Panahi, Hanieh

Article

Model selection test for the heavy-tailed distributions under censored samples with application in financial data

International Journal of Financial Studies

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Panahi, Hanieh (2016) : Model selection test for the heavy-tailed distributions under censored samples with application in financial data, International Journal of Financial Studies, ISSN 2227-7072, MDPI, Basel, Vol. 4, Iss. 4, pp. 1-14, https://doi.org/10.3390/ijfs4040024

This Version is available at:
https://hdl.handle.net/10419/167817

International Journal of
Financial Studies

MDPI

Article

Model Selection Test for the Heavy-Tailed Distributions under Censored Samples with Application in Financial Data

Hanieh Panahi

Department of Mathematics and Statistics, Lahijan Branch, Islamic Azad University, Lahijan 4416939515, Iran; Panahi@liau.ac.ir; Tel.: +98-1342229081

Academic Editor: Nicholas Apergis
Received: 19 June 2016; Accepted: 1 December 2016; Published: 13 December 2016

Abstract: Numerous heavy-tailed distributions are used for modeling financial data and in problems related to the modeling of economics processes. These distributions have higher peaks and heavier tails than normal distributions. Moreover, in some situations, we cannot observe complete information about the data. Employing the efficient estimation method and then choosing the best model in this situation are very important. Thus, the purpose of this article is to propose a new interval for comparing the two heavy-tailed candidate models and examine its suitability in the financial data under complete and censored samples. This interval is equivalent to encapsulating the results of many hypotheses tests. A maximum likelihood estimator (MLE) is used for evaluating the parameters of the proposed heavy-tailed distribution. A real dataset representing the top 30 companies of the Tehran Stock Exchange indices is used to illustrate the derived results.

Keywords: asymptotic distribution; censored sample; heavy-tailed distribution; model selection test; Tehran Stock Exchange

JEL Classification: C12; C13; C24

1. Introduction

During recent years, heavy-tailed distributions have been considered in the form of an attractive title for various research and studies. For some works on these distributions, we refer to, among others, [1–4]. These distributions have good statistical and reliability properties. Due to its practicality, the heavy-tailed distributions can be used for many applied sciences including economics, finance, econometrics, statistics, risk management and insurance. The inferential results under financial modeling have been developed by several authors; see, for example, [5–8]. There are different heavy-tailed distributions. The question then arises which of them is the best for modeling the proposed financial data. Thus, in this paper, we want to choose the best distribution using the new model selection test. There are different model selection tests for discriminating between two complete models. In almost all of the tests and criteria for model selection, the maximum likelihood estimator and maximized likelihood function have an essential role. For example, Kundu et al. [9] compared the log-normal and generalized exponential distribution using maximized likelihood method, Dey and Kundu [10] considered the problem of discriminating among the log-normal, Weibull and generalized exponential distributions, Cox [11] modified the classical hypothesis testing to compare the non-nested hypothesis and Vuong [12] tested the two models using the log-likelihood ratio of the models. The results in Vuong [12] have been extended and applied in a number of ways, including [13–18]. Moreover, in experimental study, it is quite common that complete data are not observed. Data obtained from such experiments are called censored data. Based on the studies of the

real data such as financial data, it is observed that some of the first data may not always be available. This incomplete data is called left censored data. Some of the work on left censoring was conducted by [19–22]. Based on the censored data, two heavy-tailed distributions may provide very similar data fit to a given data set. In other words, the distance between the two fitted distributions can be very small, and it may be very difficult to discriminate between them. Therefore, the main aim of this paper is to propose a new model selection test for comparing the heavy-tailed distribution under censored data. Although several articles have been done on the heavy-tailed distribution, we have not come across any articles under the model selection test for the heavy-tailed distributions under censored samples (HTDC). Thus, the main objective of this paper is the determination the best model for the financial data. In Section 2, we first provide the main definitions and assumptions. Section 3 provides an interval as a new model selection test (NMST) under censored sample. The heavy tail properties and the method for determining the heavy-tailed distribution are presented in Section 4. The application of the NMST of the Tehran Stock Exchange is presented in Section 5, which provides a comparison of different heavy-tailed rival models as well as different censoring schemes, and we finally conclude the paper in Section 6.

2. Main Definitions and Assumptions

In this section, we present the definitions and assumptions that are necessary for the proposed model selection test. Consider a sample of random variables $X_1, ..., X_n$ having probability density function $h(.)$. Let us consider two rival models:

$$F^{\alpha} = \{f^{\alpha}(.),\ \alpha \in \mathrm{M} \subset R^{p}\} = (f) \text{ and } G^{\beta} = \left\{g^{\beta}(.),\ \beta \in \mathrm{B} \subset R^{q}\right\} = (g).$$

Definition 1. *(i) (f) and (g) are non overlapping if $(f) \cap (g) = \phi$; (ii) (f) is nested in (g) if $(f) \subset (g)$; (iii) (f) is well-specified if there is a value $\alpha_0 \in \mathrm{M}$ such that $f^{\alpha_0}(.) = h$; otherwise, it is misspecified.*

Definition 2. *Given two probability distributions, $\widetilde{\nu} << \widetilde{\mu}$, the relative entropy of ν with respect to μ, or the Kullback-Leibler divergence (KL) of $\widetilde{\nu}$ from $\widetilde{\mu}$, is*

$$KL(\widetilde{\nu}, \widetilde{\mu}) = D(\widetilde{\mu} \| \widetilde{\nu}) = -E_{\mu}\left(\ln \frac{d\widetilde{\nu}}{d\widetilde{\mu}}\right).$$

If $\widetilde{\nu}$ is not absolutely continuous with respect to $\widetilde{\mu}$, then $KL(\widetilde{\nu}, \widetilde{\mu}) = D(\widetilde{\mu} \| \widetilde{\nu}) = \infty$.

The minimum assumptions for non-degenerate interval M are:

$\Re_1$: The parameter space M is an open interval in R.

$\Re_2$: $(\partial/\partial\alpha) f(x, \alpha)$ is a strictly monotone function on M for each x.

$\Re_3$: For all $\alpha \in M$, the partial derivative $(\partial/\partial\alpha) f(x, \alpha)$, is integrable on R, and the partial derivative, $(\partial/\partial\alpha) F(x, \alpha)$, exists for $x \in \chi$, and satisfies

$$(\partial/\partial\alpha) F(x, \alpha) = \int_{-\infty}^{x} (\partial/\partial\alpha) f(u, \alpha)\, du.$$

$\Re_4$: For every α, we have,

$$\left|\frac{\partial}{\partial\alpha} f^{\alpha}(x)\right| \leq \Upsilon_1,\ \left|\frac{\partial^2}{\partial\alpha^2} f^{\alpha}(x)\right| \leq \Upsilon_2 \text{ and } \left|\frac{\partial^3}{\partial\alpha^3} f^{\alpha}(x)\right| \leq \Upsilon_3;\ \text{where, } \int \Upsilon_i d\mu(x) < \infty;\ i = 1, 2, 3$$

and μ is taken to be a Lebesgue measure.

$\Re_5$: For every α, $[F^{\alpha}(x)]^{-1}$ is bounded by $\aleph(x)$ respectively, where $E(\aleph(X)) \leq \iota$; and ι is a positive constant.

$\Re_6$: For every α, we have, $\wp = \int \left(\frac{\partial}{\partial\alpha} \ln f(x, \alpha)\right)^2 f(x, \alpha) d\mu(x) < \infty$.

3. New Model Selection Test (NMST) For HTDC

Let $X_{c:n}, ..., X_{n:n}$ denote the truncated order statistics observed from an experimental test involving n units taken from an $f^{\alpha}(x)$ distribution. To simplify the notation, we will use X_i in place of $X_{i:n}$. Then, the likelihood function of $(X_c, ..., X_n)$ can be obtained as

$$l(\alpha) \propto \prod_{i=c}^{n} f^{\alpha}(x_i)\ [F^{\alpha}(x_c)]^{c-1}, \tag{1}$$

where $f^{\alpha}(x)$ and $F^{\alpha}(x)$ are the probability density function and cumulative distribution function of heavy-tailed distribution respectively. We are interested in testing the following hypotheses set to discriminate between H_0 and H_f or H_g, the NMST for left censored data.

H_0 : The two proposed heavy-tailed models (F^{α} and G^{β}) are equivalent, and against, H_f: F^{α} is better than G^{β} in the sense of the closeness to the true model, or H_g: F^{α} is worse than G^{β}.

Theorem 1. ***(NMST for HTDC):*** *Using the conditions $\Re_1$–$\Re_6$ and the asymptotic distribution of the MLE (see Appendix A), the new interval as a model selection test for HTDC is given by*

$$\left[\eta_c(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) - n^{-1/2} z_{\alpha/2}\hat{\omega}_c,\ \eta_c(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) + n^{-1/2} z_{\alpha/2}\hat{\omega}_c\right], \tag{2}$$

where $\eta_c(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) = -\frac{1}{n}\left[L_n^{f/g}(\hat{\alpha}_n, \hat{\beta}_n) - (p-q)\right] = -\frac{1}{n}\left[L_n^{f}(\hat{\alpha}_n) - L_n^{g}(\hat{\beta}_n) - (p-q)\right]$. Now, using the Equation (1), we have

$$\eta_c(f^{\hat{\alpha}_n}, g^{\hat{\beta}_n}) = -\frac{1}{n}\left[\left(\sum_{i=c}^{n} \ln\frac{f^{\hat{\alpha}_n}(x_i)}{g^{\hat{\beta}_n}(x_i)} + (c-1)\ln\frac{F^{\hat{\alpha}_n}(x_c)}{G^{\hat{\beta}_n}(x_c)}\right) - (p-q)\right],$$

where p and q are the number of parameters in the heavy-tailed models and $\hat{\alpha}_n$ and $\hat{\beta}_n$ are the quasi maximum likelihood estimators under censored sample. In addition, Z_{α} is α^{th} quantile of standard normal distribution and $\hat{\omega}_c^2$ satisfies

$$\hat{\omega}_c^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\ln\frac{f^{\hat{\alpha}_n}(w_i)}{g^{\hat{\beta}_n}(w_i)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\ln\frac{f^{\hat{\alpha}_n}(w_i)}{g^{\hat{\beta}_n}(w_i)}\right)\right)^2 + \left(\frac{c-1}{n}\right)\left[\frac{1}{c-1}\sum_{i=1}^{c-1}\left(\ln\frac{f^{\hat{\alpha}_n}(z_i)}{g^{\hat{\beta}_n}(z_i)}\right)^2 - \left(\frac{1}{c-1}\sum_{i=1}^{c-1}\ln\frac{f^{\hat{\alpha}_n}(z_i)}{g^{\hat{\beta}_n}(z_i)}\right)^2\right]. \tag{3}$$

Proof. Based on Theorem B in Appendix B, it is observed that the difference of the log-likelihood functions of the two truncated rival models (data are left censored) converges in distribution to the normal distribution. Thus, it is sufficient to find the empirical form of the ω_{*c}^2 as $\hat{\omega}_c^2$.

Using the missing information principle [23], the observed information can be written as

$$\sum_{i=c}^{n}\ln f^{\alpha}(x_i) = \sum_{i=1}^{n}\ln f^{\alpha}(w_i) - \sum_{i=1}^{c-1}\ln f^{\alpha}(z_i|X), \tag{4}$$

where $w = (w_1, ..., w_n)$ = the complete data, $z = (z_1, ..., z_{c-1})$ is the complete data of size from the right population with density functions:

$$h_1^* = \frac{f^{\alpha}(z)}{F^{\alpha}(x_c)};\ z < x_c.$$

For simplicity, we use $f^{\alpha}(z_i)$ instead of $f^{\alpha}(z_i|X)$ in what follows. Thus, the $Var\left(\frac{1}{n}L_n^{f/g}(\hat{\alpha}_n, \hat{\beta}_n)\right)$ can be expressed as

$$\begin{aligned} Var\left[\tfrac{1}{n}L_n^{f/g}(\hat{\alpha}_n,\hat{\beta}_n)\right] &= Var\left[\tfrac{1}{n}\left(\sum_{i=c}^{n}\ln\tfrac{f^{\hat{\alpha}_n}(X_i)}{g^{\hat{\beta}_n}(X_i)}+(c-1)\ln\tfrac{F^{\hat{\alpha}_n}(X_c)}{G^{\hat{\beta}_n}(X_c)}\right)\right] \\ &= Var\left[\tfrac{1}{n}\left(\sum_{i=1}^{n}\ln\tfrac{f^{\hat{\alpha}_n}(W_i)}{g^{\hat{\beta}_n}(W_i)}-\sum_{i=1}^{c-1}\ln\tfrac{f^{\hat{\alpha}_n}(Z_i)}{g^{\hat{\beta}_n}(Z_i)}+(c-1)\ln\tfrac{F^{\hat{\alpha}_n}(x_c)}{G^{\hat{\beta}_n}(x_c)}\right)\right]. \end{aligned}$$

If $\frac{c-1}{n}\to p$ as $n\to\infty$ such that $X_c=\zeta_n\to\zeta$ in probability and $\hat{\alpha}_n\xrightarrow{P}\alpha_*$, then

$$\omega_{*c}^2 = Var\left(\ln\frac{f^{\alpha_*}(W)}{g^{\beta_*}(W)}\right)+pVar\left(\ln\frac{f^{\alpha_*}(Z)}{g^{\beta_*}(Z)}\right), \tag{5}$$

where the empirical form of (5) satisfies

$$\hat{\omega}_c^2=\tfrac{1}{n}\sum_{i=1}^{n}\left(\ln\tfrac{f^{\hat{\alpha}_n}(w_i)}{g^{\hat{\beta}_n}(w_i)}\right)^2-\left(\tfrac{1}{n}\sum_{i=1}^{n}\left(\ln\tfrac{f^{\hat{\alpha}_n}(w_i)}{g^{\hat{\beta}_n}(w_i)}\right)\right)^2+(\tfrac{c-1}{n})\left[\tfrac{1}{c-1}\sum_{i=1}^{c-1}\left(\ln\tfrac{f^{\hat{\alpha}_n}(z_i)}{g^{\hat{\beta}_n}(z_i)}\right)^2-\left(\tfrac{1}{c-1}\sum_{i=1}^{c-1}\ln\tfrac{f^{\hat{\alpha}_n}(z_i)}{g^{\hat{\beta}_n}(z_i)}\right)^2\right].$$

The proposed interval has the property of

$$P_h\Big[\mathrm{A_n}<\Delta_c(f^{\hat{\alpha}_n},g^{\hat{\beta}_n})<\mathrm{B_n}\Big]\to 1-\alpha\,,$$

where $\Delta_c(f^{\hat{\alpha}_n},g^{\hat{\beta}_n})$ is the difference of the expected Kullback-Leibler divergence (KL) of $f^{\hat{\alpha}_n}$ and $g^{\hat{\beta}_n}$ under censored data and

$$\mathrm{A_n}=\eta_c(f^{\hat{\alpha}_n},g^{\hat{\beta}_n})-n^{-1/2}z_{\alpha/2}\hat{\omega}_c;\ \mathrm{B_n}=\eta_c(f^{\hat{\alpha}_n},g^{\hat{\beta}_n})+n^{-1/2}z_{\alpha/2}\hat{\omega}_c,$$

where P_h represents the probability with density h.

Decision Rule

An important problem in statistics concerning a sample of n observations is to test whether these observations come from a specified distribution. The Vuong test is one of the important tests for model selection. However, if the rival models are very close (not equivalent) to the true model, then this test can suffer from distortions. Therefore, this section suggests a simple model selection procedure based on the likelihood ratio statistic under censored data that is easy to compute and has an asymptotic standard normal distribution. The proposed interval is easy to compute and interpret as the following steps:

Step 1: Choose the rival models and calculate the quasi maximum likelihood estimates for the unknown parameters.

Step 2: Compute the likelihood ratio statistic under censored data ($l=\eta_c(f^{\hat{\alpha}_n},g^{\hat{\beta}_n})$).

Step 3: Obtain $\hat{\omega}_c$ and then construct the proposed interval $\Im=\Big[l-n^{-1/2}z_{\alpha/2}\hat{\omega}_c,\ l+n^{-1/2}z_{\alpha/2}\hat{\omega}_c\Big]$ using the α^{th} quantile of standard normal distribution (Z_α).

Step 4: Interpret the proposed test ($\Im$) as

- ***(i)*** If the calculated interval $\Im$ includes zero, it can be concluded that both proposed models (F^α and G^β) are equivalent.
- ***(ii)*** If both bounds of $\Im$ are negative, which indicates that F^α is better than G^β to estimate the true model.
- ***(iii)*** Finally, if both bounds of $\Im$ are negative, then we conclude that G^β is better than F^α to estimate the true model.

Our approach enlightens the variability of any criterion based on log-likelihood function.

4. Heavy Tail Properties

Heavy-tailed distributions are the important distributions in economics and finance. In this section, we check the heavy tail properties for different distributions.

Definition 3. *The distribution F(.) from the random variable X is considered to be heavy tail if and only if*

$$\int_R e^{-\lambda x} F(x)dx = \infty;\ for\ all\ \lambda > 0.$$

Definition 4. *A continuous distribution function is considered to be heavy tail if the generating moment function is infinite.*

Thus, we can check the heaviness using different criteria such as:

i. Based on definition 4, if only some or if none of the moments of distributions exist, then it has the heavy tail.

ii. If $\limsup_{x\to\infty} \frac{\hbar(x)}{x} = 0$, then the distribution has the heavy tail. Here, $\hbar(x)$ is the hazard function.

iii. If $\hbar*(t)$ is the decreasing function for increasing value of t, then the distribution has the heavy tail, where $\hbar*(t) = \frac{d}{dt}\hbar(\mathrm{t})$.

iv. If the distribution is heavy tail, then $\Im = \frac{Var(X)}{E(X)^2} \geq 1$. Note that the converse does not hold.

v. The distribution has the heavy tail, if

$$1 - F(x) \leq ae^{-bx};\ x \geq 0, a > 0, b > 0.$$

Here, we say that $F(x)$ has a light tail.

4.1. Heavy-Tailed Distributions

In this subsection, we consider different heavy-tailed distributions and then check the heaviness property using the different criteria.

4.1.1. Generalized Extreme Value Distribution (GEVD)

The cumulative distribution function (CDF) of GEVD is given by

$$F(x) = \begin{cases} e^{-(1-\mathrm{k}\frac{x-\xi}{\alpha})^{1/k}}; & k \neq 0 \\ e^{-e^{-\frac{x-\xi}{\alpha}}}; & k = 0 \end{cases}. \tag{6}$$

Variable X is bounded by $(\xi + \alpha)/k$ from above, if $k > 0$ and from below if $k < 0$, where $\xi \in R$ and $\alpha > 0$. We have three cases of this distribution as

- Weibull distribution ($k > 0$),
- Ferechet distribution ($k < 0$),
- Gumbel distribution ($k = 0$).

We now want to check the heavy tail property. Using the *ii* and *iii* criteria, it is observed that the Ferechet-Weibull distribution and the Weibull distribution with $0 < \beta < 1$ have a heavy tail. However, based on the v criterion, the Gumbel distribution does not have the heavy tail property.

4.1.2. Pareto Distribution

The Pareto distribution is a skewed, heavy-tailed distribution that is sometimes used to model the distribution of incomes. This distributional model is important in applications because many datasets

are observed to follow a power law probability tail, at least approximately, for large values of x. Stable distributions with index α are also asymptotically Pareto in their probability tails, and this fact has been frequently used to develop estimators for those distributions. The CDF of Pareto is given by

$$F(x) = 1 - \left(\frac{\alpha}{x+\alpha}\right)^k;\ x > 0. \tag{7}$$

The hazard function of the Pareto distribution, $\frac{k}{x+\alpha}$, is a decreasing function for positive values of k and α. Thus, using the *ii* and *iii* criteria, it has a heavy tail.

4.1.3. Log-Normal Distribution

A log-normal distribution is applied as the standard model for financial data. It is used in many different fields of study, such as economics, metrology, biology, neuroscience and engineering. The density function of a log-normal distribution, with shape parameter $\sigma > 0$ and scale parameter $\mu > 0$ is

$$f(x) = \frac{1}{\sqrt{2\pi}x\sigma} e^{-\frac{1}{2}\left(\frac{\ln x - \ln \mu}{\sigma}\right)^2};\ x > 0. \tag{8}$$

The tail heaviness property of the log-normal distribution depends on the variance. In other words, based on the *iv* criterion, we can write,

$$\varpi = \frac{Var(X)}{E(X)^2} = e^{\sigma^2} - 1. \tag{9}$$

The ϖ is longer than 1, if $\sigma > 0.8226$ or $\sigma < 0.8226$.

4.1.4. Burr Type XII Distribution

Burr [24] introduced twelve cumulative distribution functions with the primary purpose of fitting distributions to real data. One of the most important of them is the Burr Type XII distribution. The cumulative distribution function of the Burr Type XII is given by

$$F(x) = 1 - (1 + x^\beta)^{-\alpha};\ x > 0. \tag{10}$$

Here, $\alpha > 0$ and $\beta > 0$ are the two shape parameters. The shape of the hazard rate function of the Burr Type XII distribution depends only on parameter β. Its capacity to assume various shapes often permits a good fit when used to describe biological, financial, engineering or other experimental data. It also approximates the distributional form of normal, log-normal, gamma, logistic, and several Pearson-type distributions. For instance, the normal density function may be approximated as a Burr Type XII distribution with β = 4.8544 and α = 6.2266 and the gamma distribution with shape parameter 16 can be approximated as a Burr Type XII distribution with β = 3 and α = 6, and the log-logistic distribution is a special case of the Burr Type XII distribution. In addition, using the *i*, *ii* and *iii* criteria, it is observed that the Burr Type XII has a heavy tail.

4.1.5. Dugum and Singh-Maddala Distribution

The Dugum and Singh-Maddala distributions are the special case of the generalized Beta kind 2 (GB_2) distribution. The CDFs of these distributions are given by, respectively:

$$F(x) = \left(1 + \left(\frac{\beta}{x}\right)^\alpha\right)^{-\gamma};\ x > 0, \tag{11}$$

and

$$F(x) = 1 - \left(1 + \left(\frac{x}{\beta}\right)^\alpha\right)^{-\gamma};\ x > 0, \tag{12}$$

Here, all three of the parameters are positive. In addition, the r^{th} moment of these distributions can be written as

$$E(X^r) = \frac{\beta^r}{(\gamma-1)!}\Gamma(\frac{r}{\alpha}+1)\Gamma(1-\frac{r}{\alpha})$$

and

$$E(X^r) = \beta^r\gamma B(\frac{r}{\alpha}+\gamma, 1-\frac{r}{\alpha}),$$

where $\Gamma(.)$ and $B(.,.)$ denote the Gamma distribution and the Beta distribution, respectively. Using the i criterion, it is observed that the moments of these distributions only exist for values of $-\gamma\alpha < r < \alpha$. It indicates that these distributions have potentially tail heaviness properties.

5. Application of the NMST of Tehran Stock Exchange

In this section, the data set of daily returns of the top 30 companies from the Tehran Stock Exchange indices was used to study the performance of the proposed model selection test. All the programs are written in *R*. The mean, standard deviation, skewness and kurtosis of this data are 2.937687, 0.261701, 0.273005 and 1.790659, respectively. It is observed that the skewness and kurtosis are not close to zero and three, respectively. Thus, the data set has a higher peak, fatter tail and skewness in comparison to the Normal distribution. For more study, we demonstrate how it deviates from the Normal distribution using the Shapiro-Wilk (S-W) test, Kolmogrov-Smirnov (K-S) test, Anderson-Darling (A-D) test and Jarque-Bera (J-B) test. For each test, the null hypothesis is that the data are normally distributed. If the p-value is less than the significant level (0.05) of the given hypothesis, the null hypothesis will be rejected. For computing the mentioned test, we use the *nortest* and *tseries* in the *R* package. The results are provided in Table 1. Based on Table 1, we observe that the data do not follow the Normal distribution. Thus, we use the heavy-tailed distribution for modeling the data. First, we check the adequacy of the Weibull (*We*) distribution, Pareto (*Pa*) distribution, Burr Type XII (*BXII*) distribution, log-normal (*LN*) distribution, Dagum (*Da*) distribution and Singh-Maddala (*S-M*) distribution using three different well-established model selection criteria such as K-S minimum distance criterion, Akaike information criterion ($AIC = -2\sum_{i=1}^{n} log\, f^{\hat{\alpha}_n}(x_i) + 2p$), Bayesian information criterion ($BIC = -2\sum_{i=1}^{n} log\, f^{\hat{\alpha}_n}(x_i) + p\log n$) and maximum log-likelihood criterion (LL). We select the best model among all competitive distributions that has the smallest AIC, BIC and K-S distance and the greatest LL values. We first estimate the unknown parameters using MLEs. The results are presented in Table 2. It is clear that the *Da* and *LN* distributions have comparatively better fitting for the present data set. The *We* and *S-M* distributions also have a good fit. We provide the Probability-Probability (P-P) plots for different distributions in Figures 1–6. Moreover, the empirical survival function and the fitted survival functions are presented in Figure 7. Therefore, based on Figures 1–7, it is observed that the *BXII* and *Pa* distributions do not fit the data reasonably well, and hence, they cannot be used to obtain inferential results from the considered data set. Using different model selection criteria, we can compare the proposed distributions. However, these criteria have some disadvantages. For example, the LL criterion assumes that the number of parameters in each competitive model is the same. In addition, one problem with AIC and BIC are that their values have no intrinsic meaning; in particular, AIC and BIC are not invariant to a one-to-one transformation of the random variables and values of AIC and BIC depend on the number of observations. Thus, we consider the NMST for comparing the heavy-tailed distributions.

Table 1. Different Normality tests for the proposed data.

	S-W	K-S	A-D	J-B
Value of test	0.4659	0.2957	6.0576	953.5819
p-value	9.605×10^{-11}	2.493×10^{-9}	3.471×10^{-15}	$<2.2 \times 10^{-16}$

S-W: Shapiro-Wilk test; K-S: Kolmogrov-Smirnov test; A-D: Anderson-Darling test; J-B: Jarque-Bera test.

Table 2. Estimated parameters, AIC values, BIC values and log-likelihood values for different distribution functions.

Models	Parameters	MLE	AIC	BIC	LL
We	α β	12.07339 3.059887	64.83431	71.86922	−30.41716
Pa	α k	5037634 1715197	1038.656	1045.691	−517.3281
BXII	α β	0.071667 12.98343	1072.124	1079.159	−534.0619
LN	mean s.d.	1.073703 0.088294	36.6373	43.67221	−16.31865
Da	α β γ	13.39737 2.142820 37.14274	36.8876	47.43996	−17.4438
S-M	α β γ	12.22984 4.134818 40.52715	66.50122	77.05358	−30.25061

MLE: maximum likelihood estimator; AIC: Akaike information criterion; BIC: Bayesian information criterion; LL: maximum log-likelihood criterion.

Now, we check the results using the proposed interval for model selection. We consider four cases of rival models as:

(1) *Da* (f) and *LN* (g),
(2) *Da* (f) and *We* (g),
(3) *We* (f) and *S-M* (g),
(4) *Da* (f) and *S-M* (g).

Based on the estimated values, we construct the proposed interval. This interval for the above four cases are (1.9878339, 2.0104310), (−2.748678, −2.614607), (−4.2395868, −4.1287672) and (−172.43378, −161.83113), respectively. For Case 1, it is observed that both limits of the tracking interval are positive, which indicates that the *LN* is better than the *Da* distribution to estimate the true model for this data. However, the length of this interval is small, so we can conclude that the two models are similar to estimate the true model (as expected). For Cases 2–4, both limits of the tracking interval are negative, so the model (f) is better than the model (g). It is observed that this interval selects the correct model well. In addition, computational steps of this interval are simple. Now, we suppose that some of the data are missed (censored). We generate artificially left censored data from the data set as

- *Scheme 1*: $n = 249,\ c = 5$ (The first 5 pieces of data are not observed).
- *Scheme 2*: $n = 249,\ c = 20$ (The first 20 pieces of data are not observed).
- *Scheme 3*: $n = 249,\ c = 80$ (The first 60 pieces of data are not observed).

Here, n is the complete sample size and c is the number of the left censored data. Based on Case 1, the proposed intervals for Schemes 1–3 are (1.952583, 2.036342), (1.974464, 2.060377) and (2.141464, 2.212098), respectively. Similarly, for Case 2, the intervals are (−2.757090, −2.620894), (−2.768494,

−2.623881) and (−2.680060, −2.497014), respectively. For Case 3, the results are (−4.246638, −4.134079), (−4.264661, −4.145781) and (−4.304638, −4.144865), respectively, and for Case 4, the proposed intervals are (−171.1225, −160.5491), (−167.1976, −156.6924) and (−148.70633, −139.26066), respectively. It is observed that the results are similar to the complete data.

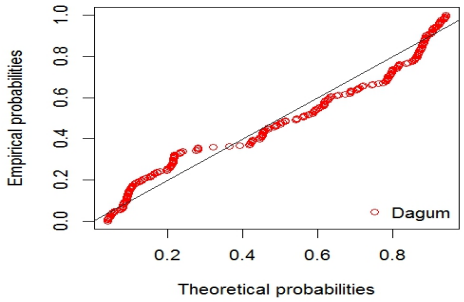

Figure 1. The Probability-Probability (P-P) plot for Dagum distribution.

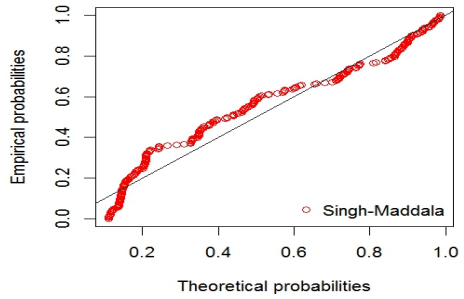

Figure 2. The P-P plot for Singh-Maddala distribution.

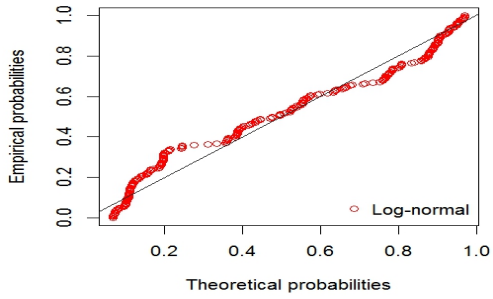

Figure 3. The P-P plot for log-normal distribution.

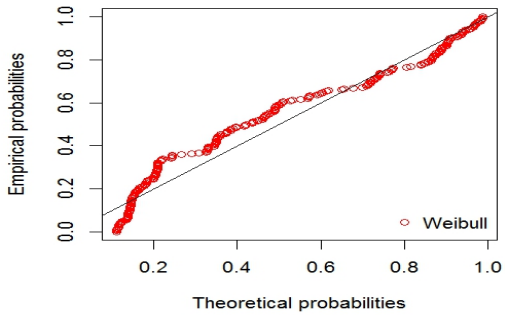

Figure 4. The P-P plot for Weibull distribution.

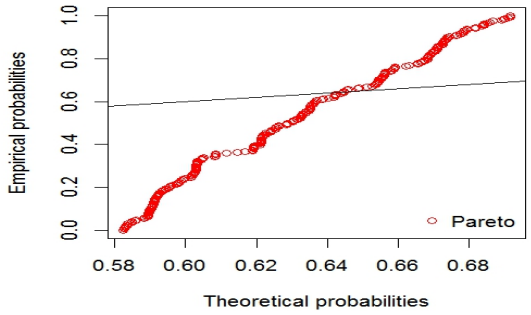

Figure 5. The P-P plot for Pareto distribution.

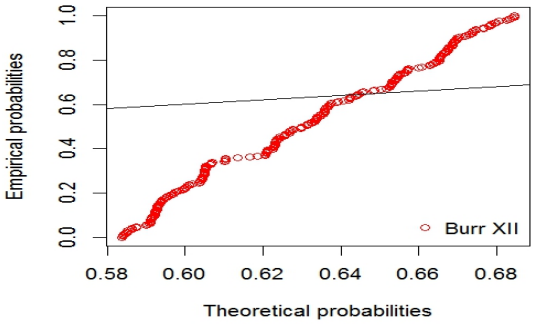

Figure 6. The P-P plot for Burr XII distribution.

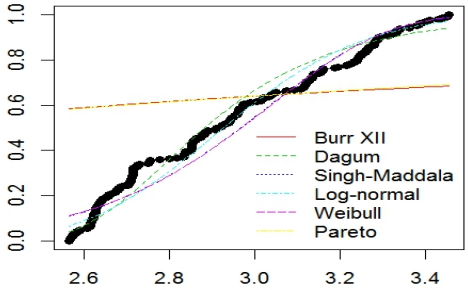

Figure 7. Empirical survival function and the fitted survival functions.

6. Conclusions

The heavy-tailed distributions are the most important distributions in several applied sciences such as economics, financial engineering and mathematical finance. Moreover, in many situations, we cannot observe the complete information about the data. In this situation, the problem of choosing the correct distribution becomes more difficult. There are different criteria such as AIC, BIC, LL and K-S distance for comparing the models. These criteria have some disadvantages. Thus in this paper, we have proposed a new model selection test for comparing the heavy-tailed distributions under complete and censored data. This interval enlightens the unavoidable variability of any criterion based on log-likelihood ratio such as AIC, BIC and their variants. Based on this test, we can make the best possible decision based on whatever data are available at hand. The computational steps of NMST are easy to compute and could be very useful for censored data. We hope that the new model selection test will attract wider application in all areas of research.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Theorem A. ***(Asymptotic distribution of the Maximum Likelihood Estimator):*** *Assume that $f^{\alpha}(x)$ is a well-specified model satisfying the conditions $\Re_1$–$\Re_6$ and $(\hat{\alpha}_n = \max_{\alpha \in \mathrm{M}} L_n^f(\alpha))$. Then, as $n \to \infty$, the asymptotic distribution of the MLE, $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$, is, $N(0, J_f^{-1})$, where, $J_f \equiv \wp + p\xi$.*

Proof. Suppose that $X_c, ..., X_n$ are distributed as the order statistic of a random sample of size n-c + 1 from a truncated distribution at x_c by probability density function (pdf) h^*. Now, let $\frac{c}{n} \to p$ as $n \to \infty$, such that $X_c (= \zeta_n)$ converges in probability to ζ, the p^{th} percentile of true distribution. Now, based on the Taylor expansion of $n^{-1}\frac{\partial L_n^f(\alpha)}{\partial \alpha}$ around $\alpha = \alpha_0$ as ([25]):

$$\begin{aligned} n^{-1}\tfrac{\partial L_n^f(\alpha)}{\partial \alpha} &= n^{-1}\tfrac{\partial L_n^f(\alpha)}{\partial \alpha}|_{\alpha=\alpha_0} + n^{-1}(\alpha - \alpha_0)\tfrac{\partial^2 L_n^f(\alpha)}{\partial \alpha\, \partial \alpha'}|_{\alpha=\alpha_0} + o_p(1) \\ &= A_1 + A_2(\alpha - \alpha_0) + o_p(1), \end{aligned} \tag{A1}$$

where, using the observed information (4), we can write

$$A_1 = \frac{1}{n}\left\{ (\sum_{i=1}^{n} \frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(w_i)) - \sum_{i=1}^{c-1} \frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(z_i) - (c-1)\frac{\partial}{\partial \alpha_0} \ln(F^{\alpha}(x_c)) \right\} \equiv \frac{1}{n}(A_1^* - A_1^{**}).$$

Here, $\frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(x)$ means that $\frac{\partial}{\partial \alpha} \ln f^{\alpha}(x)|_{\alpha=\alpha_0}$. Thus, from Cramér [26], $\frac{1}{n}A_1^* = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(w_i) \xrightarrow{P} 0$, we will prove that

$$\frac{1}{n}A_1^{**} = \frac{1}{n}\sum_{i=1}^{c-1} \frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(z_i) - (c-1)\frac{\partial}{\partial \alpha_0} \ln(F^{\alpha}(x_c)) \xrightarrow{P} 0.$$

We can rewrite A_1^{**} as

$$A_1^{**} = \sum_{i=1}^{c-1} \tfrac{\partial}{\partial \alpha_0} \ln f^{\alpha}(z_i) - \sum_{i=1}^{c-1} E\Big(\tfrac{\partial}{\partial \alpha_0} \ln f^{\alpha}(Z_i)\Big) + \sum_{i=1}^{c-1} E\Big(\tfrac{\partial}{\partial \alpha_0} \ln f^{\alpha}(Z_i)\Big) - (c-1)\tfrac{\partial}{\partial \alpha_0} \ln(F^{\alpha}(x_c)).$$

From $\Re_3$, we have

$$E\left(\frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(Z)\right) = \int_{-\infty}^{x_c} \frac{\partial}{\partial \alpha_0} \ln f^{\alpha}(z) \frac{f^{\alpha}(z)}{F^{\alpha}(x_c)}\, d\mu(z) = \frac{\frac{\partial}{\partial \alpha_0} F^{\alpha}(x_c)}{F^{\alpha}(x_c)} = \frac{\partial}{\partial \alpha_0} \ln(F^{\alpha}(x_c)). \tag{A2}$$

Thus, $\frac{A_1^{**}}{n} \xrightarrow{P} 0$. Now, by using Slutsky's theorem, the result follows ($A_1 \xrightarrow{P} 0$). Similarly, we consider, $A_2 = \frac{1}{n}(A_2^* - A_2^{**})$, where

$$A_2^* = \sum_{i=1}^{n} \frac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln f^{\alpha}(w_i)$$

and

$$A_2^{**} = \sum_{i=1}^{c-1} \frac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln f^{\alpha}(z_i) - (c-1)\frac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln(F^{\alpha}(x_c)),$$

We know that $\frac{A_2^*}{n} \xrightarrow{P} -\wp$ and

$$\begin{aligned} \tfrac{A_2^{**}}{n} = \quad & \tfrac{c-1}{n}\left\{ \tfrac{1}{c-1}\left(\sum_{i=1}^{c-1} \tfrac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln f^{\alpha}(z_i) - \sum_{i=1}^{c-1} E\Big(\tfrac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln f^{\alpha}(Z_i)\Big) \right) \right\} \\ & -\tfrac{1}{n}\left\{ (c-1)\tfrac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln(F^{\alpha}(x_c)) - \sum_{i=1}^{c-1} E\Big(\tfrac{\partial^2}{\partial \alpha_0 \partial \alpha_0'} \ln f^{\alpha}(Z_i)\Big) \right\}. \end{aligned} \tag{A3}$$

The first term in (A3) converges in probability to zero. Thus, based on (A2) and after some simplifications, we have

$$\frac{1}{c-1}\sum_{i=1}^{c-1}\Big\{\frac{\partial^2}{\partial\alpha_0\partial\alpha_0'}\ln(F^{\alpha}(y_c)) - E\Big(\frac{\partial^2}{\partial\alpha_0\partial\alpha_0'}\ln f^{\alpha}(Z_i)\Big)\Big\} = \frac{1}{c-1}\sum_{i=1}^{c-1} Var\Big(\frac{\partial}{\partial\alpha_0}\ln f^{\alpha}(Z_i)\Big) = \Im*,$$

where $\Im*$ converges to bounded values, ξ; consequently, we can write

$$A_2 = \frac{1}{n}(A_2^* - A_2^{**}) \xrightarrow{P} -J_f,$$

where

$$J_f \equiv \wp + p\xi.$$

Now, from (A1), we have

$$\sqrt{nJ_f}(\hat{\alpha}_n - \alpha_0) = \frac{\sqrt{n}A_1/\sqrt{J_f}}{-A_2/J_f},$$

where $-A_2/J_f \xrightarrow{P} 1$ and, using the Slutsky theorem, we can show that the numerator is asymptotically $N(0,1)$. Thus, we conclude that $\sum_{i=c}^{n}\frac{\partial}{\partial\alpha_0}\ln f^{\alpha}(x_i) + (c-1)\frac{\partial}{\partial\alpha_0}\ln(F^{\alpha}(x_i)) \xrightarrow{D} N(0,\ \wp + p\xi)$ and the proof is complete.

Appendix B

Theorem B. *If $f^{\alpha_*}(.)$ and $g^{\beta_*}(.)$ are the two candidate models ($f^{\alpha_*}(.) \neq g^{\beta_*}(.)$), we can write*

$$\sqrt{n}\left(\frac{1}{n}L_n^{f/g}(\hat{\alpha}_n,\hat{\beta}_n) - pE\left[\ln\frac{f^{\alpha_*}(X)}{g^{\beta_*}(X)}\right] - p\ln\frac{F^{\alpha_*}(\zeta)}{G^{\beta_*}(\zeta)}\right) \xrightarrow{D} N(0,\ \omega_{*c}^2), \tag{B1}$$

where $\alpha_* = \text{argmax}_{\alpha\in\text{M}}\text{E}_h(L_n^f(\alpha))$ and $\beta_* = \text{argmax}_{\beta\in B}\text{E}_h(L_n^g(\beta))$ are the pseudo-true values of the α. In addition, ω_{*c}^2 is the variance of the difference of log-likelihood functions.

Proof. The proof of this theorem can be obtained by using the multivariate central theorem and by routine calculations. From the Taylor expansion of $L_n^f(\alpha_*)$ around the $\hat{\alpha}_n$, we have

$$L_n^f(\alpha_*) = L_n^f(\hat{\alpha}_n) + \frac{n}{2}\Psi' A_f \Psi + o_p(1),$$

and we can also write

$$L_n^g(\beta_*) = L_n^f(\hat{\beta}_n) + \frac{n}{2}\Omega' A_g \Omega + o_p(1),$$

where $\Psi = (\hat{\alpha}_n - \alpha_*)$, $\Omega = (\hat{\beta}_n - \beta_*)$, and A_f and A_g are the Fisher information matrix.

Since $LR_n(\alpha_*,\beta_*) = L_n^{f/g}(\alpha_*,\beta_*) = L_n^f(\alpha_*) - L_n^g(\beta_*)$, we obtain

$$L_n^{f/g}(\hat{\alpha}_n,\hat{\beta}_n) = L_n^{f/g}(\alpha_*,\beta_*) + \tfrac{n}{2}(\hat{\alpha}_n - \alpha_*)' A_f(\hat{\alpha}_n - \alpha_*) - \tfrac{n}{2}(\hat{\beta}_n - \beta_*)' A_g(\hat{\beta}_n - \beta_*) + o_p(1).$$

Using Theorem A, we observed that the distribution of $\hat{\alpha}_n$, and, similarly $\hat{\beta}_n$, for large n is approximately normal. Thus, we have that $\sqrt{n}(\hat{\alpha}_n - \alpha_*)$ and $\sqrt{n}(\hat{\beta}_n - \beta_*)$ are $O_p(1)$. Therefore, we have

$$\begin{aligned}&\sqrt{n}\Big(\tfrac{1}{n}LR_n(\hat{\alpha}_n,\hat{\beta}_n) - (1-p)E_{h*}\Big[\ln\tfrac{f^{\alpha_*}(X)}{g^{\beta_*}(X)}\Big] - p\ln\tfrac{F^{\alpha_*}(\zeta)}{G^{\beta_*}(\zeta)}\Big)\\ &= \sqrt{n}\Big\{\tfrac{1}{n}L_n^{f/g}(\alpha_*,\beta_*) - (1-p)E_{h*}\Big[\ln\tfrac{f^{\alpha_*}(X)}{g^{\beta_*}(X)}\Big] - p\ln\tfrac{F^{\alpha_*}(\zeta)}{G^{\beta_*}(\zeta)}\Big\} + o_p(1).\end{aligned}$$

From the central limit theorem, the first term on the right-hand side converges in distribution to $N(0, \omega_{*c}^2)$.

References

1. Ahna, S.; Kim Joseph, H.T.; Ramaswami, V. A new class of models for heavy tailed distributions in finance and insurance risk. *Insur. Math. Econ.* **2012**, *51*, 43–52. [CrossRef]
2. Burnecki, K.; Wylomanska, A.; Chechkin, A. Discriminating between light- and heavy-tailed distributions with limit theorem. *PLoS ONE* **2015**, *10*, e0145604. [CrossRef] [PubMed]
3. Hao, X.; Tang, Q. Asymptotic ruin probabilities for a bivariate Leavy-driven risk model with heavy-tailed claims and risky investments. *J. Appl. Probab.* **2012**, *49*, 939–953.
4. Pastor, G.; Mora-Jiménez, I.; Caamaño Antonio, J.; Jäntti, R. Asymptotic expansions for heavy-tailed data. *IEEE Signal Process. Lett.* **2016**, *23*, 444–448. [CrossRef]
5. Barndor-Nielsen, O.E. Superposition of Ornstein-Uhlenbeck type processes. *Theory Probab. Appl.* **2001**, *45*, 175–194. [CrossRef]
6. Chandra, S.R.; Mukherjee, D.; SenGupta, I. PIDE and solution related to pricing of levy driven arithmetic type floating Asian options. *Stoch. Anal. Appl.* **2015**, *33*, 630–652. [CrossRef]
7. Sen Gupta, I. Generalized BN-S stochastic volatility model for option pricing. *Int. J. Theor. Appl. Financ.* **2016**, *19*, 1650014. [CrossRef]
8. Barndor-Nielsen, O.E.; Shephard, N. Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser B* **2001**, *63*, 167–241. [CrossRef]
9. Kundu, D.; Gupta, R.D.; Manglick, A. Discriminating between the log-normal and generalized exponential distribution. *J. Stat. Plan. Inference* **2005**, *127*, 213–227. [CrossRef]
10. Dey, A.K.; Kundu, D. Discriminating among the Log-Normal, Weibull and Generalized Exponential distributions. *IEEE Trans. Reliab.* **2009**, *58*, 416–424. [CrossRef]
11. Cox, D.R. Test of Separate Families of Hypothesis. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Statistical Laboratory of the University of California, Berkeley, CA, USA, 20 June–30 July 1960; pp. 105–123.
12. Vuong, Q.H. Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica* **1989**, *57*, 307–333. [CrossRef]
13. Vuong, Q.H.; Wang, W. Minimum chi-square estimation and tests for model selection. *J. Econom.* **1993**, *56*, 141–168. [CrossRef]
14. Commenges, D.; Liquet, B.; Proust-Lima, C. Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback-Leibler risks. *Biometrics* **2012**, *68*, 380–387. [CrossRef] [PubMed]
15. Panahi, H.; Asadi, S. A model selection test with application to the censored data of carbon nanotubes coating. *Prog. Color Colorants Coat.* **2016**, *9*, 17–28.
16. Panahi, H.; Sayyareh, A. Parameter estimation and prediction of order statistics for the Burr Type XII distribution with Type II censoring. *J. Appl. Stat.* **2014**, *41*, 215–232. [CrossRef]
17. Panahi, H.; Sayyareh, A. Estimation and prediction for a unified hybrid-censored Burr Type XII distribution. *J. Stat. Comput. Simul.* **2016**, *86*, 55–73. [CrossRef]
18. Panahi, H.; Sayyareh, A. Tracking interval for type II hybrid censoring scheme. *JIRSS* **2014**, *13*, 187–208.
19. Cain, K.C.; Harlow, S.D.; Little, R.J.; Nan, B.; Yosef, M.; Taffe, J.R.; Elliott, M.R. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *Am. J. Epidemiol.* **2011**, *25*, 1–7. [CrossRef] [PubMed]
20. Mitra, S.; Kundu, D. Analysis of left censored data from the generalized exponential distribution. *J. Stat. Comput. Simul.* **2008**, *78*, 669–679. [CrossRef]
21. Singh, U.; Kumar, A. Bayesian estimation of the exponential parameter under a multiply type-II Censoring scheme. *Aust. J. Stat.* **2007**, *36*, 227–238.
22. Thompson, E.M.; Hewlett, J.B.; Baise, L.G.; Voge, R.M. The Gumbel hypothesis test for left censored observations using regional earthquake records as an example. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 115–126. [CrossRef]
23. Louis, T.A. Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc. Ser. B* **1982**, *44*, 226–233.

24. Burr, I.W. Cumulative frequency functions. *Ann. Math. Stat.* **1942**, *13*, 215–232. [CrossRef]
25. Panahi, H.; Sayyareh, A. Tracking interval for doubly censored data with application of plasma droplet spread samples. *J. Stat. Res. Iran* **2015**, *11*, 147–176. [CrossRef]
26. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.