypothesis for Datasets 1 and 2, respectively. Further, on the contrary, the JEL and AJEL tests retain the null hypothesis at 5% SL, but reject the null hypothesis at 10% SL for Dataset 2.

Table 4. Observed value of test statistic values along with their p -values.

		V_n^* -test	JEL	AJEL
Dataset 1	TS-value	0.782	0.316	0.269
	$p_{\text{Bootstrap}}$ -Value	0.547	0.554	0.554
Dataset 2	TS-value	10.952	3.083	2.770
	$p_{\text{Bootstrap}}$ -Value	< 0.001	0.081	0.079

6. Conclusion

Realizing that there are few tests available to test the one-sided Lévy distribution, the proposed test is useful to practitioners. Our Monte Carlo simulation study also supports the claim that the proposed tests show higher power than JEL and AJEL based tests for various alternatives. Finally, the applicability of the proposed test has been shown by considering two real-world datasets. The proposed test is obtained to test whether the sample belongs to one-sided Lévy distribution with zero location parameter, however, this test can be extended to inverse gamma distribution which nests one-sided Lévy distribution.

Notes

1. The bootstrap- p value is obtained similar to the procedure provided in this section later.

Acknowledgments

The authors would like to thank the Associate editor and anonymous referees for suggesting modifications which help us to improve the presentation of this article.

References

- [1] R.J. Adler, R.E. Feldman, and M.S. Taqqu, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Springer Science & Business Media, 1998.
- [2] A. Batsidis, N. Martin, L. Pardo, and K. Zografos, *A necessary power divergence-type family of tests for testing elliptical symmetry*, *J. Stat. Comput. Simul.* 84 (2014), pp. 57–83.
- [3] D. Bhati and S.K. Kattumannil, *Jackknife empirical likelihood test for testing one-sided Lévy distribution*, *J. Appl. Stat.* 47 (2020), pp. 1208–1219.
- [4] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer, 2008.
- [5] I. Dumé, *Geomagnetic flip may not be random after all*, *Physicsworld.com*, March edition, 2006.
- [6] B. Efron, *Bootstrap methods: another look at the jackknife*, *Ann. Stat.* 7 (1979), pp. 1–26.
- [7] B. Efron and R. Tibshirani, *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy*, *Stat. Sci.* (1986), pp. 54–75.
- [8] J.P. Keating, R.E. Glaser, and N.S. Ketchum, *Testing hypotheses about the shape parameter of a gamma distribution*, *Technometrics* 32 (1990), pp. 67–82.
- [9] G.L. Rogers, *Multiple path analysis of reflectance from turbid media*, *J. Opt. Soc. Am. A* 25 (2008), pp. 2879–2883.
- [10] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall/CRC, London, 1994.
- [11] S.S. Shapiro and M.B. Wilk, *An analysis of variance test for normality (complete samples)*, *Biometrika* 52 (1965), pp. 591–611.
- [12] J.A. Villaseñor and E. González-Estrada, *On testing exponentiality based on a new estimator for the scale parameter*, *Braz. J. Probab. Stat.* 34 (2020), pp. 809–820.