Mathematics

Research article

On fitting and forecasting the log-returns of cryptocurrency exchange rates using a new logistic model and machine learning algorithms

Zubair Ahmad[1], **Zahra Almaspoor**[1], **Faridoon Khan**[2], **Sharifah E. Alhazmi**[3], **M. El-Morshedy**[4,5,*], **O. Y. Ababneh**[6] and **Amer Ibrahim Al-Omari**[7]

[1] Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran

[2] PIDE School of Economics, Islamabad 44000, Pakistan

[3] Mathematics Department, Al-Qunfudah University College, Umm Al-Qura University, Mecca, KSA

[4] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

[5] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

[6] Department of Mathematics, Faculty of science, Zarqa university, Zarqa, Jordan

[7] Department of Mathematics, Faculty of Science, Al al-Bayt University, Mafraq, Jordan

* **Correspondence:** Email: mah_elmorshedy@mans.edu.eg.

Abstract: Cryptocurrency is a digital currency and also exists in the form of coins. It has turned out as a leading method for peer-to-peer online cash systems. Due to the importance and increasing influence of Bitcoin on business and other related sectors, it is very crucial to model or predict its behavior. Therefore, in recent, numerous researchers have attempted to understand and model the behaviors of cryptocurrency exchange rates. In the practice of actuarial and financial studies, heavy-tailed distributions play a fruitful role in modeling and describing the log returns of financial phenomena. In this paper, we propose a new family of distributions that possess heavy-tailed characteristics. Based on the proposed approach, a modified version of the logistic distribution, namely, a new modified exponential-logistic distribution is introduced. To illustrate the new modified exponential-logistic model, two financial data sets are analyzed. The first data set represents the log-returns of the Bitcoin exchange rates. Whereas, the second data set represents the log-returns of the Ethereum exchange rates. Furthermore, to forecast the high volatile behavior of the same datasets, we apply dual machine learning algorithms, namely Artificial neural network and support vector regression. The effectiveness of these models is evaluated against self exciting threshold autoregressive model.

Keywords: logistic distribution; heavy-tailed distributions; T-*X* family; cryptocurrency; artificial neural networks; support vector regression

Mathematics Subject Classification: 62F09, 62G34

1. Introduction

In recent times, cryptocurrency and blockchain technology has gained the ability to revolutionize economic phenomena worldwide. Over the past couple of years, a new fashion to make payments via cryptocurrencies has arisen. It allows money/asset transfer across the globe without consulting a centralized third party; see Phillip et al. [1], and Alzaatreh and Sulieman [2]. Since 2009, after the introduction of Bitcoin into the financial market, numerous cryptocurrencies have been introduced. Since 2014, around 4500 cryptocurrencies have been launched worldwide. Figure 1 shows the number of cryptocurrencies from 2014 to the present.
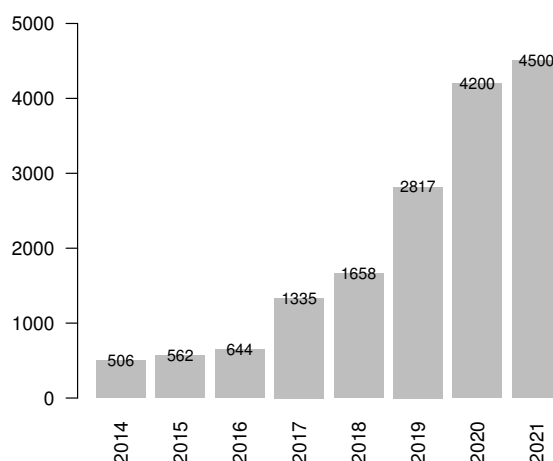

Figure 1. Number of available cryptocurrencies from 2014 to 2021.

Due to the increasing interest of investors in Bitcoin, it has attracted considerable attention in recent years. There is a vast literature on Bitcoin, for example, (*i*) Ciaian et al. [3] implemented a regression approach for predicting Bitcoin prices, (*ii*) Núñez et al. [4] performed a statistical analysis of Bitcoin, (*iii*) Punzo and Bagnato [5] modeled the cryptocurrency return data by implementing the Laplace scale mixtures, (*iv*) Ibrahim et al. [6] used different time series models for predicting the direction moment of Bitcoin, (*v*) Hachicha and Hachicha [7] performed a comparative study of stochastic volatility and MCMC (Markov Chain Monte-Carlo) algorithm using the Bitcoin stock market indexes, and (*vi*) Livieris et al. [8] used a machine learning approach for cryptocurrency forecasting. For more studies related to Bitcoin and other cryptocurrencies, we refer to Chkili et al. [9], Cebrián-Hernández et al. [10], Bazán-Palomino [11], Qin et al. [12], Ghabri et al. [13], Yan [14], Mahdi et al. [15], Liu et al. [16], Naimy et al. [17], and Umar et al. [18].

In this paper, we propose a new method for introducing new HT (heavy-tailed) distributions. The proposed method can be implemented for modeling the HT financial and other related data sets. For case studies, we analyze two data sets related to the log-returns of Bitcoin and Ethereum ERs (exchange rates) based on the USA (United States of America) dollars.

Now, we introduce a new modfied exponential-X (for short "NME-X") family of distributions. The NME-X is introduced by combining the exponential model having PDF (probability density function) $k(t) = e^{-t}$ with the T-X family method.

Consider a random variable, say T, with PDF $k(t)$, where $T \in [\pi_1, \pi_2]$ for $-\infty \leq \pi_1 < \pi_2 \leq \infty$. Let X be a random variable with CDF (cumulative distribution function) expressed by $W(x; \Delta)$ depending on the parameter vector Δ. Let suppose that $F[W(x; \Delta)]$ be a function of CDF of X, satisfying the following conditions

(*i*) $F[W(x; \Delta)] \in [\pi_1, \pi_2]$,
(*ii*) $F[W(x; \Delta)]$ is differentiable and monotonically increasing,
(*iii*) $F[W(x; \Delta)] \to \pi_1$ as $x \to -\infty$ and $F[W(x; \Delta)] \to \pi_2$ as $x \to \infty$.

Then, according to Alzaatreh et al. [19], the CDF of the T-X distributions approach is defined by

$$M(x; \Delta) = \int_{\pi_1}^{F[W(x;\Delta)]} k(t) \, dt, \qquad x \in \mathbb{R}, \tag{1.1}$$

where $F[W(x; \Delta)]$ satisfies the above conditions presented in (*i*)–(*iii*). Corresponding to Eq (1.1), the PDF of the T-X distributions is given by

$$m(x; \Delta) = \left\{ \frac{d}{dx} F[W(x; \Delta)] \right\} k\{F[W(x; \Delta)]\}, \qquad x \in \mathbb{R}.$$

Now, using $k(t) = e^{-t}$ and replacing $F[W(x; \Delta)] = \frac{\beta^2[1 - W(x;\Delta)]}{[\beta + W(x;\Delta)]^2}$ in Eq (1.1), we get the CDF of the NME-X family, given by

$$M(x; \beta, \Delta) = 1 - \frac{\beta^2 \bar{W}(x; \Delta)}{[\beta + W(x; \Delta)]^2}, \qquad \beta > 0, x \in \mathbb{R}, \tag{1.2}$$

where, $\bar{W}(x; \Delta) = 1 - W(x; \Delta)$ is the SF (survival function) of the baseline model.

Based on our study of the literature, the new method defined in Eq (1.2) has not been proposed/used so far. This is one of the key motivations of this work. Henceforth, using the proposed method numerous new distributions can also be introduced for data modeling in different sectors. The PDF $m(x; \beta, \Delta)$ corresponding to Eq (1.2) is given by

$$m(x; \beta, \Delta) = \frac{\beta^2 w(x; \Delta)}{[\beta + W(x; \Delta)]^3} [\beta + 2 - W(x; \Delta)], \qquad \beta > 0, x \in \mathbb{R}, \tag{1.3}$$

where $\frac{d}{dx} W(x; \Delta) = w(x; \Delta)$.

Corresponding to Eq (1.2) and (1.3), the SF $S(x; \beta, \Delta) = 1 - M(x; \beta, \Delta)$ and hazard function (HF) $h(x; \beta, \Delta) = \frac{m(x;\beta,\Delta)}{1-M(x;\beta,\Delta)}$, are given, respectively, by

$$S(x;\beta,\boldsymbol{\Delta}) = \frac{\beta^2 \bar{W}(x;\boldsymbol{\Delta})}{[\beta + W(x;\boldsymbol{\Delta})]^2}, \qquad \beta > 0, x \in \mathbb{R},$$

and

$$h(x;\beta,\boldsymbol{\Delta}) = \frac{w(x;\boldsymbol{\Delta})}{\bar{W}(x;\boldsymbol{\Delta})[\beta + W(x;\boldsymbol{\Delta})]}[\beta + 2 - W(x;\boldsymbol{\Delta})].$$

In this paper, we implement the proposed approach in Eq (1.2) and introduce a new modified/extended version of the logistic distribution, namely, a new modified exponential-logistic (NME-Logistic) model. The next section offers the CDF, PDF, SF, and HF of the NME-Logistic model. Furthermore, different PDF behaviors of the NME-Logisticmodel are also presented in the same section.

2. A NME-logistic distribution

Consider the CDF $W(x;\boldsymbol{\Delta})$ and PDF $w(x;\boldsymbol{\Delta})$ of the two parameters $(\lambda > 0, \eta > 0)$ logistic distribution given by

$$W(x;\boldsymbol{\Delta}) = \frac{1}{1 + e^{-\left(\frac{x-\eta}{\lambda}\right)}}, \qquad x \in \mathbb{R}, \eta \in \mathbb{R}, \lambda \in \mathbb{R}^+, \tag{2.1}$$

and

$$w(x;\boldsymbol{\Delta}) = \frac{e^{-\left(\frac{x-\eta}{\lambda}\right)}}{\left(1 + e^{-\left(\frac{x-\eta}{\lambda}\right)}\right)^2}, \qquad x \in \mathbb{R}, \eta \in \mathbb{R}, \lambda \in \mathbb{R}^+,$$

respectively, where $\boldsymbol{\Delta} = (\phi, \varphi)$. Using Eq (2.1) in Eq (1.2), we get the CDF of the NMExp-Weibull model given by

$$M(x;\beta,\boldsymbol{\Delta}) = 1 - \frac{\beta^2 \left(1 - \frac{1}{1+e^{-\left(\frac{x-\eta}{\lambda}\right)}}\right)}{\left[\beta + \frac{1}{1+e^{-\left(\frac{x-\eta}{\lambda}\right)}}\right]^2}, \qquad x \in \mathbb{R}, \eta \in \mathbb{R}, \lambda, \beta \in \mathbb{R}^+, \tag{2.2}$$

with PDF

$$m(x;\beta,\boldsymbol{\Delta}) = \frac{\beta^2 \frac{e^{-\left(\frac{x-\eta}{\lambda}\right)}}{\left(1+e^{-\left(\frac{x-\eta}{\lambda}\right)}\right)^2}}{\left[\beta + \frac{1}{1+e^{-\left(\frac{x-\eta}{\lambda}\right)}}\right]^3}\left[\beta + 2 - \frac{1}{1 + e^{-\left(\frac{x-\eta}{\lambda}\right)}}\right], \qquad x \in \mathbb{R}. \tag{2.3}$$

Different plots for PDF of the NME-Logistic model are illustrated in Figure 2. The PDF plots of the NME-Logistic model are obtained for $\eta = 4$, $\lambda = 1.5$, and different values of β.
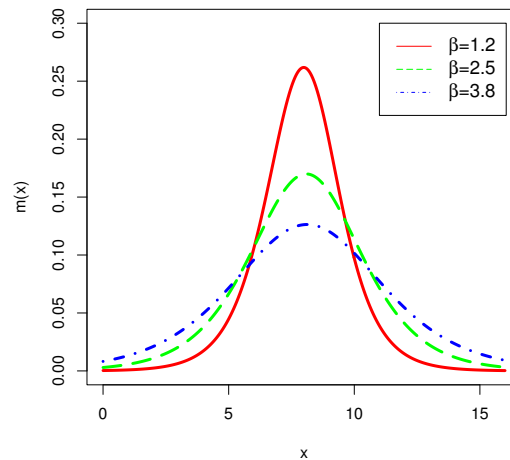
Figure 2. Different plots of $m(x; \beta, \Delta)$ for $\phi = 1.4, \varphi = 0.5$, and different values of β.

Furthermore, the SF and HF of the NME-logistic model are given by

$$S(x; \beta, \Delta) = \frac{\beta^2 \left(1 - \frac{1}{1 + e^{-\left(\frac{x-\eta}{\lambda}\right)}}\right)}{\left[\beta + \frac{1}{1 + e^{-\left(\frac{x-\eta}{\lambda}\right)}}\right]^2}, \qquad x \in \mathbb{R},$$

and

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

respectively.

3. The HT charachteristics of the NMExp-*X* family

Here, we mathematically prove that the NMExp-*X* distributions possess the HT characteristics.

3.1. *The regularly varying tail behavior*

The RVTB (regularly varying tail behavior) plays a curial role to identify HT distributions. In this subsection, we derive the RVTB of the NMExp-*X* family. As per the results of Karamata's theorem Seneta [20], in terms of SF $\bar{W}(x; \Delta)$, we have

Theorem 1. If $S(x; \Delta) = 1 - W(x; \Delta)$ is a RVF (regularly varying function), then $S(x; \beta, \Delta) = 1 - M(x; \beta, \Delta)$ is also a RVF.

Proof. Assume $\lim_{x \to \infty} \frac{S(kx; \Delta)}{S(x; \Delta)} = g(k)$ is finite but nonzero for every $a > 0$. By incorporating Eq (1.2), we have

$$\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \frac{\beta^2\,[1 - W\,(kx;\mathbf{\Delta})]}{[\beta + W\,(kx;\mathbf{\Delta})]^2} \times \frac{[\beta + W\,(x;\mathbf{\Delta})]^2}{\beta^2\,[1 - W\,(x;\mathbf{\Delta})]},$$

$$\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \frac{[1 - W\,(kx;\mathbf{\Delta})]}{[1 - W\,(x;\mathbf{\Delta})]} \times \frac{[\beta + W\,(x;\mathbf{\Delta})]^2}{[\beta + W\,(kx;\mathbf{\Delta})]^2}.$$

Applying $\lim_{x\to\infty}$ on both sides, we get

$$\lim_{x\to\infty}\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \lim_{x\to\infty}\frac{[1 - W\,(kx;\mathbf{\Delta})]}{[1 - W\,(x;\mathbf{\Delta})]} \times \frac{[\beta + W\,(x;\mathbf{\Delta})]^2}{[\beta + W\,(kx;\mathbf{\Delta})]^2}. \tag{3.1}$$

Since $\lim_{x\to\infty} W\,(x;\mathbf{\Delta}) = 1$. So, from Eq (3.1), we have

$$\lim_{x\to\infty}\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \lim_{x\to\infty}\frac{[1 - W\,(kx;\mathbf{\Delta})]}{[1 - W\,(x;\mathbf{\Delta})]} \times \frac{[\beta + W\,(\infty;\mathbf{\Delta})]^2}{[\beta + W\,(k.\infty;\mathbf{\Delta})]^2}\,,$$

$$\lim_{x\to\infty}\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \lim_{x\to\infty}\frac{[1 - W\,(kx;\mathbf{\Delta})]}{[1 - W\,(x;\mathbf{\Delta})]} \times \frac{[\beta + W\,(\infty;\mathbf{\Delta})]^2}{[\beta + W\,(\infty;\mathbf{\Delta})]^2}\,,$$

$$\lim_{x\to\infty}\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \lim_{x\to\infty}\frac{[1 - W\,(kx;\mathbf{\Delta})]}{[1 - W\,(x;\mathbf{\Delta})]} \times \frac{[\beta + 1]^2}{[\beta + 1]^2}\,,$$

$$\lim_{x\to\infty}\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = \lim_{x\to\infty}\frac{S\,(kx;\mathbf{\Delta})}{S\,(x;\mathbf{\Delta})}\,,$$

$$\lim_{x\to\infty}\frac{S\,(kx;\beta,\mathbf{\Delta})}{S\,(x;\beta,\mathbf{\Delta})} = g\,(k)\,. \tag{3.2}$$

The expression provided in Eq (3.2) is finite but nonzero for every $k > 0$. Therefore, $S\,(x;\beta,\mathbf{\Delta})$ is a RVF.

3.2. The regular variational result

Let's assume that the distribution of X has a power-law behavior, then we have

$$S\,(x;\mathbf{\Delta}) = 1 - W\,(x;\mathbf{\Delta}) = \mathbb{P}(X > x) \sim x^{-\lambda}.$$

Using the results of Karamata's characterization theorem, we can write $S\,(x;\beta,\mathbf{\Delta})$ as

$$S\,(x;\beta,\mathbf{\Delta}) = x^{-\lambda}L(x),$$

where $L(x)$ is a SVF. Note that

$$S\,(x;\beta,\mathbf{\Delta}) = \frac{\beta^2\,[1 - W\,(x;\mathbf{\Delta})]}{[\beta + W\,(x;\mathbf{\Delta})]^2}. \tag{3.3}$$

Since $1 - W\,(x;\mathbf{\Delta}) \sim x^{-\lambda}$. Therefore, from Eq (3.3), we get

$$S\,(x;\beta,\mathbf{\Delta}) = \frac{\beta^2 x^{-\lambda}}{(\beta + x^{-\lambda})^2}\,.$$

$$S(x; \beta, \Delta) = x^{-\lambda} L(x),$$

where $L(x) = \frac{\beta^2}{(\beta + x^{-\lambda})^2}$. So, if we can show that $L(x)$ is a SVF, then the RVTB obtained is true. For all $\lambda > 0$, we must show that

$$\lim_{x \to \infty} \frac{L(kx)}{L(x)} = 1. \qquad (3.4)$$

By carrying out the computation, we have

$$\frac{L(kx)}{L(x)} = \frac{\beta^2}{\left(\beta + (kx)^{-\lambda}\right)^2} \times \frac{\left(\beta + x^{-\lambda}\right)^2}{\beta^2},$$

$$\frac{L(kx)}{L(x)} = \frac{\left(\beta + x^{-\lambda}\right)^2}{\left(\beta + (kx)^{-\lambda}\right)^2}.$$

Applying $\lim_{x \to \infty}$ on both sides, we get

$$\lim_{x \to \infty} \frac{L(kx)}{L(x)} = \lim_{x \to \infty} \frac{\left(\beta + x^{-\lambda}\right)^2}{\left(\beta + (kx)^{-\lambda}\right)^2}. \qquad (3.5)$$

Since, $x \to \infty$, then $\lim_{x \to \infty} \frac{1}{x^\lambda} = 0$, and $\lim_{x \to \infty} \frac{1}{x^\lambda k^\lambda} = 0$. Therefore, from Eq (3.5), we obtain

$$\lim_{x \to \infty} \frac{L(kx)}{L(x)} = \frac{\beta^2}{\beta^2},$$

which leads to the proof of Eq (3.4), given by

$$\lim_{x \to \infty} \frac{L(kx)}{L(x)} = 1.$$

4. Analyzing the cryptocurrency data sets

In this section, we implement the NME-logistic distribution for analyzing two real-life data taken from the finance sector. The first data set represents the log-returns of the daily Bitcoin ERs. Whereas, the second data set represents the log-returns of the daily Ethereum ERs.

We fit the NME-logistic distribution to the daily Bitcoin and Ethereum ERs and compare its fitting results with the logistic and Gumble distributions. The SFs corresponding to the competing models are given by

- Logistic distribution

$$S(x; \lambda, \eta) = \frac{e^{-\left(\frac{x-\eta}{\lambda}\right)}}{1 + e^{-\left(\frac{x-\eta}{\lambda}\right)}}, \qquad x \in \mathbb{R}, \eta \in \mathbb{R}, \lambda \in \mathbb{R}^+,$$

and

- Gumble distribution

$$S(x; \lambda, \eta) = 1 - \exp\left\{-e^{-\left(\frac{x-\eta}{\lambda}\right)}\right\}, \qquad x \in \mathbb{R}, \eta \in \mathbb{R}, \lambda \in \mathbb{R}^+,$$

respectively.

To figure out analytically which competing distribution provides the close fit to the Bitcoin ERs data, three goodness of fit tests such as (*i*) Anderson–Darling (AD) test, (*ii*) Cramér–von Mises (CVM) test, and (iii) Kolmogorov–Smirnov (KS) test, were considered. Besides these tests, the p-value of the competing models has also been calculated.

4.1. Analyzing the Bitcoin ERs data

In this subsection, we analyze the log-returns of the Bitcoin ERs data using the NME-logistic distribution and the other competing distributions. The data set is available at `https://coinmarketcap.com/currencies/bitcoin/`. The considered data set represents the daily Bitcoin ERs based on the USA dollars from June 30, 2014, to June 30, 2022.

The log-returns of the daily Bitcoin ERs, r_t, is obtained as $r_t = \log(P_t) - \log(P_{t-1})$, where P_t represents the exchange rate at time t. The summary statistics (SS) of log returns of the Bitcoin ERs are $\bar{x} = -0.0014$, mininum $= -0.1740$, $Q_1 = -0.0222$, $Q_2 = 0.0003$, $Q_3 = 0.0188$, maximum $= 0.1358$, variance $= 0.0013$, range $= 0.3098$, skewness $= -0.3939$, and kurtosis $= 5.2472$. Some basic plots of the daily Bitcoin ERs are presented in Figure 3.
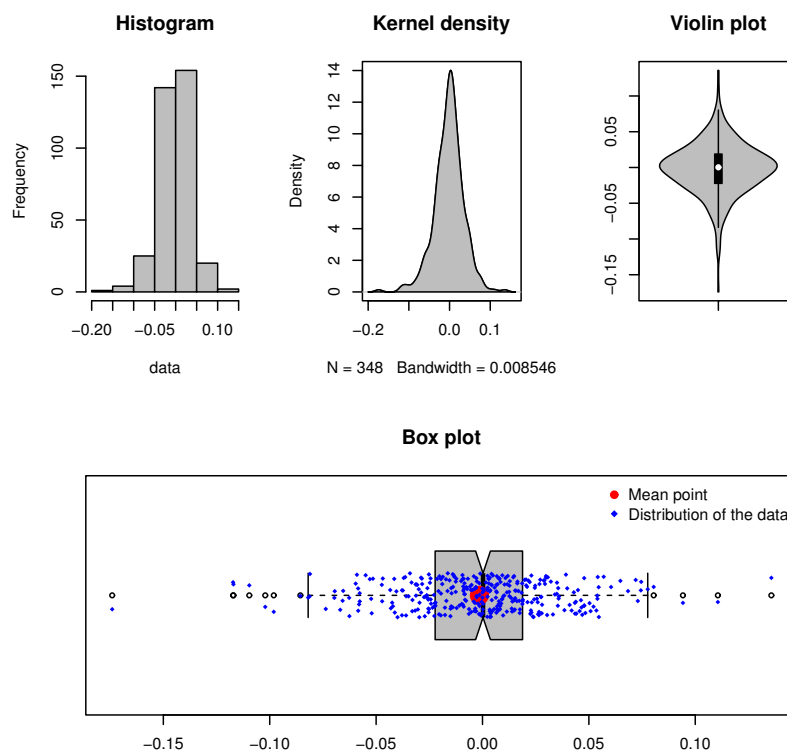

Figure 3. Some basic plots of the daily Bitcoin ERs.

Corresponding to the Bitcoin ERs data, the MLE $\left(\hat{\lambda}_{MLE}, \hat{\eta}_{MLE}, \hat{\beta}_{MLE}\right)$ results of the fitted models are reported in Table 1. Whereas, the values of AD, CVM, KS, and p-value are obtained in Table 2. Based on the reported results in Table 2, it is obvious that the NME-Logistic model provides a close fit to the Bitcoin ERs data.

Table 1. The values of $\hat{\eta}_{MLE}$, $\hat{\lambda}_{MLE}$, and $\hat{\beta}_{MLE}$ of the fitted models for the Bitcoin ERs data.

Model	$\hat{\eta}_{MLE}$	$\hat{\lambda}_{MLE}$	$\hat{\beta}_{MLE}$
NME-Logistic	0.0269	0.0197	0.8200
Logistic	-0.0021	0.0214	-
Gumbel	0.1932	0.1076	-

Table 2. The values of analytical measures of the fitted models for the Bitcoin ERs data.

Model	CVM	AD	KS	p-value
NME-Logistic	0.0690	0.3808	0.0364	0.7445
Logistic	0.1321	0.7310	0.0654	0.1011
Gumbel	0.3876	0.9765	0.1075	0.0953

Besides the numerical illustration of the fitted models, a visual comparison of the fitted models is also provided. For the visual illustration, the plots of fitted the PDFs (Figure 4), CDFs, and Kaplan-Meier (KM) survival plots (Figure 5) are considered. Based on the visual illustration provided in Figures 4 and 5, it is clear that the NME-Logistic model provides a close fit to the Bitcoin ERs data.
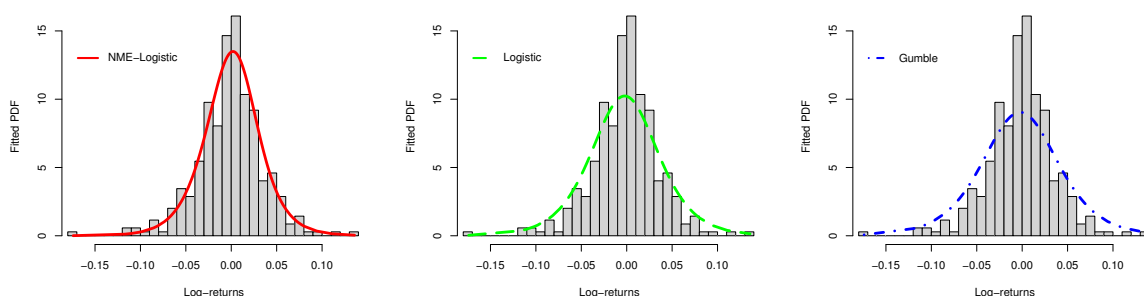

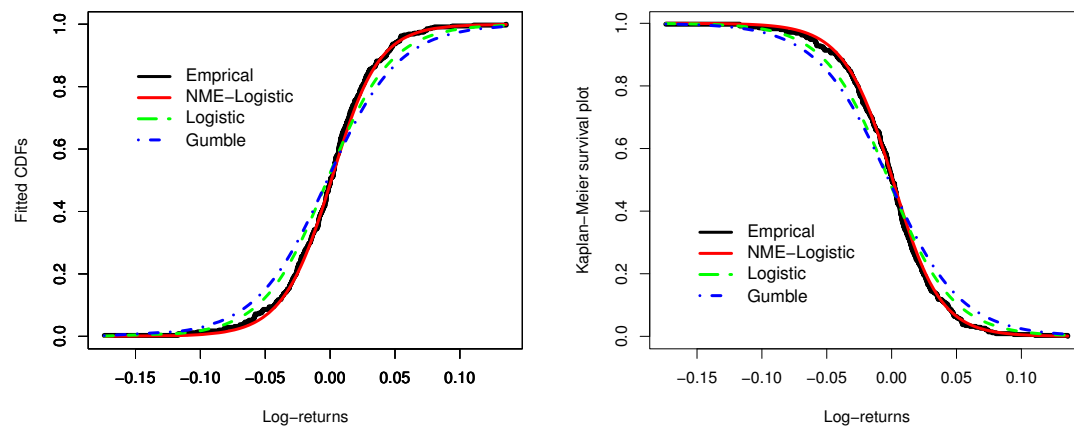**Figure 4.** The plots of the fitted PDFs of the NME-Logistic, logistic and Gumble models for the Bitcoin ERs data.

Figure 5. The plots of the fitted CDFs and SFs of the NME-Logistic, logistic and Gumble distributions for the Bitcoin ERs data.

4.2. Analyzing the Ethereum ERs data

Here, we implement the NME-logistic distribution to analyze the log-returns of the Ethereum ERs data set. The Ethereum ERs data is available at `https://www.google.com/finance/quote/ETH-USD.` The second data set represents the daily Ethereum ERs based on the USA dollars from June 30, 2017, to June 30, 2022. The SS of the log-returns of the Ethereum ERs are $\bar{x} = -0.0019$, mininum $= -0.1777$, $Q_1 = -0.0274$, $Q_2 = 0.0006$, $Q_3 = 0.0243$, maximum $= 0.1258$, variance $= 0.0020$, range $= 0.3036$, skewness $= -0.3096$, and kurtosis $= 4.0979$. Some key plots of the daily Ethereum ERs are provided in Figure 6.
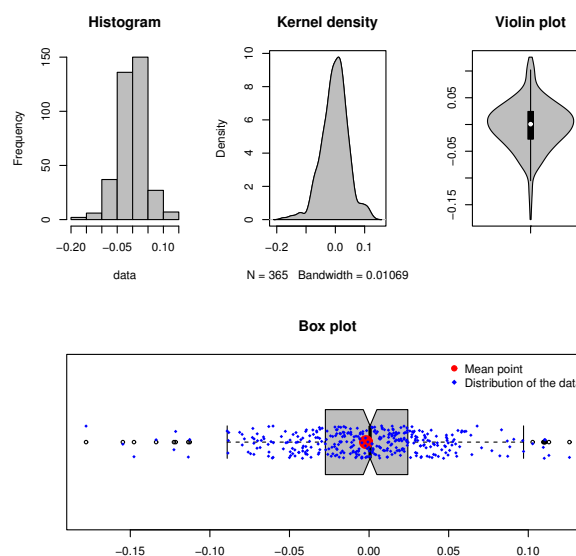

Figure 6. Some basic plots of the daily Ethereum ERs.

Using the Ethereum ERs data set, the values $\hat{\lambda}_{MLE}, \hat{\eta}_{MLE}$, and $\hat{\beta}_{MLE}$ are presented in Table 3. Whereas, the values of AD, CVM, KS and p-value of the fitted distributions are obtained in Table 4. From the numerical results in Table 4, it is clear that the NME-Logistic distribution provides the best fit for the Ethereum ERs data.

Table 3. The values of $\hat{\eta}_{MLE}$, $\hat{\lambda}_{MLE}$, and $\hat{\beta}_{MLE}$ of the fitted models for the Ethereum ERs data.

Model	$\hat{\eta}_{MLE}$	$\hat{\lambda}_{MLE}$	$\hat{\beta}_{MLE}$
NME-Logistic	0.0577	0.0272	0.3193
Logistic	-0.0008	0.0246	-
Gumbel	0.1264	0.0907	-

Table 4. The values of analytical measures of the fitted models for the Ethereum ERs data.

Model	CVM	AD	KS	p-value
NME-Logistic	0.0243	0.2030	0.0239	0.9850
Logistic	0.0737	0.4548	0.0289	0.9195
Gumbel	0.1280	0.8543	0.1015	0.8362

In addition to the numerical illustration of the competing distributions, a visual comparison of the competing models is also provided. For this purpose, the plots of fitted PDFs (Figure 7), CDFs, and KM plots (Figure 8) are obtained. Based on the visual comparison of the competing distributions in Figures 7 and 8, it is obvious that the NME-Logistic model provides the best fit for the Ethereum ERs data.
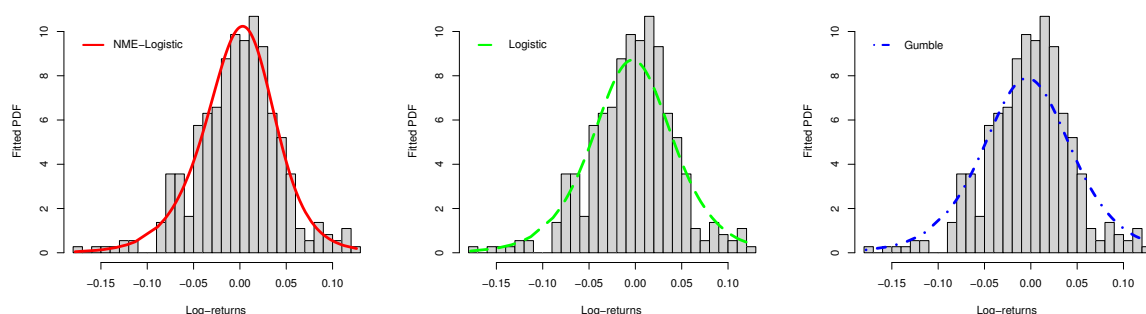

Figure 7. The plots of fitted the PDFs of the NME-Logistic, logistic and Gumble models for the Ethereum ERs data.
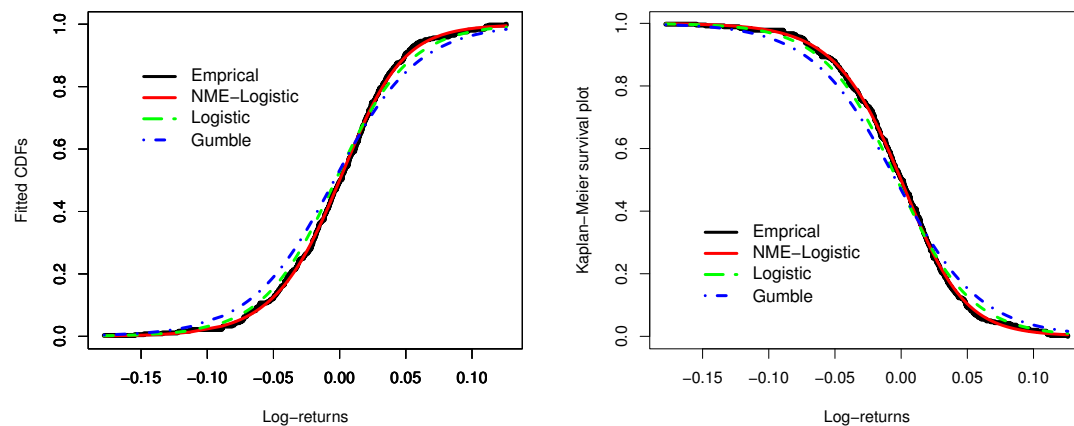
Figure 8. The plots of the fitted CDFs and SFs of the NME-Logistic, logistic and Gumble distributions for the Ethereum ERs data.

5. Forecasting models and analysis

In the previous section, the modified version of the logistic distribution is examined in contrast to some existing probability distributions through real phenomena related to the log-returns of the daily Bitcoin and Ethereum ERs data sets. This section aims to forecast the log-returns of the daily Bitcoin and Ethereum ERs data by applying two popular machine learning techniques, such as artificial neural networks (ANNs) and support vector regression (SVR) including SETAR model. The effectiveness of ML techniques is assessed against the traditional model, namely, the self exciting threshold autoregressive (SETAR) model.

5.1. Artificial neural network

The ANN is a flexible computing algorithm for analyzing a wide range of non-linear problems. A significant superiority of the ANN algorithms over competitive nonlinear tools is, that they can approximate the huge class of functions with a higher degree of accuracy. In addition to this advantage, the construction of ANNs does not require any prior assumption. Alternatively, the network is ascertained by the characteristics of the data; see Zhang [21]. Among a list of numerous networks, the single hidden layer feed-forward network is the most popular algorithm type for predictive modeling. The algorithm is basically characterized by connecting three layers, namely, input layers, hidden layers, and output layers; see Peng et al. [22].

Mathematically, the connection between input layers $(\varnothing_{t-1}, \varnothing_{t-2}, ..., \varnothing_{t-a},)$ and output layer $(\varnothing_t)$ can be expressed as

$$\varnothing_t = \beta_0 + \sum_{i=1}^{k} \beta_i D \left(z_i + \sum_{m=1}^{a} z_{im} \varnothing_{t-m} \right) + \epsilon_t,$$

where $\beta_i\,(i = 1, 2, 3, ..., k)$, $z_{im}\,(i = 1, 2, 3, ..., k, \quad m = 1, 2, 3, ..., a)$ are the unknown parameters, often

refers to connection weight, k and a are indicating the number of hidden and input layers, respectively. Using the first data set, we estimate the model with 3 hidden layers and 5 input layers (lagged variables), and for the second data set, 3 hidden layers and 3 input layers are used. Both hidden layers and input variables are chosen through the error and trial approach followed by Khashei and Hajirahimi [23].

5.2. Support vector regression

The SVR approach was initially proposed by Cortes and Vapnik [24], and till today, it is widely used for classification and regression problems. The SVR is based on the structured risk minimization principle and statistical learning theory, which in turn evades the overfitting problem, and thus yields an accurate forecast. Awan et al. [25] argued that SVR approximates linear and nonlinear real-world problems precisely.

In practice, kernel function plays a key role in the forecasting performance of SVR. The utilization of kernel functions performs operations in the input space in place of higher-dimensional space. Numerous kernel functions are utilized in the literature including linear, polynomial, and sigmoid Radial Basis Functions (RBF) and splines; see Yu et al. [26]. Among these, RBF has gained a substantial focus by dint of its exceptional performance in capturing the non-linear nexus Ghosh [27] and Raje and Mujumdar [28].

The SVR is helpful while determining the margin of error which is acceptable in the model; see Bibi et al. [29] and Ribeiro et al. [30]. The mathematical expression for SVR with kernel function is illustrated as

$$\varnothing_t = \sum_{i=1}^{k} (\beta_i - \beta_i^*) Z(D_i - D) + \epsilon_t,$$

where D illustrates the support-vector, k indicates the size of support vector, $Z(D_i - D)$ indicates the kernel function, and ϵ_t illustrates the threshold value. Herein, the RBF with a parameter $\mathfrak{R}^2$ can be expressed as

$$Z(D_i - D_k) = e^{-\frac{\|D_i - D_k\|^2}{2\mathfrak{R}^2}},$$

where $\|D_i - D_k\|^2$ denotes the Euclidean distance amidst the two covariates in squared form, $\mathfrak{R}^2$ denotes the width of RBF. In our study, we use the RBF as a kernel function for SVR.

5.3. SETAR

Unlike the ML tools delineated in the previous subsections, the SETAR model belongs to the traditional statistical procedures of time series. The SETAR model was initially proposed by Tong and Lim [31], which is basically the modified form of the autoregressive model. The SETAR model is parametric, non-linear in nature, and a special case of markov switiching models; see D'Amato et al. [32]. It assumes that the series behavior varies post entering into a different regime. In Eq (12), suppose m denotes the number of regimes, often equal to two or three (in our case, the high regime and the low regime are represented by mH and mL respectively), and τ is the threshold indicator, where a series changes the regime, a represent the lagged values, the model SETAR(m = 2, a) can be expressed as

$$\emptyset_t = 1_{(\tau_{t-1 \leq th})} \left(z_0^{mL} + z_1^{mL}\emptyset_{t-1} + ... + z_a^{mL}\emptyset_{t-a} \right) \tag{5.1}$$
$$+ 1_{(\tau_{t-1 > th})} \left(z_0^{mH} + z_1^{mH}\emptyset_{t-1} + ... + z_a^{mH}\emptyset_{t-a} \right).$$

If the value of the threshold variable is fixed, then the model is linear and can be estimated by conditional least squares.

5.4. Forecasting of the out-of-sample log-returns of daily Bitcoin and Ethereum ERs

This work uses log-return of daily Bitcoin ERs data in order to quantify the predictability of ANN and SVM algorithms against SETAR model. Therefore, we split the data into two parts, intending to facilitate the out-of-sample prediction accuracy. For estimation, we use the data from June 30, 2014, to June 30, 2022 (log-returns of daily Bitcoin ERs), and June 30, 2017, to June 30, 2022 (log-returns of daily Ethereum ERs), for checking the models' multistep ahead out-of-sample forecasting accuracy. We use 80 percent data for estimating the models, and the remaining 20 percent data, for checking the models' multistep ahead out-of-sample forecasting accuracy, as followed by Ahmad et al. [33].

The predictability is evaluated through popular statistical measures, namely, root-mean-square error (RMSE) and mean absolute error (MAE). The model with lower RMSE and MAE is considered the best model comparatively. The mathematical expressions for MAE and RMSE are, respectively, given by

$$MAE = mean\left(|\emptyset_t - \hat{\emptyset}_t|\right),$$

and

$$RMSE = \sqrt{mean\left(\emptyset_t - \hat{\emptyset}_t\right)^2},$$

where the observed and forecasted values of log-returns of daily Bitcoin and Ethereum ERs data sets are given by $\emptyset_t$ and $\hat{\emptyset}_t$, respectively.

5.4.1. Analysis of the log-returns of the daily Bitcoin ERs

Figure 9 depicts the log-returns of daily Bitcoin ERs data, which shows a high volatility nature, and thus it is a very difficult task to forecast it through traditional models. The alternative approaches to capture its high volatile behavior with satisfactory prediction accuracy, are machine learning tools. In this section, we apply ANN, SVR, and SETAR models to analyze the high volatile data.

In Figure 9, the log-returns of daily Bitcoin ERs data are divided by a vertical blue dotted line, where the training part for model estimation and the second part (testing data) is used for out-of-sample prediction.
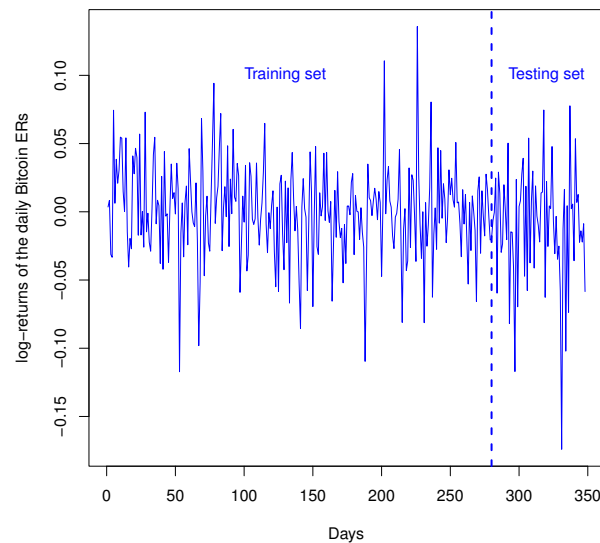
Figure 9. The log-returns of daily Bitcoin ERs data.

The standard accuracy measures, namely, the RMSE and MAE for ML algorithms and the SETAR model are visually displayed in Figure 10. The height of the bar represents the models' forecasting performance in terms of RMSE and MAE. The smaller the height of a bar, the more accurate the forecast is comparatively.

From Figure 10, it can be seen that the corresponding bars for ANN are smaller than the SVR and SETAR models. In addition, SETAR outperforms the SVR. Therefore, we can infer that ANN yields a more satisfactory forecast than the other competitive approaches (SVR and SETAR).
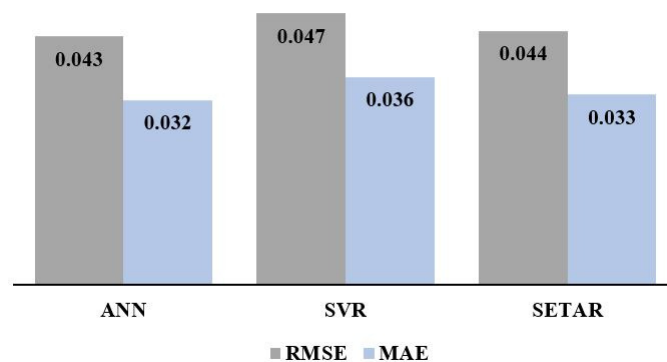

Figure 10. Post sample forecast comparison using the log-returns of the daily Bitcoin ERs data.

5.4.2. Analysis of log-returns of the daily Ethereum ERs

The log-returns of daily Ethereum ERs series are very noisy in nature, as shown by Figure 11, thereby linear models have no capability to fit such data precisely, and provide accurate forecasts. Therefore, our study adopts alternate procedures, including the ML algorithms and SETAR model to

fit the high volatile time series along with satisfactory forecasting accuracy. The vertical dotted line splits the log-returns of daily Ethereum ERs time series into two parts, as shown in Figure 11, where the first part (training set) is used for estimation and the second part (testing set) is used for post-sample prediction.
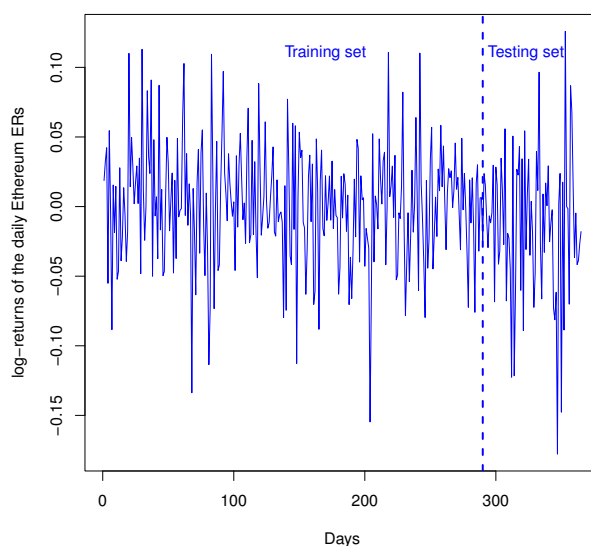

Figure 11. The log-returns of the daily Ethereum ERs data.

The forecast comparison between ML algorithms and the SETAR model is built through dual accuracy measures such as RMSE and MAE, which are presented in Figure 12. The bar height illustrates the predictability of the selected models in terms of RMSE and MAE. In general, the lesser height of a bar ensures a more reliable forecast. We can observe that the RMSE and MAE computed for SETAR are smaller than the ML algorithms, which ensures the robustness of the SETAR model. Moreover, across the ML techniques, the ANN produces a more accurate forecast than the SVR.
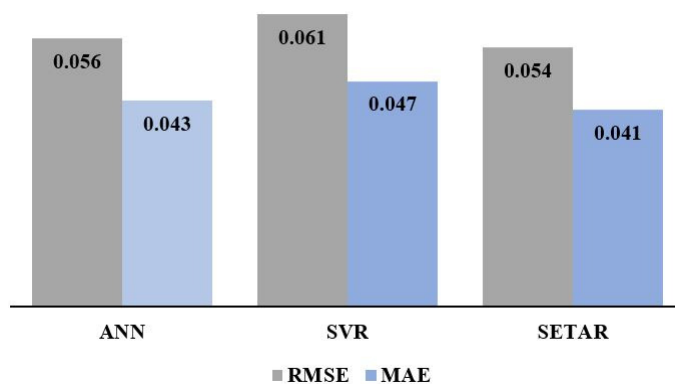

Figure 12. Post sample forecast comparison using the log-returns of the daily Ethereum ERs data.

6. Concluding remarks

Cryptocurrencies are taking place in most aspects of daily life dealing with payments. Among the cryptocurrencies, Bitcoin and Ethereum hold key places and experience heavy-tailed behaviors. This paper contributed to the literature on distribution theory by introducing a new HT version of the logistic distribution, namely, a new modified exponential-logistic distribution. The HT characteristics of the NME-logistic distribution were proved empirically and visually. To establish the applicability of the NME-logistic distribution, two financial data sets related to the log-returns of Bitcoin and Ethereum exchange rates are analyzed. Furthermore, in order to forecast the high volatile behavior of the same datasets, we implement the dual ML algorithms, namely ANN and SVR in comparison to the SETAR model. To assess the effectiveness of these models, we split the data into two parts intending to facilitate the out-of-sample prediction accuracy. For models' estimation, we utilize 80 percent data, and the remaining part of the data is used, for checking the models' multistep ahead post sample predictive capability. Using the first dataset, the ANN produces promising results in contrast to SVR and SETAR. As we move towards the second dataset for analysis, the SETAR showed an outstanding performance than the ML algorithms. The study concludes that models' performance varies from data to data, because of its varying nature. Future studies can compare the aforementioned tools with other ML techniques like bagging, boosting, and random forest. In addition, to achieve more improvement in the prediction accuracy, the ensemble procedure can be developed using ANN, SVR, and SETAR.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

References

1. A. Phillip, J. S. K. Chan, C. Peiris, A new look at cryptocurrencies, *Econ. Lett.*, **163** (2018), 6–9. http://doi.org/10.1016/j.econlet.2017.11.020

2. A. Alzaatreh, H. Sulieman, On fitting cryptocurrency log-return exchange rates, *Empir. Econ.*, **60** (2019), 1157–1174. http://doi.org/10.1007/s00181-019-01782-6

3. P. Ciaian, M. Rajcaniova, D. A. Kancs, The economics of BitCoin price formation, *Appl. Econ.*, **48** (2016), 1799–1815. http://doi.org/10.1080/00036846.2015.1109038

4. J. A. Núñez, M. I. Contreras-Valdez, C. A. Franco-Ruiz, Statistical analysis of bitcoin during explosive behavior periods, *PLoS ONE*, **14** (2019), e0213919. http://doi.org/10.1371/journal.pone.0213919

5. A. Punzo, L. Bagnato, Modeling the cryptocurrency return distribution via Laplace scale mixtures, *Physica A*, **563** (2021), 125354. http://doi.org/10.1016/j.physa.2020.125354

6. A. Ibrahim, R. Kashef, L. Corrigan, Predicting market movement direction for bitcoin: A comparison of time series modeling methods, *Comput. Electr. Eng.*, **89** (2021), 106905. http://doi.org/10.1016/j.compeleceng.2020.106905

7. A. Hachicha, F. Hachicha, Analysis of the bitcoin stock market indexes using comparative study of two models SV with MCMC algorithm, *Rev. Quant. Finan. Acc.*, **56** (2021), 647–673. http://doi.org/10.1007/s11156-020-00905-w

8. I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, P. Pintelas, An advanced CNN-LSTM model for cryptocurrency forecasting, *Electronics*, **10** (2021), 287. http://doi.org/10.3390/electronics10030287

9. W. Chkili, A. B. Rejeb, M. Arfaoui, Does bitcoin provide hedge to Islamic stock markets for pre- and during COVID-19 outbreak? A comparative analysis with gold, *Resour. Policy*, **74** (2021), 102407. http://doi.org/10.1016/j.resourpol.2021.102407

10. Á. Cebrián-Hernández, E. Jiménez-Rodríguez, Modeling of the Bitcoin volatility through key financial environment variables: An application of conditional correlation MGARCH models, *Mathematics*, **9** (2021), 267. http://doi.org/10.3390/math9030267

11. W. Bazán-Palomino, How are Bitcoin forks related to Bitcoin?, *Financ. Res. Lett.*, **40** (2021), 101723. http://doi.org/10.1016/j.frl.2020.101723

12. M. Qin, C. W. Su, R. Tao, BitCoin: A new basket for eggs?, *Econ. Model.*, **94** (2021), 896–907, http://doi.org/10.1016/j.econmod.2020.02.031

13. Y. Ghabri, K. Guesmi, A. Zantour, Bitcoin and liquidity risk diversification, *Financ. Res. Lett.*, **40** (2021), 101679. http://doi.org/10.1016/j.frl.2020.101679

14. M. Liu, H. Chen, J. Yan, Detecting roles of money laundering in Bitcoin mixing transactions: A goal modeling and mining framework, *Front. Phys.*, **9** (2021), 665399. http://doi.org/10.3389/fphy.2021.665399

15. E. Mahdi, V. Leiva, S. Mara'Beh, C. Martin-Barreiro, A new approach to predicting cryptocurrency returns based on the gold prices with support vector machines during the COVID-19 pandemic using sensor-related data, *Sensors*, **21** (2021), 6319. http://doi.org/10.3390/s21186319

16. X. F. Liu, X. J. Jiang, S. H. Liu, C. K. Tse, Knowledge discovery in cryptocurrency transactions: a survey, *IEEE Access*, **9** (2021), 37229–37254. http://doi.org/10.1109/ACCESS.2021.3062652

17. V. Naimy, O. Haddad, G. Fernández-Avilés, R. El Khoury, The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies, *PLoS ONE*, **16** (2021), e0245904. http://doi.org/10.1371/journal.pone.0245904

18. M. Umar, C. W. Su, S. K. A. Rizvi, X. F. Shao, Bitcoin: A safe haven asset and a winner amid political and economic uncertainties in the US?, *Technol. Forecast. Soc.*, **167** (2021), 120680. http://doi.org/10.1016/j.techfore.2021.120680

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

20. E. Seneta, Karamata's characterization theorem, feller and regular variation in probability theory, *Publications de l'Institut Mathématique*, **71** (2002), 79–89. http://doi.org/10.2298/PIM0271079S

21. G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, **50** (2003), 159–175, http://doi.org/10.1016/S0925-2312(01)00702-0

22. Z. Peng, F. U. Khan, F. Khan, P. A. Shaikh, Y. Dai, I. Ullah, et al., An application of hybrid models for weekly stock market index prediction: Empirical evidence from SAARC countries, *Complexity*, **2021** (2021), 5663302. http://doi.org/10.1155/2021/5663302

23. M. Khashei, Z. Hajirahimi, A comparative study of series arima/mlp hybrid models for stock price forecasting, *Commun. Stat.-Simulat. Comput.*, **48** (2019), 2625–2640. http://doi.org/10.1080/03610918.2018.1458138

24. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.*, **20** (1995), 273–297. http://doi.org/10.1007/BF00994018

25. S. M. Awan, Z. A. Khan, M. Aslam, W. Mahmood, A. Ahsan, Application of NARX based FFNN, SVR and ANN Fitting models for long term industrial load forecasting and their comparison, *2012 IEEE International Symposium on Industrial Electronics*, IEEE, 2012, 803–807. http://doi.org/10.1109/ISIE.2012.6237191

26. P. S. Yu, T. C. Yang, S. Y. Chen, C. M. Kuo, H. W. Tseng, Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting, *J. Hydrol.*, **552** (2017), 92–104. http://doi.org/10.1016/j.jhydrol.2017.06.020

27. S. Ghosh, SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output, *J. Geophys. Res.: Atmos.*, **115** (2010), D22102. http://doi.org/10.1029/2009JD013548

28. D. Raje, P. P. Mujumdar, A comparison of three methods for downscaling daily precipitation in the Punjab region, *Hydrol. Process.*, **25** (2011), 3575–3589. http://doi.org/10.1002/hyp.8083

29. N. Bibi, I. Shah, A. Alsubie, S. Ali, S. A. Lone, Electricity spot prices forecasting based on ensemble learning, *IEEE Access*, **9** (2021), 150984–150992. http://doi.org/10.1109/ACCESS.2021.3126545

30. M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, L. dos Santos Coelho, Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil, *Chaos Soliton. Fract.*, **135** (2020), 109853. http://doi.org/10.1016/j.chaos.2020.109853

31. H. Tong, K. S. Lim, Threshold autoregression, limit cycles and cyclical data, In: *Exploration of a nonlinear world: An appreciation of Howell Tong's contributions to statistics*, World Scientific Publishing, 2009, 9–56. https://doi.org/10.1142/9789812836281_0002

32. V. D'Amato, S. Levantesi, G. Piscopo, Deep learning in predicting cryptocurrency volatility, *Physica A*, **596** (2022), 127158. http://doi.org/10.1016/j.physa.2022.127158

33. Z. Ahmad, Z. Almaspoor, F. Khan, M. El-Morshedy, On predictive modeling using a new flexible Weibull distribution and machine learning approach: Analyzing the COVID-19 data, *Mathematics*, **10** (2022), 1792. http://doi.org/10.3390/math10111792