AIMS Mathematics, 8(8): 19477–19503.
DOI: 10.3934/math.2023994
Received: 27 March 2023
Revised: 16 May 2023
Accepted: 22 May 2023
Published: 09 June 2023

http://www.aimspress.com/journal/Math

Research article

Univariate and multivariate analyses of the asset returns using new statistical models and penalized regression techniques

Huda M. Alshanbari[1], Zubair Ahmad[2,*], Faridoon Khan[3], Saima K. Khosa[4], Muhammad Ilyas[5] and Abd Al-Aziz Hosni El-Bagoury[6]

[1] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[2] Department of Statistics, Quaid-i-Azam University, Islamabad 44000, Pakistan
[3] Pakistan Institute of Development Economics, Islamabad 44000, Pakistan
[4] Department of Mathematics and Statistics University of Saskatchewan, Saskatoon, SK, Canada
[5] Department of Statistics, University of Malakand, Dir (L), Chakdara, Khyber Pakhtunkhwa, Pakistan
[6] Higher Institute of Engineering and Technology at El-Mahala El-Kobra, Egypt

* **Correspondence:** Email: z.ferry21@gmail.com.

Abstract: The COVID-19 epidemic has had a profound effect on almost every aspect of daily life, including the financial sector, education, transportation, health care, and so on. Among these sectors, the financial and health sectors are the most affected areas by COVID-19. Modeling and predicting the impact of the COVID-19 epidemic on the financial and health care sectors is particularly important these days. Therefore, this paper has two aims, (i) to introduce a new probability distribution for modeling the financial data set (oil prices data), and (ii) to implement a machine learning approach to predict the oil prices. First, we introduce a new approach for developing new probability distributions for the univariate analysis of the oil price data. The proposed approach is called a new reduced exponential-X (NRE-X) family. Based on this approach, two new statistical distributions are introduced for modeling the oil price data and its log returns. Based on certain statistical tools, we observe that the proposed probability distributions are the best competitors for modeling the prices' data sets. Second, we carry out a multivariate analysis while considering some covariates of oil price data. Dual well-known machine learning algorithms, namely, the least absolute shrinkage and absolute deviation (Lasso) and Elastic net (Enet) are utilized to achieve the important features for oil prices based on the best model. The best model is established through forecasting performance.

Keywords: logistic distribution; heavy-tailed distributions; T-X family; cryptocurrency; artificial neural networks; support vector regression
Mathematics Subject Classification: 62F09, 62G34

1. Introduction

The worldwide COVID-19 outbreak has affected the production and lives of society, in addition to posing a serious threat to human health [1]. Governments have implemented countermeasures such as, lock-downs, to circumvent the transmission of the disease as the number of confirmed cases continues to rise. Global supply networks were disrupted as a result of these actions, and the world experienced an economic blockade [2]. It resulted in the collapse of the stock market crisis, the energy market, and a downturn in the world economy [3, 4]. One of the financial markets most influenced by COVID-19 is WTI (West Texas Intermediate) crude oil futures [5]. WTI crude oil futures fall to a negative value on April 20, 2020, from $18 per barrel to -$38, experienced for the first time in history [6].

The global market economy was seriously affected by the drop in crude oil futures prices. Coal futures declined from 70.68 (on March 27) to $51.62 (on April 27), indicating that they are alternatives to each other. Since then, the market has descended into depression. The US government updated a fresh round of energy perspectives in order to stabilize the market economy and mitigate the influence of the crises. However, since June, the coal prices and WTI crude oil have risen steadily. Inflationary pressures have also sparked worries about the energy market's future growth.

From Figures 1 and 2, it can be observed that the series of oil prices and oil returns are highly volatile and noisy. The existing statistical tools are not capable to fit both series of oil efficiently, and thereby this study modifies the statistical tool in order to approximate the oil prices and their return accurately. In addition, our study adopts dual penalization techniques, namely the Least absolute shrinkage and selection operator (Lasso) and elastic net (Enet) to select the important features that drive the oil prices in the case of China.

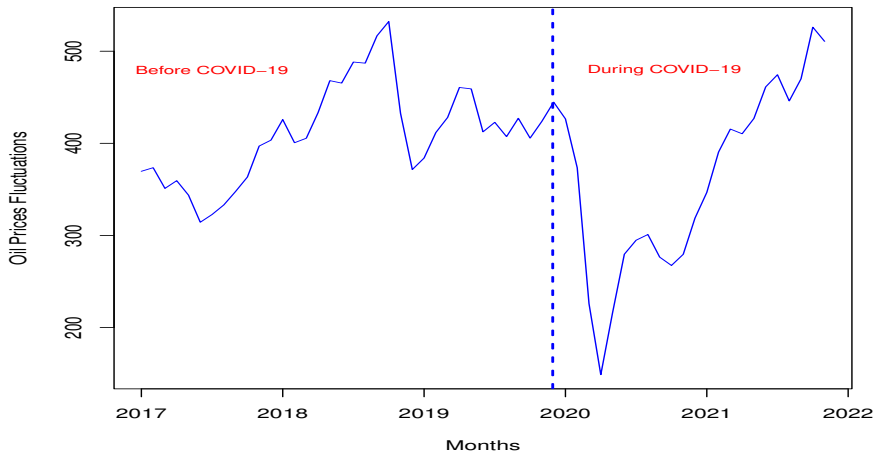

Figure 1. Before and during COVID-19 oil prices trend in China.

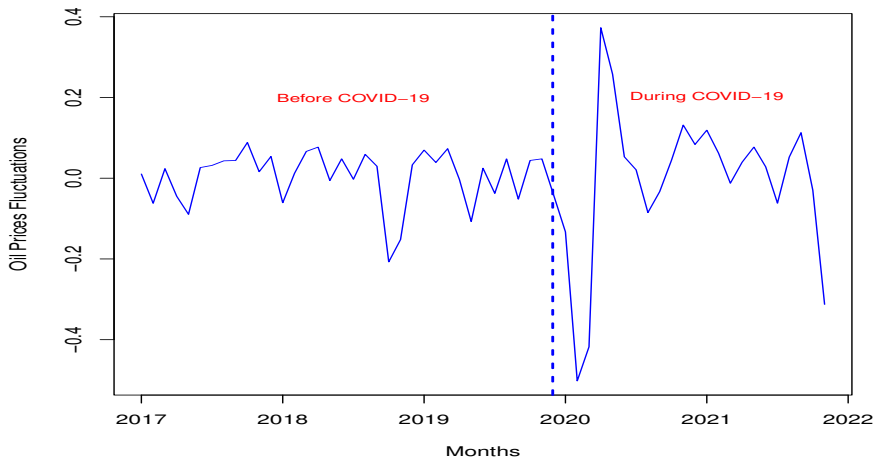

Figure 2. Before and during COVID-19 oil return trend in China.

In the financial and other closed connected sectors, statistical distributions play a vital role to provide the best description of the data under consideration. The next section offers a review of probability models that are implemented for analyzing data sets in different sectors.

2. Literature review

Probability/statistical models play a vital role in almost every sector. The applications of statistical models can be found (*i*) in financial sectors [7, 8], (*ii*) on social media platforms [9], (*iii*) oceanology and metrology [10], (*iv*) engineering-related areas [11], (*v*) medicine and healthcare-related sectors [12], and (*vi*) hydrology [13], among others.

Among these sectors, the statistical models play an important role in data modeling in different sectors related to the financial phenomena. For example, (*i*) Bhati and Ravi [14] introduced a log-Moyal distribution for modeling the financial and insurance data sets, (*ii*) Ahmad et al. [15] implemented a heavy-tailed extension of the Weibull model for analyzing the vehicle insurance loss data, (*iii*) Li et al. [16] proposed a generalized log-Moyal distribution for modeling the loss data, (*vi*) Bielak et al. [17] implemented a multidimensional and heavy-tailed-based model for market risk factors analysis. For furthermore details, we refer the interested readers to [18–24].

The development of new statistical distributions for modeling data in the above-mentioned sectors is a prominent research area. Therefore, numerous new statistical distributions have been introduced for providing the best description of the metrology, oceanology, engineering, hydrology, medical, and financial data sets [25]. In this paper, we also introduce a novel statistical approach for modeling data in the financial sector. The proposed approach is called, a new reduced exponential-X (NRE-X) family of distributions.

Genesis: Suppose a random variable, say T, with probability density function (PDF) $f(t)$, where $T \in [r_1, r_2]$ for $-\infty \leq r_1 < r_2 \leq \infty$. Let X is another random variable with CDF (cumulative distribution function) represented by $K(x)$. Let $L[K(x)]$ represent a function of CDF of X, satisfying the following three conditions

(*i*) $L[K(x)] \in [r_1, r_2]$,
(*ii*) $L[K(x)]$ is differentiable and monotonically increasing,
(*iii*) $L[K(x)] \to r_1$ as $x \to -\infty$ and $L[K(x)] \to r_2$ as $x \to \infty$.

According to the findings of [26], the CDF of the T-X distributions approach is defined by

$$M(x) = \int_{r_1}^{L[K(x)]} f(t)\, dt, \qquad x \in \mathbb{R}. \tag{2.1}$$

Corresponding to $M(x)$, the PDF $m(x)$ of the T-X distributions is given by

$$m(x) = \left\{\frac{d}{dx} L[K(x)]\right\} k\{L[K(x)]\}, \qquad x \in \mathbb{R}.$$

Consider the CDF $F(t)$ of the exponential distribution given by

$$F(t) = 1 - e^{-\phi t}, \qquad t \geq 0, \phi \in \mathbb{R}^+. \tag{2.2}$$

Let $\phi = 1$, then, Eq (2.2) becomes

$$F(t) = 1 - e^{-t}, \qquad t \geq 0. \tag{2.3}$$

Corresponding to $F(t)$, the PDF $f(t)$ is given by

$$f(t) = e^{-t}, \qquad t > 0. \tag{2.4}$$

Using Eq (2.4) and $L[K(x)] = \frac{\lambda^2[1-K(x)]}{\left(\lambda - \log[1-K(x)]\right)^2}$ in Eq (2.1), we get the CDF of the proposed the NRE-X family, given by

$$M(x;\lambda) = 1 - \frac{\lambda^2 \bar{K}(x)}{\left(\lambda - \log\left[\bar{K}(x)\right]\right)^2}, \qquad x \in \mathbb{R}, \lambda \in \mathbb{R}^+, \tag{2.5}$$

where, $\bar{K}(x) = 1 - K(x)$ is the SF (survival function) of the baseline random variable.

Based on our search of the existing literature, the NRE-X approach defined in Eq (2.5) has not been used/proposed so far. This is one of the key motivations of the current work. Therefore, using the NRE-X approach numerous new models can also be introduced for data modeling in different sectors. Corresponding to $M(x;\lambda)$, the PDF $m(x;\lambda)$ is given by

$$m(x;\lambda) = \frac{\lambda^2 k(x)}{\left(\lambda - \log\left[\bar{K}(x)\right]\right)^3}\left(2 + \lambda - \log\left[\bar{K}(x)\right]\right), \quad x \in \mathbb{R}, \tag{2.6}$$

where $\frac{d}{dx}K(x) = k(x)$.

In link to Eqs (2.5) and (2.6), the SF $\bar{M}(x;\lambda) = 1 - M(x;\lambda)$ and hazard function (HF) $h(x;\lambda) = \frac{m(x;\lambda)}{\bar{M}(x;\lambda)}$, are given, respectively, by

$$\bar{M}(x;\lambda) = \frac{\lambda^2 \bar{K}(x)}{\left(\lambda - \log\left[\bar{K}(x)\right]\right)^2}, \qquad x \in \mathbb{R},$$

and

$$h(x;\lambda) = \frac{k(x)}{\bar{K}(x)\left(\lambda - \log\left[\bar{K}(x)\right]\right)}\left(2 + \lambda - \log\left[\bar{K}(x)\right]\right), \quad x \in \mathbb{R}.$$

In this paper, we implement the proposed NRE-X distributions and introduce updated versions of the logistic and Fréchet distributions. The updated version of the logistic distribution is called, a new reduced exponential-logistic (NRE-logistic) model. Whereas, the updated version of the Fréchet model is called, a new reduced exponential-Fréchet (NRE-Fréchet) model. In the next section, we provide expressions for the CDF, PDF, SF, and HF of the NRE-logistic and NRE-Fréchet distributions. Furthermore, different PDF behaviors of the NRE-logistic and NRE-Fréchet distributions are also presented.

3. Special models of the NRE-X method

This section is devoted to introducing two special models of the NRE-X method by taking the logistic and Fréchet distributions as baseline models.

3.1. A NRE-logistic distribution

Consider the CDF $K(x;\theta,\sigma)$ of the two parameters ($\theta \in \mathbb{R}, \sigma \in \mathbb{R}^+$) logistic distribution is given by

$$K(x;\theta,\sigma) = \left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1} \qquad x \in \mathbb{R}. \tag{3.1}$$

For $\theta \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$, the PDF, SF and HF of the logistic distribution are given by

$$k(x;\theta,\sigma) = \frac{1}{\sigma}e^{-\left(\frac{x-\theta}{\sigma}\right)}\left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-2}, \qquad x \in \mathbb{R},$$

$$\bar{K}(x;\theta,\sigma) = 1 - \left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}, \qquad x \in \mathbb{R},$$

and

$$h(x;\theta,\sigma) = \frac{e^{-\left(\frac{x-\theta}{\sigma}\right)}\left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-2}}{1 - \left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}}, \qquad x \in \mathbb{R},$$

respectively.

Using Eq (3.1) in Eq (2.5), we get the CDF of the NRE-logistic distribution, given by

$$M(x;\lambda,\theta,\sigma) = 1 - \frac{\lambda^2\left(1 - \left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right)}{\left(\lambda - \log\left[1 - \left(1 + e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]\right)^2}, \qquad x \in \mathbb{R}, \tag{3.2}$$

with PDF

$$m(x;\lambda,\theta,\sigma)=\frac{\frac{\lambda^2}{\sigma}e^{-\left(\frac{x-\theta}{\sigma}\right)}\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-2}}{\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]\right)^3}\left(2+\lambda-\log\left[1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]\right),\qquad x\in\mathbb{R}. \quad (3.3)$$

Some possible plots for the PDF of the NRE-logistic distribution are presented in Figure 3. In the plots in Figure 3, the additional parameter λ has a positive impact on the shape of the PDF of the NRE-logistic distribution. These plots reveal that as the value of λ increases, the NRE-logistic distribution possesses the heavy-tailed characteristics which are very useful in the financial and other related sectors.

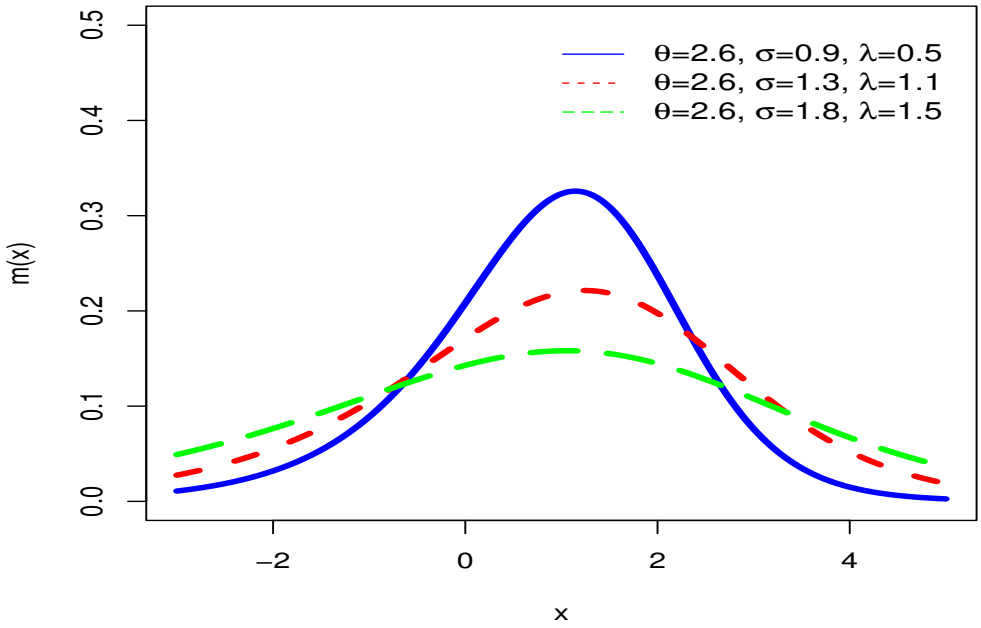

Figure 3. Plots for the PDF of the NRE-logistic distribution for a fixed value of σ and different values of θ and λ.

Furthermore, the SF and HF of the NRE-logistic distribution are, respectively, given by

$$\bar{M}(x;\lambda,\theta,\sigma)=\frac{\lambda^2\left(1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right)}{\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]\right)^2},$$

and

$$h(x;\lambda,\theta,\sigma)=\frac{\frac{1}{\sigma}e^{-\left(\frac{x-\theta}{\sigma}\right)}\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-2}\left(2+\lambda-\log\left[1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]\right)}{\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]\right)\left(1-\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right)}.$$

3.2. A NRE-Fréchet distribution

Let $K(x;\eta,\varphi)$ be the CDF of the two parameters ($\eta \in \mathbb{R}^+, \varphi \in \mathbb{R}^+$) Fréchet distribution, given by

$$K(x;\eta,\varphi) = e^{-\varphi x^{-\eta}}, \qquad x \in \mathbb{R}^+. \tag{3.4}$$

For $\eta \in \mathbb{R}^+$ and $\varphi \in \mathbb{R}^+$, the PDF, SF, and HF of the Fréchet distribution are given by

$$k(x;\eta,\varphi) = \eta\varphi x^{-\eta-1} e^{-\varphi x^{-\eta}}, \qquad x \in \mathbb{R}^+,$$

$$\bar{K}(x;\eta,\varphi) = 1 - e^{-\varphi x^{-\eta}}, \qquad x \in \mathbb{R}^+,$$

and

$$h(x;\eta,\varphi) = \frac{\eta\varphi x^{-\eta-1} e^{-\varphi x^{-\eta}}}{1 - e^{-\varphi x^{-\eta}}}, \qquad x \in \mathbb{R}^+,$$

respectively.

Using Eq (3.4) in Eq (2.5), we get the CDF of the NRE-Fréchet distribution, given by

$$M(x;\lambda,\eta,\varphi) = 1 - \frac{\lambda^2\left(1 - e^{-\varphi x^{-\eta}}\right)}{\left[\lambda - \log\left(1 - e^{-\varphi x^{-\eta}}\right)\right]^2}, \qquad x \in \mathbb{R}^+, \lambda \in \mathbb{R}^+, \tag{3.5}$$

with PDF

$$m(x;\lambda,\eta,\varphi) = \frac{\lambda^2\eta\varphi x^{-\eta-1} e^{-\varphi x^{-\eta}}}{\left[\lambda - \log\left(1 - e^{-\varphi x^{-\eta}}\right)\right]^3}\left[2 + \lambda - \log\left(1 - e^{-\varphi x^{-\eta}}\right)\right], \ x \in \mathbb{R}^+. \tag{3.6}$$

Some possible behaviors for the PDF of the NRE-Fréchet distribution are presented in Figure 4. The plots in Figure 4, show that the additional parameter λ, has a significant impact on the PDF shape of the NRE-Fréchet distribution. In a nutt shell, we can see that as the value of λ increases, the NRE-Fréchet distribution tends to be a heavy-tailed distribution.

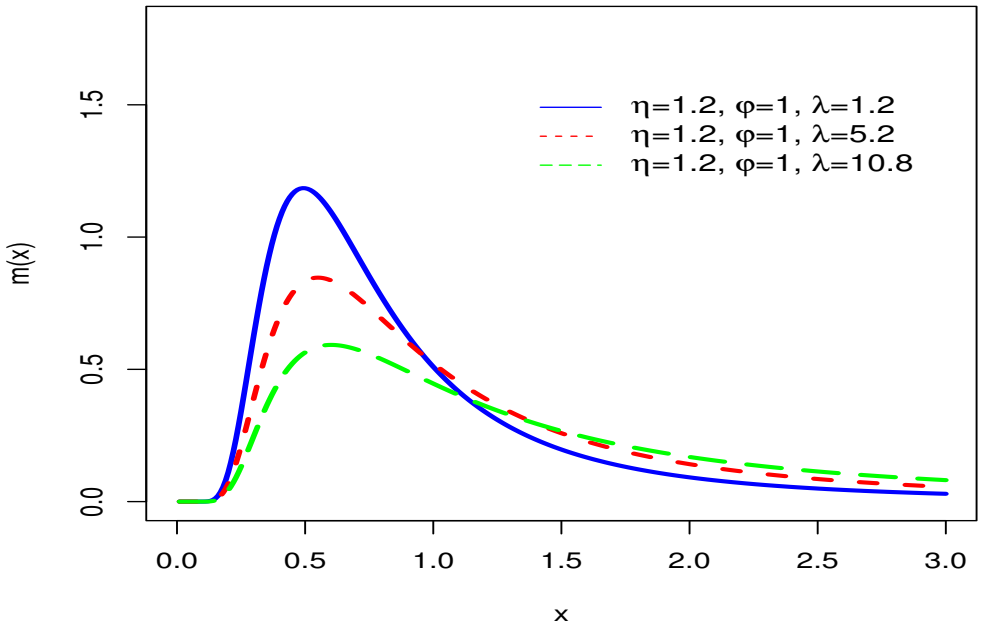

Figure 4. Plots for the PDF of the NRE-Fréchet distribution for fixed values of (η, φ) and different values of λ.

Furthermore, the SF and HF of the NRE-Fréchet distribution are given by

$$\bar{M}(x;\lambda,\eta,\varphi) = \frac{\lambda^2\left(1-e^{-\varphi x^{-\eta}}\right)}{\left[\lambda-\log\left(1-e^{-\varphi x^{-\eta}}\right)\right]^2}, \qquad x\in\mathbb{R}, \lambda\in\mathbb{R}^+,$$

and

$$h(x;\lambda,\eta,\varphi) = \frac{\eta\varphi x^{-\eta-1}e^{-\varphi x^{-\eta}}\left[2+\lambda-\log\left(1-e^{-\varphi x^{-\eta}}\right)\right]}{\left[\lambda-\log\left(1-e^{-\varphi x^{-\eta}}\right)\right]\left(1-e^{-\varphi x^{-\eta}}\right)}, \quad x\in\mathbb{R}^+,$$

respectively.

4. Estimation and simulation of the NRE-logistic and NRE-Fréchet distributions

In this section, we use the maximum likelihood estimation method to obtain the maximum likelihood estimators (MLEs) for the parameters of the NRE-logistic and NRE-Fréchet distributions. Furthermore, simulation studies are also provided to assess the performances of the MLEs of the NRE-logistic and NRE-Fréchet distributions.

4.1. Estimation and simulation of the NRE-logistic distribution

Assume a set of random samples, say $X_1, X_2, ..., X_n$, taken from the NRE-logistic distribution with PDF $m(x;\lambda,\theta,\sigma)$ in Eq (3.3). Then, corresponding to $m(x;\lambda,\theta,\sigma)$, the log-likelihood function (LLF),

say $\ell(\lambda,\theta,\sigma)$, is given by

$$\ell(\lambda,\theta,\sigma) = 2n\log\lambda - n\log\sigma - \sum_{i=1}^{n}\left(\frac{x_i-\theta}{\sigma}\right) - 2\sum_{i=1}^{n}\log\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right) + \sum_{i=1}^{n}\log\left(2+\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right) + 3\sum_{i=1}^{n}\log\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right).$$

Corresponding to $\ell(\lambda,\theta,\sigma)$, the partial derivatives with respect to θ,σ, and λ are given, respectively, by

$$\frac{\partial}{\partial\theta}\ell(\lambda,\theta,\sigma) = +\frac{n}{\sigma} - \frac{2n}{\sigma}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i-\theta}{\sigma}\right)}}{\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)} - \sum_{i=1}^{n}\frac{\frac{n}{\sigma}e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-2}}{\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\left(2+\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right)} - \frac{3n}{\sigma}\sum_{i=1}^{n}\frac{e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-2}}{\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right)},$$

$$\frac{\partial}{\partial\sigma}\ell(\lambda,\theta,\sigma) = -\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{x_i}{\sigma^2} - 2\sum_{i=1}^{n}\frac{\frac{x_i}{\sigma^2}e^{-\left(\frac{x_i-\theta}{\sigma}\right)}}{\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)} - \sum_{i=1}^{n}\frac{\frac{x_i}{\sigma^2}e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-2}}{\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\left(2+\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right)} - 3\sum_{i=1}^{n}\frac{\frac{x_i}{\sigma^2}e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-2}}{\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right)}$$

and

$$\frac{\partial}{\partial\lambda}\ell(\lambda,\theta,\sigma) = \frac{2n}{\lambda} + \sum_{i=1}^{n}\frac{1}{\left(2+\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right)} + 3\sum_{i=1}^{n}\frac{1}{\left(\lambda-\log\left[1-\left(1+e^{-\left(\frac{x_i-\theta}{\sigma}\right)}\right)^{-1}\right]\right)}.$$

On solving $\frac{\partial}{\partial\theta}\ell(\lambda,\theta,\sigma) = 0$, $\frac{\partial}{\partial\sigma}\ell(\lambda,\theta,\sigma) = 0$, and $\frac{\partial}{\partial\lambda}\ell(\lambda,\theta,\sigma) = 0$, we get the MLEs $\left(\hat{\theta}_{MLE},\hat{\sigma}_{MLE},\hat{\lambda}_{MLE}\right)$ of the parameters (θ,σ,λ) of the of the NRE-logistic distribution.

Next, we provide a Monte Carlo simulation study (MCSS) to evaluate the performances of $\hat{\theta}_{MLE},\hat{\sigma}_{MLE}$, and $\hat{\lambda}_{MLE}$. For the NRE-logistic distribution, the MCSS is conducted by generating the random numbers such as $n = 25, 50, ..., 500$ using the inverse CDF method. The MCSS s conducted for two different combinations of θ,σ, and λ. These combination values are given by (i) $\theta = 1.5, \sigma = 0.9, \lambda = 0.5$ and (ii) $\theta = 1.3, \sigma = 0.7, \lambda = 1.1$.

The performances of $\hat{\theta}_{MLE},\hat{\sigma}_{MLE}$, and $\hat{\lambda}_{MLE}$ are evaluated using two decisive tools. The numerical values of the selected decisive tools are computed as

- Bias

$$Bias\left(\hat{\theta}_{MLE}\right) = \sum_{i=1}^{n}\left(\hat{\theta}_i - \theta\right).$$

- Mean square error (MSE)

$$MSE\left(\hat{\theta}_{MLE}\right) = \sum_{i=1}^{n}\left(\hat{\theta}_i - \theta\right)^2.$$

The MLEs and the above evaluation criteria are computed using the optim()R with argument method = "L-BFGS-B". After carrying out the MCSS, the numerical results are provided in Tables 1 and 2. Whereas, the results of the MCSS are illustrated visually in Figures 5 and 6.

The results of the MCSS of the NRE-logistic distribution presented in Tables 1 and 2, Figures 5 and 6, reveal that as the size of the sample n increases

- The MLEs $\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}$, and $\hat{\lambda}_{MLE}$ tend to stable.
- The MSEs of $\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}$, and $\hat{\lambda}_{MLE}$ decreases.
- The biases of $\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}$, and $\hat{\lambda}_{MLE}$ decay to zero.

Table 1. The results for the MCSS of the NRE-logistic for $\theta = 1.5, \sigma = 0.9, \lambda = 0.5$.

n	Parameters	MLEs	MSEs	Biases
25	θ	1.5306540	0.0845499	0.03065429
	σ	1.6616360	1.6801157	0.76163600
	λ	2.1413169	6.5317397	1.64131690
50	θ	1.4878190	0.0405659	-0.01218063
	σ	1.3621325	0.8178465	0.46213251
	λ	1.7129257	4.8094673	1.21292570
75	θ	1.4833020	0.0266625	-0.01669770
	σ	1.2395189	0.5670010	0.33951893
	λ	1.3942232	3.2730729	0.89422320
100	θ	1.4907440	0.0200039	-0.00925587
	σ	1.1644847	0.4874031	0.26448468
	λ	1.2066578	2.6246481	0.70665780
200	θ	1.4939300	0.0128720	-0.00606980
	σ	1.0687798	0.3153811	0.16877977
	λ	0.9406648	1.4242122	0.44066480
300	θ	1.4850080	0.0079217	-0.01499184
	σ	1.0421690	0.2105968	0.14216899
	λ	0.7456199	0.7993679	0.30561990
400	θ	1.4895150	0.0067249	-0.01048548
	σ	1.0060013	0.1530510	0.10600131
	λ	0.5966013	0.3541581	0.19660130
500	θ	1.4963270	0.0048694	-0.00367308
	σ	0.9351764	0.1209646	0.07517640
	λ	0.5406189	0.2645338	0.14061890

Table 2. The results for the MCSS of the NRE-logistic for $\theta = 1.3, \sigma = 0.7, \lambda = 1.1$.

n	Parameters	MLEs	MSEs	Biases
25	θ	1.3409810	0.06089320	4.0981e-02
	σ	0.9684732	0.32269754	0.26847323
	λ	2.4804880	6.35740400	1.58048760
50	θ	1.3207820	0.03098274	2.0781e-02
	σ	0.8257357	0.18517215	0.12573567
	λ	2.0018830	4.78013720	1.10188330
75	θ	1.3131030	0.02263240	1.3103e-02
	σ	0.8056960	0.16232288	0.10569596
	λ	1.8454520	3.95911760	0.94545150
100	θ	1.2969670	0.01678957	-3.0325e-03
	σ	0.8074029	0.13781325	0.10740290
	λ	1.8371310	3.76135120	0.93713060
200	θ	1.3000190	0.01000074	1.9164e-05
	σ	0.7502970	0.09277251	0.05029699
	λ	1.4371740	2.04954990	0.53717410
300	θ	1.2876490	0.00735997	-1.2350e-02
	σ	0.7879704	0.07972821	0.08797043
	λ	1.4682940	1.76350150	0.56829450
400	θ	1.2992960	0.00527875	-7.0367e-04
	σ	0.7273118	0.05311106	0.02731182
	λ	1.1906500	0.97227000	0.29065020
500	θ	1.3000400	0.00404874	3.9596e-05
	σ	0.7113346	0.04542952	0.01133456
	λ	1.1060500	0.72525460	0.20605010

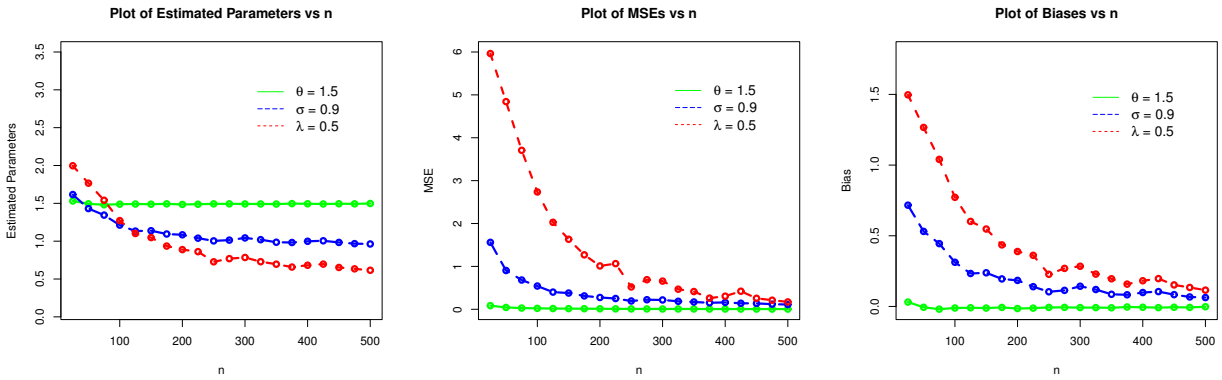

Figure 5. The visual illustration of the simulation results of the NRE-logistic for $\theta = 1.5, \sigma = 0.9, \lambda = 0.5$.

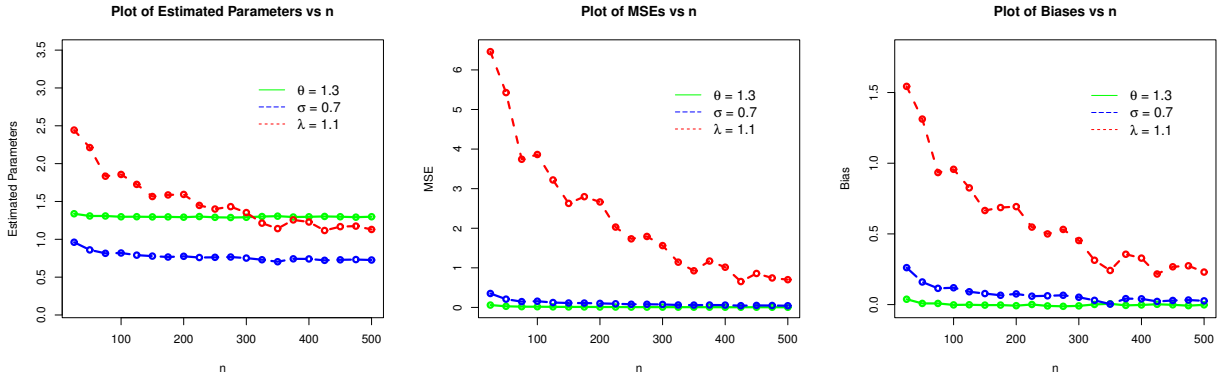

Figure 6. The visual illustration of the simulation results of the NRE-logistic for $\theta = 1.3, \sigma = 0.7, \lambda = 1.1$.

4.2. Estimation and simulation of the NRE-Fréchet distribution

Consider a random sample, say $X_1, X_2, ..., X_n$, observed from the NRE-Fréchet distribution with PDF $m(x; \lambda, \eta, \varphi)$ expressed by Eq (3.6). Then, corresponding to $m(x; \lambda, \eta, \varphi)$, the LLF, say $\ell(\lambda, \eta, \varphi)$, is given by

$$m(x; \lambda, \eta, \varphi) = \frac{\lambda^2 \eta \varphi x^{-\eta-1} e^{-\varphi x^{-\eta}}}{\left[\lambda - \log\left(1 - e^{-\varphi x^{-\eta}}\right)\right]^3} \left[2 + \lambda - \log\left(1 - e^{-\varphi x^{-\eta}}\right)\right],$$

$$\ell(\lambda, \eta, \varphi) = 2n \log \lambda + n \log \eta + n \log \varphi - (\eta + 1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \varphi x_i^{-\eta} + \sum_{i=1}^{n} \log\left[2 + \lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]$$
$$- 3 \sum_{i=1}^{n} \log\left[\lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right].$$

Corresponding to $\ell(\lambda, \eta, \varphi)$, the partial derivatives with respect to η, φ, and λ are given, respectively, by

$$\frac{\partial}{\partial \eta} \ell(\lambda, \eta, \varphi) = \frac{n}{\eta} - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} (\log x_i) \varphi x_i^{-\eta} + \sum_{i=1}^{n} \frac{(\log x_i) \varphi x_i^{-\eta} e^{-\varphi x_i^{-\eta}}}{\left(1 - e^{-\varphi x_i^{-\eta}}\right)\left[2 + \lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]}$$
$$- 3 \sum_{i=1}^{n} \frac{(\log x_i) \varphi x_i^{-\eta} e^{-\varphi x_i^{-\eta}}}{\left(1 - e^{-\varphi x_i^{-\eta}}\right)\left[\lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]},$$

$$\frac{\partial}{\partial \varphi} \ell(\lambda, \eta, \varphi) = \frac{n}{\varphi} - \sum_{i=1}^{n} x_i^{-\eta} - \sum_{i=1}^{n} \frac{x_i^{-\eta} e^{-\varphi x_i^{-\eta}}}{\left(1 - e^{-\varphi x_i^{-\eta}}\right)\left[2 + \lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]} + 3 \sum_{i=1}^{n} \frac{x_i^{-\eta} e^{-\varphi x_i^{-\eta}}}{\left(1 - e^{-\varphi x_i^{-\eta}}\right)\left[\lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]},$$

and

$$\frac{\partial}{\partial \lambda} \ell(\lambda, \eta, \varphi) = \frac{2n}{\lambda} + \sum_{i=1}^{n} \frac{1}{\left(1 - e^{-\varphi x_i^{-\eta}}\right)\left[2 + \lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]} - 3 \sum_{i=1}^{n} \frac{1}{\left(1 - e^{-\varphi x_i^{-\eta}}\right)\left[\lambda - \log\left(1 - e^{-\varphi x_i^{-\eta}}\right)\right]}.$$

Equating the expressions $\frac{\partial}{\partial\eta}\ell(\lambda,\eta,\varphi)$, $\frac{\partial}{\partial\varphi}\ell(\lambda,\eta,\varphi)$, and $\frac{\partial}{\partial\lambda}\ell(\lambda,\eta,\varphi)$ to zero and solving we get the MLEs $\left(\hat{\eta}_{MLE},\hat{\varphi}_{MLE},\hat{\lambda}_{MLE}\right)$ of the parameters (η,φ,λ) of the of the NRE-Fréchet distribution.

After obtaining the MLEs $\hat{\eta}_{MLE},\hat{\varphi}_{MLE}$, and $\hat{\lambda}_{MLE}$, now we check their performances through a Monte Carlo simulation study (MCSS). The evaluation of $\hat{\eta}_{MLE},\hat{\varphi}_{MLE}$, and $\hat{\lambda}_{MLE}$ is carried out by generating random numbers from the NRE-Fréchet distribution by implementing the NRE-Fréchet distribution. For the NRE-Fréchet distribution, the MCSS is conducted for two sets of parameters, such as (i) $\theta = 1.5,\sigma = 0.9,\lambda = 0.5$ and (ii) $\theta = 0.9,\sigma = 1.1,\lambda = 1.5$.

Again, we use the Bias and MSE as decisive tools to judge the performances of $\hat{\eta}_{MLE},\hat{\varphi}_{MLE}$, and $\hat{\lambda}_{MLE}$. After conducting the MCSS, the numerical results for the MLEs, MSEs, and biases of the NRE-Fréchet distribution are presented in Tables 3 and 4. Whereas, the results of the MCSS are illustrated visually in Figures 7 and 8.

From the results of the MCSS of the NRE-Fréchet distribution provided in Tables 3 and 4, Figures 7 and 8, we can see that the MLEs of the NRE-Fréchet distribution tend to sable as the size of the sample increases. With the increase in the sample size, the MSEs of the $\hat{\eta}_{MLE},\hat{\varphi}_{MLE}$, and $\hat{\lambda}_{MLE}$ decreases and biases of $\hat{\eta}_{MLE},\hat{\varphi}_{MLE}$, and $\hat{\lambda}_{MLE}$ decay to zero.

Table 3. The results for the MCSS of the NRE-Fréchet for $\eta = 0.7,\varphi = 1.6,\lambda = 1.3$.

n	Parameters	MLEs	MSEs	Biases
25	η	0.8295898	0.08826282	0.12958978
	ϕ	1.6435660	0.20131711	0.04356598
	λ	1.8830760	1.75815404	0.58307614
500	η	0.7636202	0.03381550	0.06362024
	ϕ	1.6144310	0.11071804	0.01443136
	λ	1.5688950	0.60710519	0.26889488
75	η	0.7413941	0.02058671	0.04139414
	ϕ	1.6032170	0.07300298	0.00321683
	λ	1.4934430	0.38750257	0.19344311
100	η	0.7308403	0.01473447	0.03084032
	ϕ	1.6124770	0.05724885	0.01247658
	λ	1.4101360	0.15857094	0.11013555
200	η	0.7159275	0.00628118	0.01592747
	ϕ	1.5983640	0.02611875	-0.00163616
	λ	1.3440490	0.02956159	0.04404876
300	η	0.7083867	0.00356276	0.00838667
	ϕ	1.6127650	0.01877270	0.01276520
	λ	1.3214120	0.01317318	0.02141173
400	η	0.7037929	0.00282884	0.00379289
	ϕ	1.6012280	0.01354377	0.00122802
	λ	1.3194250	0.01055359	0.01942524
500	η	0.7062376	0.00234377	0.00623763
	ϕ	1.5949220	0.01018247	-0.00507846
	λ	1.3100780	0.00762077	0.01907786

Table 4. The results for the MCSS of the NRE-Fréchet for $\eta = 0.9, \varphi = 1.1, \lambda = 1.5$.

n	Parameters	MLEs	MSEs	Biases
25	η	1.0040981	0.083711686	0.10409810
	ϕ	1.2140710	0.098602278	0.11407104
	λ	1.8282900	0.975918910	0.3282901
500	η	0.9628862	0.037615081	0.06288619
	ϕ	1.1532450	0.036819558	0.05324535
	λ	1.6785700	0.420855040	0.17856980
75	η	0.9416555	0.026231108	0.04165548
	ϕ	1.1467750	0.027392599	0.04677475
	λ	1.5794770	0.139978210	0.07947676
100	η	0.9243457	0.015663488	0.02434573
	ϕ	1.1333150	0.021317467	0.03331549
	λ	1.5717710	0.105736930	0.07177117
200	η	0.9111239	0.007281454	0.01112393
	ϕ	1.1173830	0.012048905	0.01738310
	λ	1.5347830	0.045606720	0.03478277
300	η	0.9082706	0.005246932	0.00827062
	ϕ	1.1059050	0.007748683	0.00590497
	λ	1.5355600	0.032318420	0.03556033
400	η	0.9051241	0.003407605	0.00512410
	ϕ	1.1116230	0.006835205	0.01162262
	λ	1.5194740	0.024970670	0.01947445
500	η	0.9068855	0.003215347	0.00688549
	ϕ	1.1049340	0.005313549	0.00493448
	λ	1.5219570	0.020792440	0.02195670

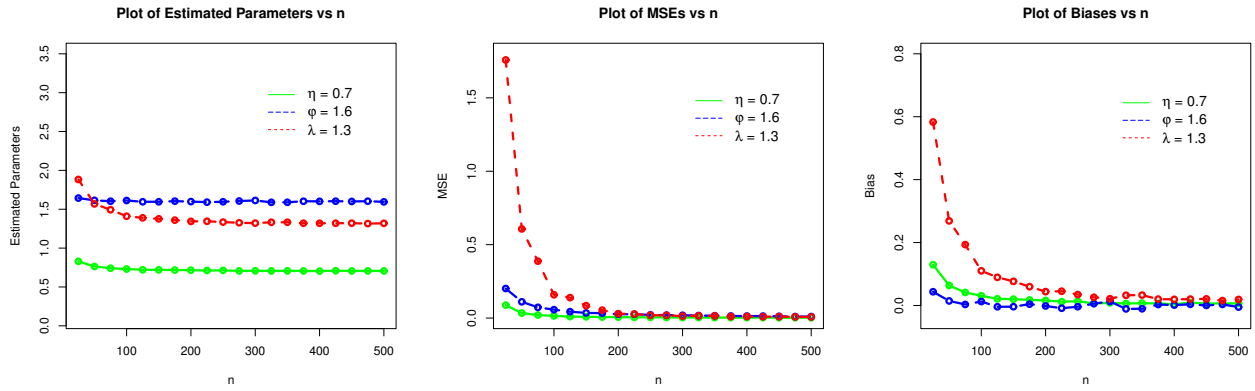

Figure 7. The visual illustration of the simulation results of the NRE-Fréchet for $\eta = 0.7, \varphi = 1.6, \lambda = 1.3$.

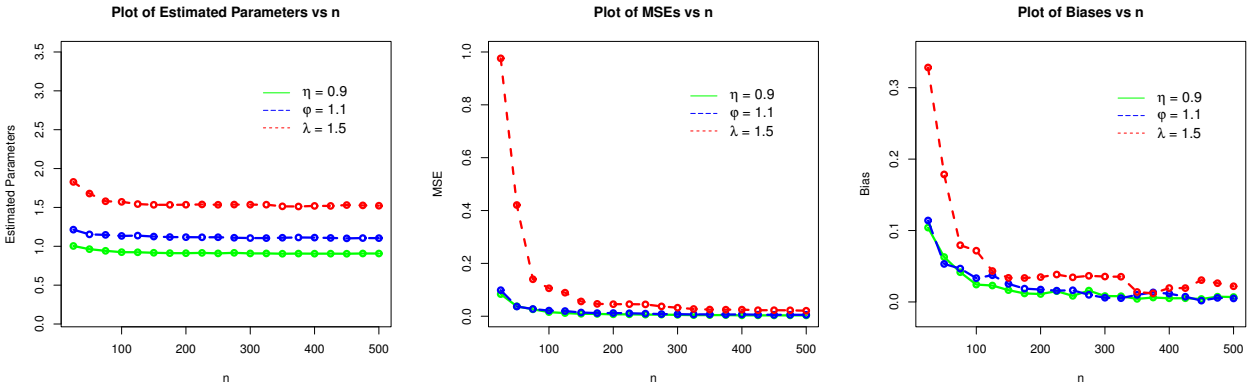

Figure 8. The visual illustration of the simulation results of the NRE-Fréchet distribution for $\eta = 0.9, \varphi = 1.1, \lambda = 1.5$.

5. Data analyses

In this this section, we implement the NRE-logistic and NRE-Fréchet distributions for analyzing a data set taken from the healthcare economic sector. The data set consists of fifty-nine (59) observations and is taken from an oil company in China during the COVID-19 pandemic. First, we analyze the price data taken from the oil company of China. Then, we obtain the price returns of the same oil company in China during the COVID-19 epidemic.

The comparison of the NRE-logistic is made with the (*i*) logistic (a two-parameter model; see Eq (7)), (*ii*) exponentiated logistic (Exp-logistic) which is a three-parameter competing distribution, and (*iii*) Kumaraswamy logistic (Kum-logistic) which is a four-parameter competing distribution. The CDFs of the Exp-logistic and Kum-logistic distributions are given by

$$K(x;\alpha,\theta,\sigma) = \left(\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right)^{\alpha}, \qquad x \in \mathbb{R}, \alpha \in \mathbb{R}^{+},$$

and

$$K(x;\alpha,\beta,\theta,\sigma) = 1-\left(1-\left[\left(1+e^{-\left(\frac{x-\theta}{\sigma}\right)}\right)^{-1}\right]^{\alpha}\right)^{\beta}, \qquad x \in \mathbb{R}, \alpha \in \mathbb{R}^{+}, \beta \in \mathbb{R}^{+},$$

respectively.

Whereas, the results of the NREi-Fréchet distribution are compared with (*i*) the Fréchet distribution (a two-parameter model; see Eq (10)) and alpha power transformed Fréchet (APT-Fréchet) which is a three-parameter competing distribution. The CDF of the APT-Fréchet distribution is given by

$$K(x;\alpha_1,\eta,\varphi) = \frac{\alpha_1^{e^{-\varphi x^{-\eta}}}-1}{\alpha_1-1}, \qquad x \in \mathbb{R}, \alpha_1 \in \mathbb{R}^{+}, \alpha_1 \neq 1.$$

To figure out analytically/numerically which competing model provides the close fit to the economic data collected during the COVID-19 pandemic, three goodness-of-fit tests such as (*i*) Anderson-Darling (AD) test, (*ii*) Cramér-von Mises (CVM) test, and (iii) Kolmogorov-Smirnov (KS) test, are considered. Besides these tests, the p-value of the competing models has also been calculated. The numerical

values of the goodness-of-fit tests and MLEs of the fitted distributions are obtained using the R-script Adequacy Model with BFGS algorithm.

5.1. An application of the NRE-Fréchet model

This section is devoted to implementing the NRE-Fréchet distribution for modeling the oil prices data, given by: 369.66, 373.55, 351.07, 359.46, 343.74, 314.33, 322.73, 333.15, 347.80, 363.43, 397.06, 403.46, 425.79, 400.65, 405.52, 433.35, 468.06, 465.40, 488.27, 487.03, 516.71, 532.23, 432.55, 371.54, 384.04, 411.75, 428.13, 460.63, 459.14, 412.43, 422.83, 407.27, 427.24, 405.65, 423.91, 444.75, 426.54, 373.43, 225.98, 148.81, 216.04, 279.61, 294.83, 301.01, 276.45, 267.42, 279.53, 318.90, 346.71, 390.55, 415.50, 410.38, 427.01, 461.28, 474.47, 446.01, 470.01, 526.12, 510.69.

For easily carrying out the numerical computation, we transformed the oil prices data using $X = \frac{Y}{10}$, where Y represents the data set. Some summary measures (SMs) of the transformed data are given by: is is given by minimum = 1.488, 1st Quartile (Q_1) = 3.473, Median (Q_2) = 4.056, Mean =3.912, 3rd Quartile (Q_1) = 4.391, maximum = 5.322, skewness = -0.6735137, kurtosis = 3.442013, variance = 0.6276279, and range = 3.8342. In line with the oil price data, some key plots are obtained in Figure 9.

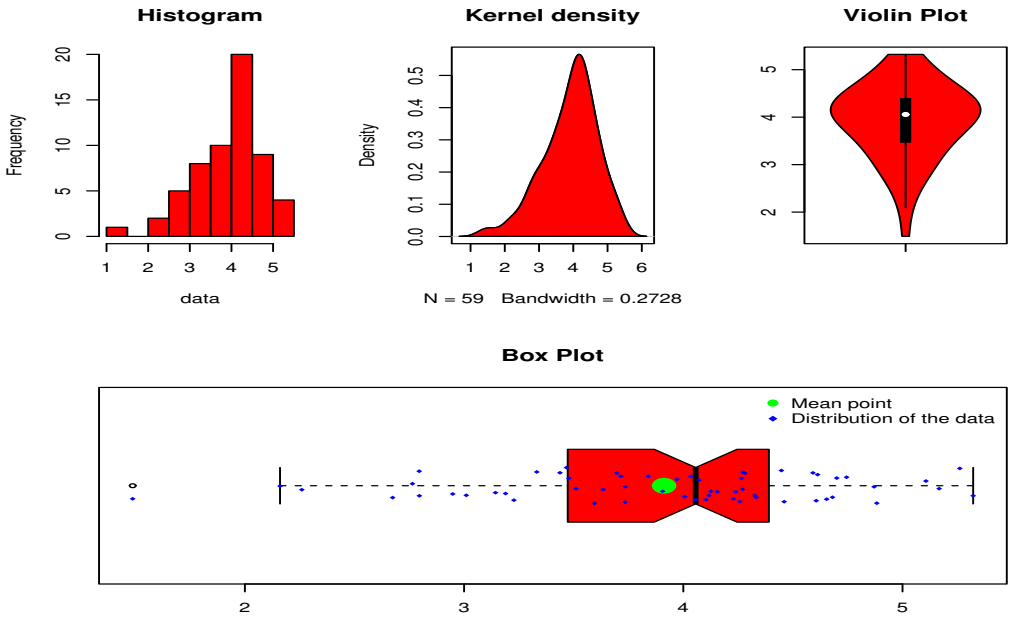

Figure 9. Some basic plots of the oil prices data.

Corresponding to the oil prices data, the values of the maximum likelihood estimators (MLE) $\left(\hat{\lambda}_{MLE}, \hat{\eta}_{MLE}, \hat{\varphi}_{MLE}, \hat{\alpha}_{1\,MLE}\right)$ of the fitted models are reported in Table 5. For the NRE-Fréchet distribution, the profiles of the log-likelihood function of the MLES are displayed visually in Figure 10.

Furthermore, the values of AD, CVM, KS and p-value are obtained in Table 6. From the given results in Table 6, we can see that (i) for the NRE-Fréchet model, the values of the selection criteria are CVM = 0.27746, AD = 1.59522, KS = 0.11565, p-value = 0.38000, (ii) for the Fréchet model, the values of the selection criteria are CVM = 0.70603, AD = 4.02970, KS = 0.20811, p-value =

0.60170, and (iii) for the APT-Fréchet model, the values of the selection criteria are CVM = 0.50369, AD = 2.873401, KS = 0.16504, p-value = 0.07141. From the above discussion, it is obvious that the NRE-Fréchet distribution provides a close fit to the oil prices data.

Besides the numerical illustration, visual proof of the best fitting of the NRE-Fréchet distribution is also provided. For the visual illustration, the plots of fitted PDF, CDF, SF, QQ (quantile-quantile), and PP (probability-probability) are considered; see Figure 11. Based on the visual illustration provided in Figure 11, it is clear that the NRE-Fréchet distribution provides a close fit to the oil prices data.

Table 5. The values of $\hat{\lambda}_{MLE}, \hat{\eta}_{MLE}, \hat{\varphi}_{MLE}$, and $\hat{\alpha}_{1\,MLE}$ of the fitted models for the oil prices data.

Model	$\hat{\lambda}_{MLE}$	$\hat{\eta}_{MLE}$	$\hat{\varphi}_{MLE}$	$\hat{\alpha}_{1\,MLE}$
NRE-Fréchet	0.00317	0.97677	25.00396	-
Fréchet	-	3.14566	44.87837	-
APT-Fréchet	-	4.20771	41.89421	139.89786

Table 6. The values of analytical measures of the fitted models for the oil prices data.

Model	CVM	AD	KS	p-value
NRE-Fréchet	0.27746	1.59522	0.11565	0.38000
Fréchet	0.70603	4.02970	0.20811	0.06170
APT-Fréchet	0.50369	2.873401	0.16504	0.07141

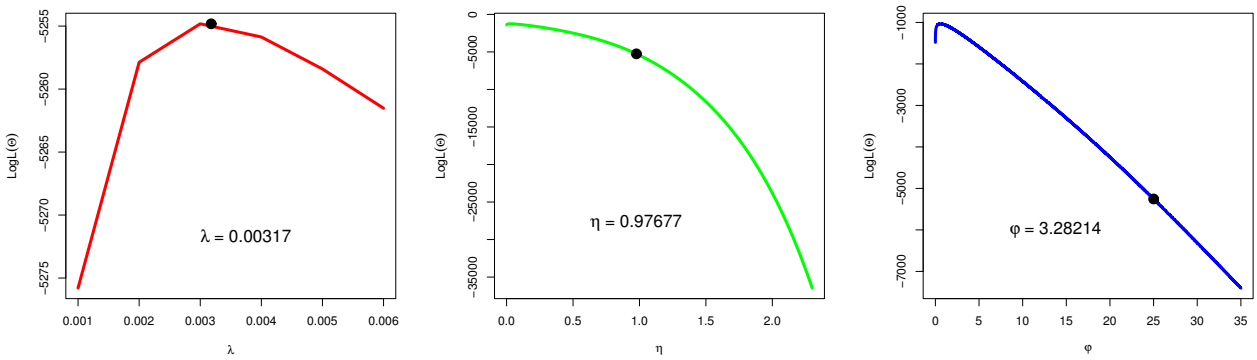

Figure 10. The profiles of the LLF of λ, η, and φ of the NRE-Fréchet distribution using the oil prices data.

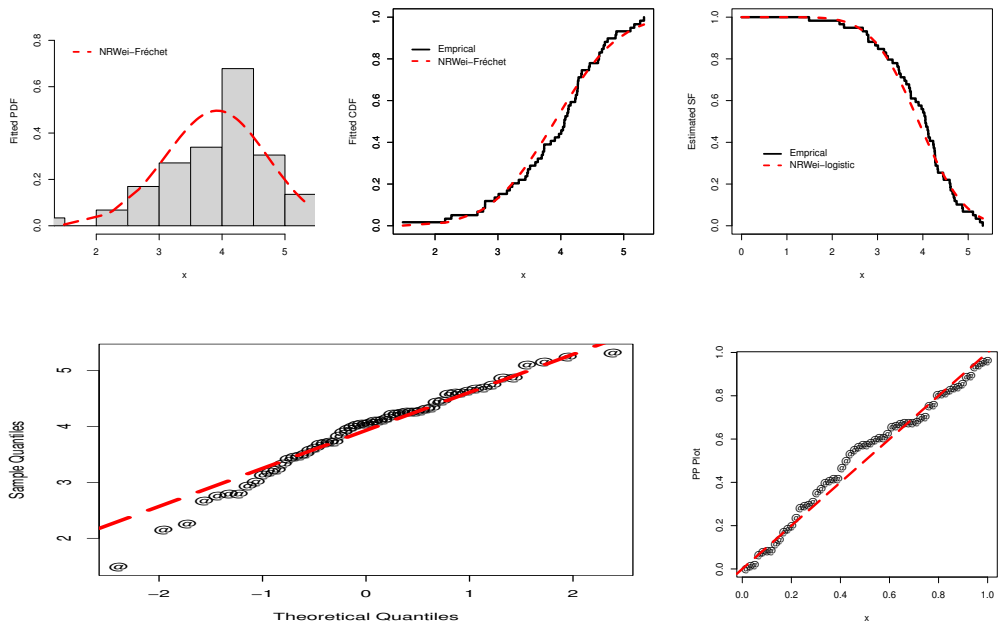

Figure 11. The estimated PDF, CDF, SF, QQ, and PP plots of the NRE-Fréchet distribution for the oil prices data.

5.2. *An application of the NRE-logistic model*

This subsection is devoted to apply the NRE-logistic distribution for analyzing the prices returns of an oil company. The prices returns (log-returns) of the oil company, r_t, is obtained as $r_t = \log(P_t) - \log(P_{t-1})$, where P_t represents the oil price at time t. The SMs of this is given by minimum = -0.502284, 1st Quartile (Q_1) = -0.036423, Median (Q_2) = 0.028893, Mean = 0.005572, 3rd Quartile (Q_3) = 0.053652, maximum = 0.372793, skewness = -1.396718, kurtosis = 9.199987, variance = 0.01558175, and range = 0.8750773. Corresponding to the log-returns of the oil prices data, some basic plots are presented in Figure 12.

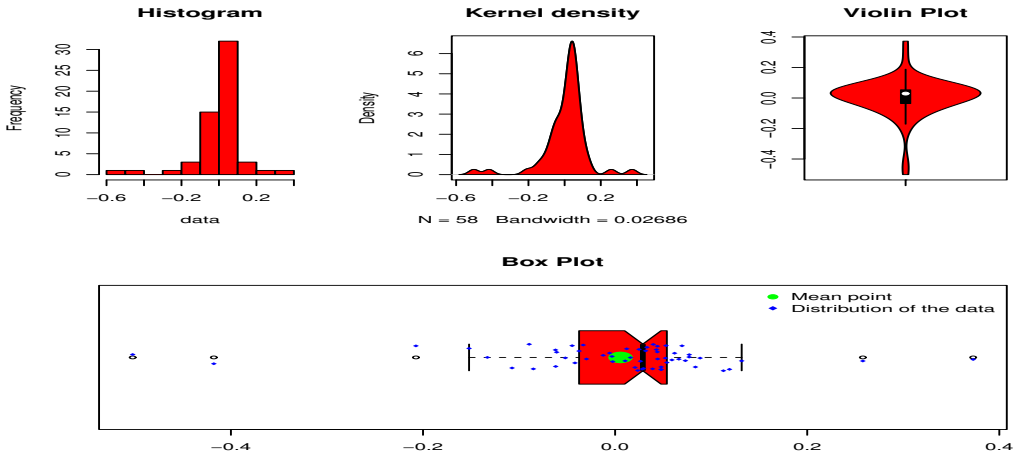

Figure 12. Some basic plots of the log returns of the oil prices data.

In the link to the prices returns data, the values $\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}, \hat{\lambda}_{MLE}$, and $\hat{\alpha}_{MLE}$ of the logistic, Exp-logisitc, and NRE-logistic distributions are presented in Table 7. For the NRE-logistic distribution, the profiles of the log-likelihood function of the MLEs based on the prices returns data are displayed visually in Figure 13.

Furthermore, the values of AD, CVM, KS and p-value are reported in Table 8. From the obtained results in Table 8, we can see that (i) for the NRE-logistict model, the values of the selection criteria are CVM = 0.28420, AD = 1.68629, KS = 0.11902, p-value = 0.35560, (ii) for the Logistic model, the values of the selection criteria are CVM = 0.33210, AD = 1.91355, KS = 0.12817, p-value = 0.27250, (iii) for the Exp-logistic model, the values of the selection criteria are CVM = 0.30118, AD = 1.84708, KS = 0.12076, p-value = 0.31480, and (iv) for the Kum-logistic model, the values of the selection criteria are CVM = 0.32518, AD = 1.96717, KS = 0.13960, p-value = 0.18960. From the above discussion, it is obvious that the NRE-Fréchet distribution provides a close fit to the oil prices data. Based on the obtained results in Table 8, it is clear that the NRE-logistic distribution provides a close fit to the prices returns data of an oil company.

After the proof of the best fitting capability of the NRE-logistic distribution, a visual illustration of the performances of the NREi-logistic model is also provided. For this purpose, the plots of the fitted PDF, CDF, PP, QQ, and SF are again considered; see Figure 14. The plots, in Figure 14, show that the NRE-logistic distribution provides the best fit for the price returns data.

Table 7. The values of $\hat{\lambda}_{MLE}, \hat{\theta}_{MLE}, \hat{\sigma}_{MLE}$, $\hat{\alpha}_{MLE}$, and $\hat{\beta}_{MLE}$ of the fitted models for the price returns data.

Model	$\hat{\lambda}_{MLE}$	$\hat{\theta}_{MLE}$	$\hat{\sigma}_{MLE}$	$\hat{\alpha}_{MLE}$	$\hat{\beta}_{MLE}$
NRE-logistic	1.11894	0.08833	0.06210	-	-
Logistic	-	0.01619	0.05521	-	-
Exp-logistic	-	0.05203	0.03959	0.55644	-
Kum-logistic	-	0.09952	0.07700	1.06195	2.81753

Table 8. The values of analytical measures of the fitted models for the price returns data.

Model	CVM	AD	KS	p-value
NRE-logistic	0.28244	1.68355	0.12088	0.33750
Logistic	0.33210	1.91355	0.12817	0.27250
Exp-logistic	0.30118	1.84708	0.12096	0.31480
Kum-logistic	0.32518	1.96717	0.13960	0.18960

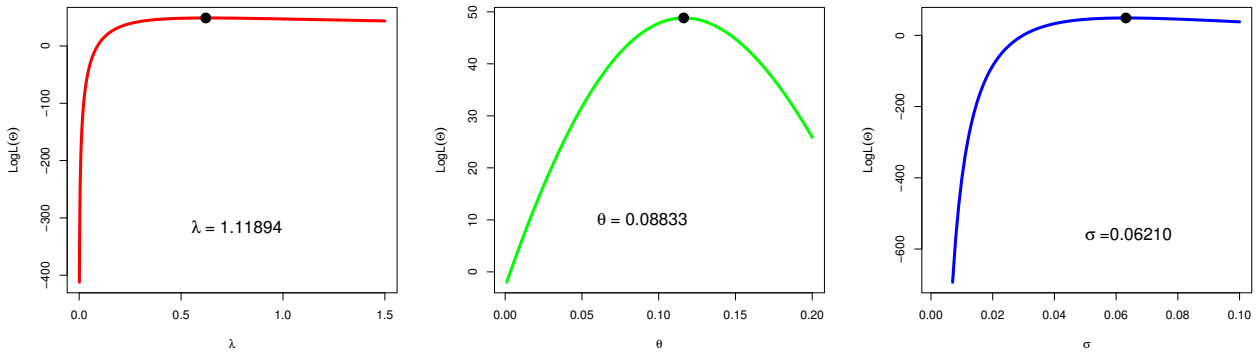

Figure 13. The profiles of the LLF of λ, θ, and σ of the NRE-logistic distribution using the price returns data.

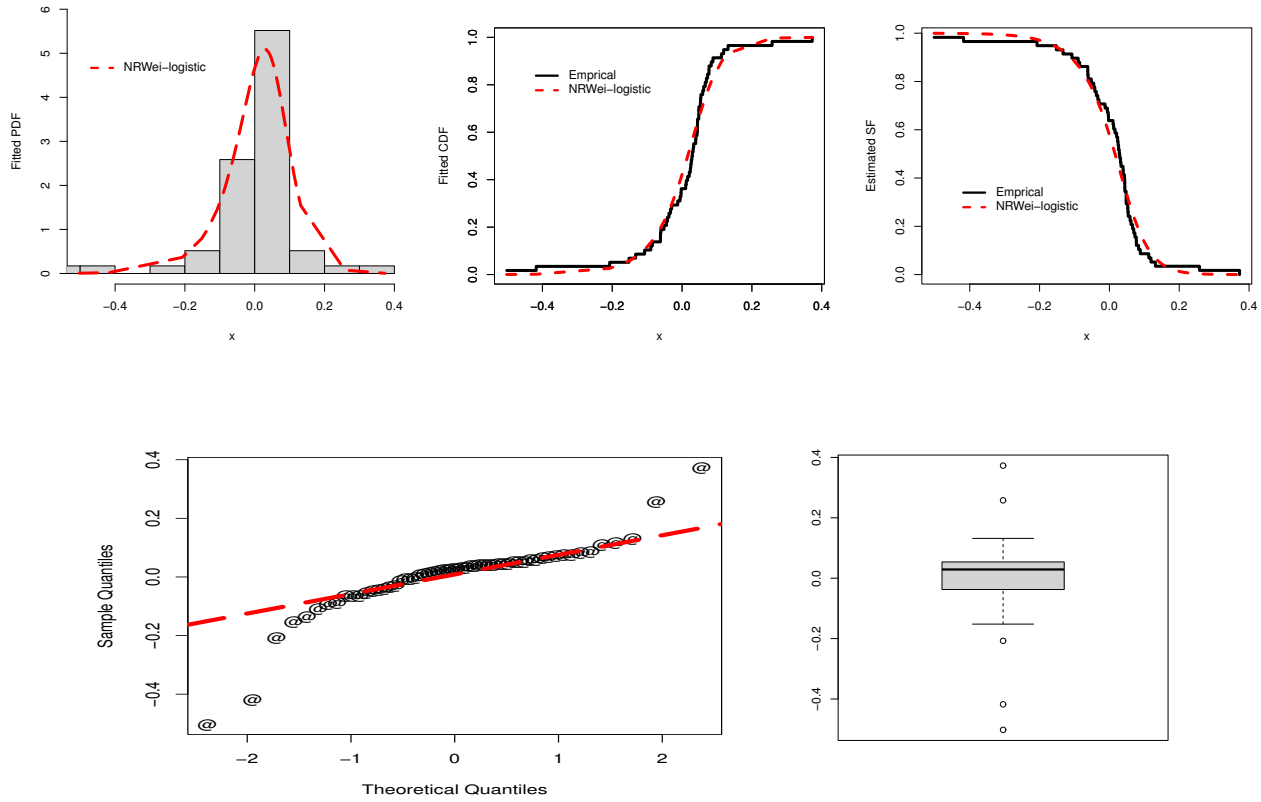

Figure 14. The estimated PDF, CDF, PP, QQ, and SF of the NRE-logistic distribution for the price returns data.

6. The penalized regression techniques

In the classical linear regression model, it is assumed that there is no severe correlation among the set of explanatory variables, and such an ideal situation usually does not exist in practice. The violation of such an assumption induces a problem that refers to multicollinearity. In the presence of severe multicollinearity, the estimates computed from the ordinary least squares (OLS) approach are not reliable. In other words, the estimates of the unknown parameters have a large standard error with the false signs, which adversely affect estimation as well as forecasting. In such circumstances, the most popular way of estimating the unknown parameters is the penalization family of tools.

Generally, penalized regression methods are modified versions of ordinary least squares regression methods (OLS) and are undeniably deemed an essential part of Machine Learning (ML) techniques. It has already been demonstrated in the literature that ML methods are effective for analyzing large datasets [27].

This study adopts the dual most well-known tools of penalized regression, including the least absolute shrinkage and selection operator (Lasso) and elastic net (Enet) for variable selection [28]. Penalized the least square regression performs variable selection and coefficients estimation by minimizing

$$\frac{1}{2} \parallel \vartheta - z\theta \parallel_2^2 + \tau \left[\rho \sum_{k=1}^{m} |\theta_k| + (1-\rho) \sum_{k=1}^{m} |\theta_k|^2 \right],$$

where $\vartheta = (\vartheta_1, \vartheta_2, ..., \vartheta_n)$, $z = (z_1, z_2, ..., z_n)$, and θ is the coefficient matrix with $\theta = (\theta_1, \theta_2, ..., \theta_n)$. Here, m and n represent the number of explanatory variables and data points, respectively. The second term indicates a penalty function, which plays a crucial role in variable selection. The penalty function puts some coefficients exactly equal to zero in order to reduce the dimensionality. The component τ is so-called the hyperparameter that controls the amount of shrinkage, and its range lies between zero and infinity [29, 30]. The different values of parameter ρ are responsible for providing different models, given as follows:

- Ridge regression if $\rho = 0$.
- Lasso regression if $\rho = 1$.
- Elastic net if $0 < \rho < 1$.

The selection of a hyperparameter, also known as a tuning parameter is very crucial and therefore several approaches have been followed in the literature while selecting its optimum value for the model [31]. One of them is cross-validation (CV), that commonly used for obtaining the optimum model. Our study uses the CV approach for the selection of tuning parameters.

Data analysis

After performing univariate analysis, now we carry out multivariate analysis to investigate the impact of macroeconomic variables on oil prices (OP) in the case of China for the period January 2017 to November 2021. The macroeconomic variables include the gross domestic product (GDP), the real effective exchange rate (REER), consumer price index (CPI) and money supply (MS). Although the data is time series, it is more probable to face the problem of autocorrelation, because autocorrelation

is a pure phenomenon of time series data. Therefore, we proceed with our analysis by incorporating the common factors in the models. This study takes into account five lags for each variable (response and explanatory variables), which in turn induce a total of 29 variables. Before modeling, it is more appropriate to determine the correlation pattern among the set of 29 variables. For that purpose, we use the graphical approach.

In Figure 15, red and blue colors manifest the negative and positive association, respectively. The circle area and severity are positively related to the correlation coefficients. There are numerous dark colors in Figure 15, which ensures the problem of multicollinearity in the data.

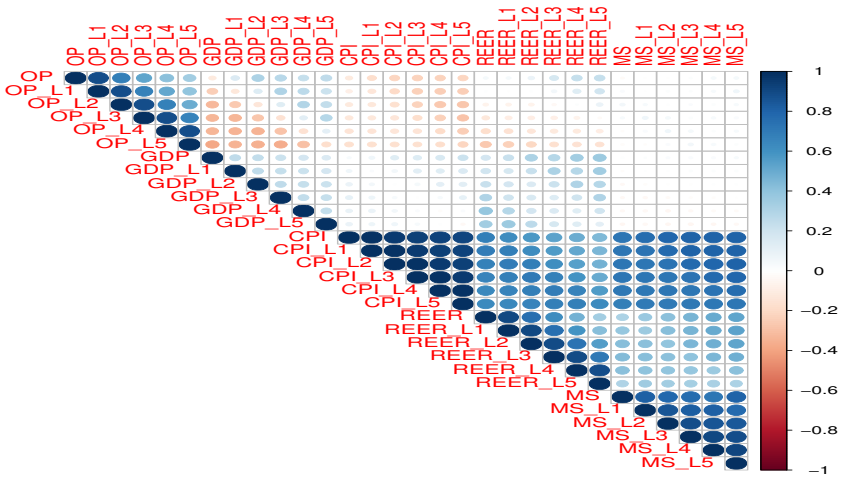

Figure 15. The pairwise correlation of the data.

We observed that the length of explanatory variables is large, and they are highly correlated to each other, and thereby it is plausible to apply the traditional models like vector autoregression, etc., to the underlying dataset [32]. Therefore, this study adopts dual advanced machine learning techniques, namely Lasso and Enet, to resolve these problems and obtain reliable results.

The feature selection plots are presented in Figures 16 and 17. The Lasso holds 22 features out of 29 and the rival approach Enet selects 18 features. The RMSE, MAE and R-squared values are given in Table 9. The Lasso produces the RMSE, MAE and R-squared values of 44.77, 33.84 and 0.75. In contrast, the Enet produces the RMSE, MAE and R-squared values of 41.71, 31.30 and 0.78, respectively. We can observe that the error metrics associated with Enet are smaller than the error metrics attached with Lasso, and similarly the value of Enet R-squared exceeds the Lasso R-squared value. These results clearly reveal that Enet fitted the data more accurately as compared to the competitor counterpart. Figures 18 and 19 are based on the findings given in Table 9. It can be concluded that Lasso holds more irrelevant features and thereby over-specify the model, which in turn adversely affects the performance of Lasso.

Table 9. Error metrics for Lasso and Enet using oil prices data.

Method	RMSE	MAE	R-squared
Lasso	44.77	33.84	0.75
Enet	41.71	31.30	0.78

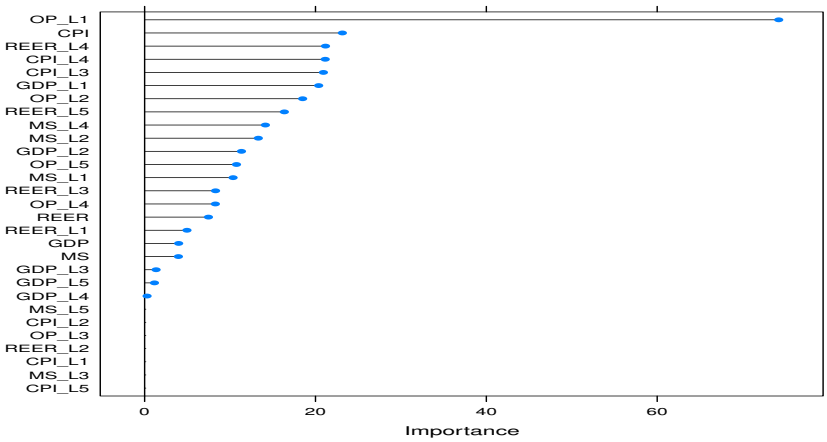

Figure 16. Feature selection via Lasso.

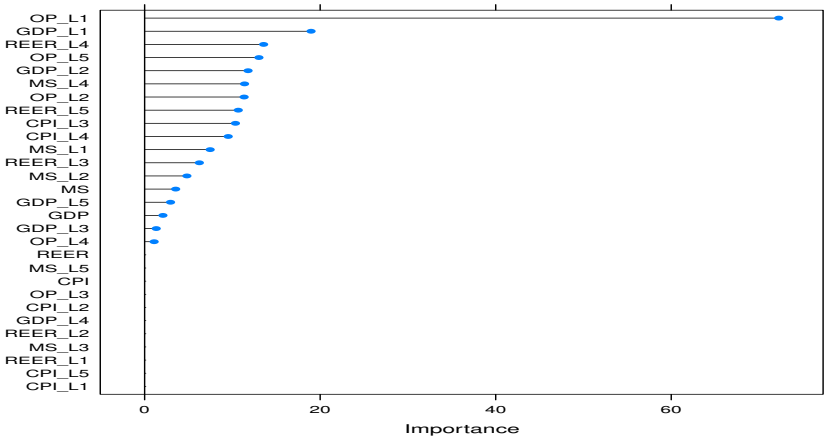

Figure 17. Feature selection via Enet.

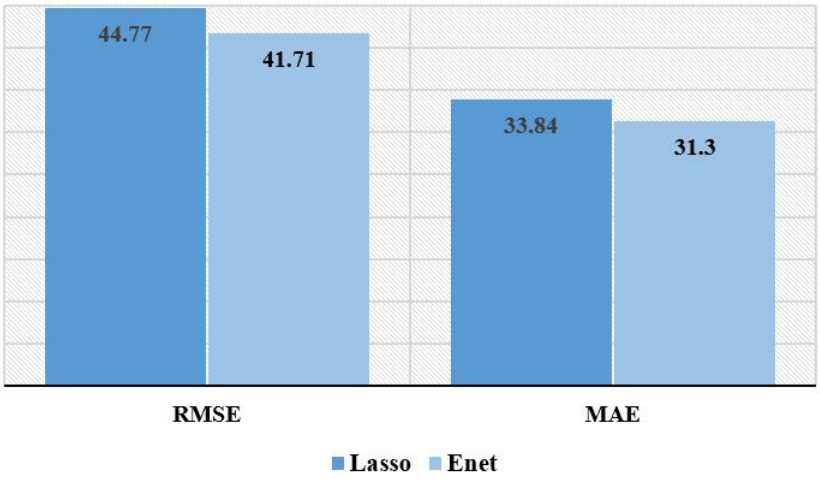

Figure 18. In a sample comparison of Lasso and Enet using oil prices data.

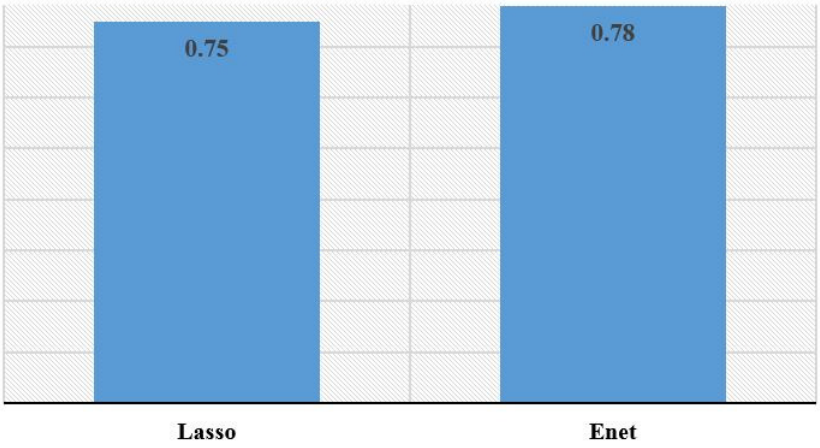

Figure 19. R-squared for Lasso and Enet.

7. Conclusions

The COVID-19 pandemic has highly damaged the healthcare and business sectors. To have enough knowledge and description of the events related to the COVID-19 epidemic, numerous research studies have been performed. Most of these research studies have been carried out by introducing new statistical methodologies. This paper further contributed to the development of the statistical methodologies by introducing a novel statistical approach for modeling the oil price data. The proposed method was named the NRE-*X* approach. Based on the NRE-*X* approach, two new statistical models were introduced. The first model was named the NRE-Fréchet distribution. The proposed NRE-Fréchet distribution was applied for analyzing the oil price data, whereas, the second model was named the NRE-logistic distribution. The suggested NRE-logistic distribution was applied for modeling the log-returns of the oil price data. By considering three different analytical measures (with p-value) as decisive tools, it is observed that the proposed probability distributions were the best competitors for the oil price and log-returns of the oil price data set. Complementing the univariate analysis, we performed the multivariate analysis to examine the influence of macroeconomic variables on oil price in the case of China. For this purpose, our study implemented dual penalization techniques, including Lasso and Enet. The empirical analysis reveals, that Lasso keeps 22 covariates out of 29 and Enet holds 18 covariates. On the basis of three popular statistical accuracy measures, namely, the RMSE, MAE, and R-squared, Enet and Lasso were compared, and it can be inferred that the forecasting ability of Enet is superior to Lasso. It clearly demonstrates that Lasso over-specified the model. In other words, Lasso holds more irrelevant variables in contrast to Enet, which adversely affects its forecasting performance.

In the future, we are motivated to introduce the bivariate and multivariate modifications of the NRE-logistic and NRE-Fréchet distributions for bivariate and multivariate analyses.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R 299), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare no conflict of interest.

References

1. B. B. Ahundjanov, S. B. Akhundjanov, B. B. Okhunjanov, Risk perception and oil and gasoline markets under COVID-19, *J. Econ. Bus.*, **115** (2021), 105979. https://doi.org/10.1016/j.jeconbus.2020.105979
2. C. T. Vidya, K. P. Prabheesh, Implications of COVID-19 pandemic on the global trade networks, *Emerg. Mark. Financ. Tr.*, **56** (2020), 2408–2421. https://doi.org/10.1080/1540496X.2020.1785426
3. M. Ali, N. Alam, S. A. R. Rizvi, Coronavirus (COVID-19). An epidemic or pandemic for financial markets, *J. Behav. Expe. Financ.*, **27** (2020), 100341. https://doi.org/10.1016/j.jbef.2020.100341
4. W. Zhao, S. K. Khosa, Z. Ahmad, M. Aslam, A. Z. Afify, Type-I heavy tailed family with applications in medicine, engineering and insurance, *Plos One*, **15** (2020), e0237462. https://doi.org/10.1371/journal.pone.0237462
5. R. Gerlagh, R. J. Heijmans, K. E. Rosendahl, COVID-19 tests the market stability reserve, *Environ. Resour. Econ.*, **76** (2020), 855–865. https://doi.org/10.1007/s10640-020-00441-0
6. C. Gharib, S. Mefteh-Wali, S. B. Jabeur, The bubble contagion effect of COVID-19 outbreak: Evidence from crude oil and gold markets, *Financ. Res. Lett.*, **38** (2021), 101703. https://doi.org/10.1016/j.frl.2020.101703
7. Z. Ahmad, E. Mahmoudi, S. Dey, S. K. Khosa, Modeling vehicle insurance loss data using a new member of TX family of distributions, *J. Stat. Theory Appl.*, **19** (2020), 133–147. https://doi.org/10.2991/jsta.d.200421.001
8. D. Zhang, M. Hu, Q. Ji, Financial markets under the global pandemic of COVID-19, *Financ. Res. Lett.*, **36** (2020), 101528. https://doi.org/10.1016/j.frl.2020.101528
9. E. Arrigo, C. Liberati, P. Mariani, Social media data and users' preferences: A statistical analysis to support marketing communication, *Big Data Res.*, **24** (2021), 100189. https://doi.org/10.1016/j.bdr.2021.100189
10. M. Z. Rashid, A. S. Akhter, Survival weighted power function distribution with applications to medical, oceanology and metrology data, *Adv. Appl. Stat.*, **67** (2021), 117–253. https://doi.org/10.17654/AS067020133
11. W. Wang, Z. Ahmad, O. Kharazmi, C. B. Ampadu, E. H. Hafez, M. M. M. El-Din, New generalized-X family: Modeling the reliability engineering applications, *Plos One*, **16** (2021), e0248312. https://doi.org/10.1371/journal.pone.0248312

12. C. S. Kumar, S. R. Nair, A generalized Log-Weibull distribution with bio-medical applications, *Int. J. Stat. Med. Res.*, **10** (2021), 10–21. https://doi.org/10.6000/1929-6029.2021.10.02

13. H. S. Bakouch, C. Chesneau, O. A. Elsamadony, The Gumbel kernel for estimating the probability density function with application to hydrology data, *J. Data Inform. Manag.*, **3** (2021), 261–269. https://doi.org/10.1007/s42488-021-00058-y

14. D. Bhati, S. Ravi, On generalized log-Moyal distribution: A new heavy tailed size distribution, *Insur. Math. Econ.*, **79** (2018), 247–259. https://doi.org/10.1016/j.insmatheco.2018.02.002

15. Z. Ahmad, E. Mahmoudi, G. G. Hamedani, O. Kharazmi, New methods to define heavy-tailed distributions with applications to insurance data, *J. Taibah Univ. Sci.*, **14** (2020), 359–382. https://doi.org/10.1080/16583655.2020.1741942

16. Z. Li, J. Beirlant, S. Meng, Generalizing the log-Moyal distribution and regression models for heavy-tailed loss data, *ASTIN Bull. J. IAA*, **51** (2021), 57–99. https://doi.org/10.1017/asb.2020.35

17. Ł. Bielak, A. Grzesiek, J. Janczura, A. Wyłomańska, Market risk factors analysis for an international mining company. Multi-dimensional, heavy-tailed-based modelling, *Resour. Policy*, **74** (2021), 102308. https://doi.org/10.1016/j.resourpol.2021.102308

18. D. B. Madan, Multivariate distributions for financial returns, *Int. J. Theor. Appl. Fin.*, **23** (2020), 20500417. https://doi.org/10.1142/S0219024920500417

19. C. O. Cepoi, Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil, *Financ. Res. Lett.*, **36** (2020), 101658. https://doi.org/10.1016/j.frl.2020.101658

20. W. Wang, W. Li, N. Zhang, K. Liu, Portfolio formation with preselection using deep learning from long-term financial data, *Expert Syst. Appl.*, **143** (2020), 113042. https://doi.org/10.1016/j.eswa.2019.113042

21. S. M. Carta, S. Consoli, A. S. Podda, D. R. Recupero, M. M. Stanciu, Ensembling and dynamic asset selection for risk-controlled statistical arbitrage, *IEEE Access*, **9** (2021), 29942–29959. https://doi.org/10.1109/ACCESS.2021.3059187

22. L. W. Cong, K. Tang, J. Wang, Y. Zhang, Deep sequence modeling: Development and applications in asset pricing, *J. Financ. Data Sci.*, **3** (2021), 28–42. https://doi.org/10.3905/jfds.2020.1.053

23. F. Schuhmacher, H. Kohrs, B. R. Auer, Justifying mean-variance portfolio selection when asset returns are skewed, *Manag. Sci.*, **67** (2021), 7812–7824. https://doi.org/10.1287/mnsc.2020.3846

24. S. Karim, M. U. Akhtar, R. Tashfeen, M. R. Rabbani, A. A. A. Rahman, A. AlAbbas, Sustainable banking regulations pre and during coronavirus outbreak: The moderating role of financial stability, *Econ. Res.-Ekon. Istraž.*, **35** (2022), 3360–3377. https://doi.org/10.1080/1331677X.2021.1993951

25. Z. Ahmad, G. G. Hamedani, N. S. Butt, Recent developments in distribution theory: A brief survey and some new generalized classes of distributions, *Pak. J. Stat. Oper. Res.*, **15** (2019), 87–110. https://doi.org/10.18187/pjsor.v15i1.2803

26. A. Alzaatreh, C. Lee, F. Famoye, A new method for generating families of continuous distributions, *Metron*, **71** (2013), 63–79. https://doi.org/10.1007/s40300-013-0007-y

27. J. L. Castle, J. A. Doornik, D. F. Hendry, Modelling non-stationary 'big data', *Int. J. Forecasting*, **37** (2021), 1556–1575. https://doi.org/10.1016/j.ijforecast.2020.08.002

28. F. Khan, A. Urooj, K. Ullah, B. Alnssyan, Z. Almaspoor, A comparison of Autometrics and penalization techniques under various error distributions: Evidence from Monte Carlo simulation, *Complexity*, 2021. https://doi.org/10.1155/2021/9223763

29. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B*, **58** (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

30. F. Khan, A. Urooj, S. A. Khan, S. K. Khosa, S. Muhammadullah, Z. Almaspoor, Evaluating the performance of feature selection methods using huge big data: A Monte Carlo simulation approach, *Math. Prob. Eng.*, 2022. https://doi.org/10.1155/2022/6607330

31. S. Smeekes, E. Wijler, Macroeconomic forecasting using penalized regression methods, *Int. J. Forecasting*, **34** (2018), 408–430. https://doi.org/10.1016/j.ijforecast.2018.01.001

32. J. H. Stock, M. W. Watson, Macroeconomic forecasting using diffusion indexes, *J. Bus. Econ. Stat.*, **20** (2002), 147–162. https://doi.org/10.1198/073500102317351921

AIMS Press