ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

Estuscience – Se, 2024, 25 [2] pp. 180-192, DOI: 10.18038/estubtda.1317322

RESEARCH ARTICLE

MODELLING STOCK PRICES OF A BANK WITH EXTREME VALUE DISTRIBUTIONS

Ceren ÜNAL [1,*] Gamze ÖZEL KADILAR [2]

[1] Department of Statistics, Faculty of Science, Hacettepe University, 06800, Ankara, Türkiye
cerenunal@hacettepe.edu.tr - 0000-0002-9357-1771

[2] Department of Statistics, Faculty of Science, Hacettepe University, 06800, Ankara, Türkiye
gamzeozl@hacettepe.edu.tr - 0000-0003-3886-3074

Abstract

The study investigates the application of Extreme Value Theory in modelling stock prices, aiming to capture the tail behaviour and extreme movements that conventional distributions often fail to represent accurately. The use of Extreme Value Theory has gained considerable attention in the field of finance due to its ability to model rare events, such as financial crises or market crashes. By incorporating Extreme Value Theory, researchers aim to improve risk management, portfolio optimization, and pricing of financial derivatives. In this study, the Log-normal, Weibull, Gamma, and Normal distributions were used to model the stock price closing data, with a specific focus on extreme value distributions. Both graphical explorations and goodness-of-fit criteria were considered together to evaluate the suitability of these distributions. When assessing the data, it was observed that the Weibull distribution provided the best fit for the given stock price closing data.

Keywords

Extreme value theory,
Generalized extreme value distribution,
Tail behaviour,
Stock prices,
Risk management

Time Scale of Article

Received :20 June 2023
Accepted : 31 May 2024
Online date :28 June 2024

1. INTRODUCTION

One of the most crucial challenges in mathematical finance revolves around determining the distribution of speculative prices. This problem holds significant theoretical and practical implications, yet it remains unsolved. The initial solution proposed by Bachelier [1] was the random walk process, where price movements were modelled as independent rises and falls [2]. According to this model, returns were assumed to follow a Normal (Gaussian) distribution. Stock price modelling with statistical distributions involves fitting historical stock price data to a specific distribution to capture the underlying patterns and behaviour of the stock.

The Normal distribution is widely employed in finance for measuring security prices. However, in reality, the distribution of financial assets is typically non-normal [3, 4]. Empirical studies consistently demonstrate that actual stock price returns exhibit skewness, kurtosis, and heavy tails [5]. The presence of heavy tails in price distributions holds significant importance in financial risk analysis. Consequently, it is crucial to obtain a reliable distribution [6]. Thus, alternative distributions are explored as substitutes for the Normal distribution.

Mandelbrot [7] was among the first to propose an explanation for the presence of long tails in financial data, leading to the development of a leptokurtic distribution. However, this initial attempt resulted in probability density function moments that were not finite. Subsequently, the Log-normal distribution

*Corresponding Author: cerenunal@hacettepe.edu.tr

gained widespread usage for describing the distributions of financial assets, including share prices. The Log-normal distribution provided more reliable outcomes than the Normal distribution since asset prices could not be negative. This viewpoint influenced the development of the Black-Scholes model, a widely used framework for pricing options in finance, which is based on the assumption of a Log-normal distribution for determining option prices. Building upon the Log-normal distribution, extreme value distributions (EVDs) emerged as a tool for analysing the tail behaviour of distributions and estimating tail probabilities, enabling the modelling of asymptotic behaviour. EVDs primarily address probabilistic and statistical inquiries related to exceptionally small or large values within a set of random variables. Extreme events, which occur with low probabilities, hold significant importance in financial risk management. In order to quantify the level of risk involved, it becomes necessary to estimate probabilities in the tail region of the distribution for the given process.

Extreme Value Theory (EVT) is a statistical approach used for modelling extreme events, such as extreme stock price movements, which are of great interest in financial risk management. EVT is based on the analysis of the tail behaviour of probability distributions and focuses on estimating the probabilities of rare events that fall beyond the threshold of Normal distributions. In EVT, three types of extreme value distributions are commonly used: the Gumbel distribution, the Fréchet distribution, and the Weibull distribution. These distributions are known as the Generalized Extreme Value (GEV) distributions. The Weibull distribution is one of the alternative distributions used in financial analysis and risk management. It is particularly useful for modelling extreme events and tail values. This distribution is defined by the Weibull function and has shape and scale parameters. In financial markets, the Weibull distribution is commonly employed for extreme value analysis and risk management [8]. Extreme value analysis involves studying unexpected and rare events, and the Weibull distribution provides a suitable framework for capturing such events. The distribution allows for a flexible modelling approach that can account for the heavy-tailed nature of financial data. The Weibull distribution has applications in various areas of finance, including risk modelling, asset pricing, and insurance. It can be used to estimate tail probabilities, quantify the likelihood of extreme events, and assess the associated risks. By understanding the behaviour of extreme values, financial practitioners can make more informed decisions regarding risk management strategies, portfolio optimization, and pricing of financial derivatives [9].

The Log- normal distribution became widely adopted in finance due to its ability to provide more accurate results compared to the Normal distribution when describing the distributions of financial assets, including share prices. The Log-normal distribution is particularly advantageous as it ensures that asset prices cannot take negative values, which is a fundamental characteristic in financial markets [10]. The application of the Log-normal distribution extended to the development of the Black-Scholes model, a widely used framework for pricing options in finance. The model relies on the assumption of a Log-normal distribution for the underlying asset price to determine option prices. This assumption has proven to be effective in capturing the behaviour of financial markets and has played a significant role in options pricing and risk management [11].

The Gamma distribution is a widely used probability distribution in finance due to its ability to model positive random variables with skewed and positively skewed distributions. The Gamma distribution provides a versatile framework for modelling various financial variables and risk measures in finance. Its ability to capture skewness and positive skewness makes it a valuable tool for understanding the distributions of asset returns, estimating risks, and pricing financial derivatives [12].

In this study, we compare the statistical distributions to model stock price data. The paper is organised as follows: In Section 2, statistical distributions used for stock modelling is defined. Section 3 describes data set and its’ statistical properties. Section 4 presents the results. Conclusion is given in Section 5.

2. METHODOLOGY

Extreme value distributions, such as the Log-normal, Gamma, and Weibull distributions, are commonly used in various fields to model extreme events or rare occurrences. In this section, we define normal distribution and extreme value distributions used in the study briefly.

The normal distribution is often used as a starting point for modelling stock prices. It assumes that stock returns follow a bell-shaped curve, with most returns clustering around the mean and tails becoming less likely as they move away from the mean. However, stock returns often exhibit characteristics such as skewness and excess kurtosis, which are not captured by the normal distribution [13]. If a random variable X has a normal distribution with mean μ and variance σ^2 $(\sigma^2 > 0)$, then probability density function (pdf) of the form is given by

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

The log-normal distribution is widely used to model variables that are the result of the exponential growth of underlying processes. It is characterized by its skewed and right-tailed shape. In finance, the log-normal distribution is often applied to model stock prices, as it accounts for the positive-only nature of prices and the tendency for large price movements. The log-normal distribution is commonly used for modeling stock price data due to its ability to handle variables that are inherently positive and have skewed distributions. It assumes that the logarithm of the variable follows a normal distribution. The pdf of the log-normal distribution is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(lnx-\mu)^2}{\sigma^2}\right)}. \tag{2}$$

The gamma distribution is a generalization of the exponential distribution which is a versatile distribution commonly used to model positive continuous variables. It has two shape parameters that allow for various shapes, including right-skewness and left-skewness. In finance, the gamma distribution is often used to model waiting times, insurance claims, and the distribution of returns. The pdf of the gamma distribution with a shape parameter (θ) and scale parameter (λ) is given by

$$f(x) = \frac{\lambda^\theta}{\Gamma(\theta)} x^{\theta-1} exp(-\lambda x), \tag{3}$$

where $x > 0, \theta > 0, \lambda > 0$ and $\Gamma(\theta)$ is the gamma function.

The Weibull distribution was introduced by the Swedish physicist Weibull [14] and is a flexible distribution that can model a wide range of shapes, including both right-skewed and left-skewed distributions [15]. It is often used to model time-to-failure data or the distribution of extreme events. In finance, the Weibull distribution is sometimes employed to model the tail behaviour of asset returns or the distribution of extreme returns. The Weibull distribution is a statistical distribution commonly used in reliability analysis to model the time to failure of a system or an event. However, it is not commonly used for directly modelling stock price data. The pdf of the Weibull distribution is given by

$$f(x) = \lambda\theta x^{\theta-1} exp\left(-\lambda x^\theta\right), \quad x > 0. \tag{4}$$

The Weibull distribution is characterized by its shape parameter (θ) and scale parameter (λ). It has a flexible shape that can be either positively skewed $(\theta > 1)$ or negatively skewed $(\theta < 1)$, and it allows for modelling different hazard rates (the probability of an event occurring over time).

3. DATA SET

HDFC Bank is one of the largest and most prominent private sector banks in India. As a publicly traded company, HDFC Bank's stock price data refers to the historical prices at which its shares have traded on the stock market over a specific period. This data includes the opening price, closing price, high price, low price, and trading volume of HDFC Bank's stock on different trading days.

The stock price of HDFC Bank is influenced by various factors, including the overall performance of the banking sector, the bank's financial results, macroeconomic conditions, interest rates, market sentiment, and news or events that may affect the banking industry or the company specifically. By analysing HDFC Bank's stock price data, investors, traders, and analysts can gain insights into the past performance, volatility, trends, and potential future movements of the stock. This data can be used for various purposes, including technical analysis, fundamental analysis, risk assessment, and investment decision-making [16]. In this study, previous close price is used to analyse stock price data and previous close price refers to the prior day's value of a stock, bond, commodity, futures or any other security. Descriptive statistics is presented in Table 1.

Table 1. Descriptive statistics for previous close price of stock price.

Min. Value	Q_1	Median	Mean	Q_3	Max. Value	Variance	St. Deviation
157.4	479.9	934.8	1007.1	1421.0	2565.8	404187.9	635.757

In the analysis, the Agostino test for skewness and the Anscombe test for kurtosis were performed using R software with "moments" package. The skewness value was calculated as 0.5637, with a p-value that is very close to zero. This indicates that the data is positively skewed, as the p-value is less than the significance level of 0.05. The kurtosis value obtained was three, which suggests a Normal distribution. If the kurtosis value is greater than three, it indicates a leptokurtic distribution with thicker tails and sharper peaks. On the other hand, if the kurtosis value is less than 3, it indicates a platykurtic distribution with thinner tails and flatter peaks. In this case, since the kurtosis value is less than three, the data exhibits a platykurtic distribution. According to the Anscombe test, the kurtosis value is determined to be 2.2974. This further confirms that the data follows a platykurtic distribution with thin tails. Additionally, the p-value obtained from the Anscombe test is less than 0.05, suggesting that the data is not normally distributed at a significance level of 0.05.

To visually assess the data for outliers and its adherence to a normal distribution, a box plot in Figure 1 was constructed. From the box plot, it can be observed that there are no outliers present in the dataset. However, the data does not exhibit a Normal distribution, as evidenced by the non-symmetrical shape of the box plot. Overall, the analysis indicates that the data is positively skewed, has a platykurtic distribution with thin tails, and does not follow a Normal distribution.

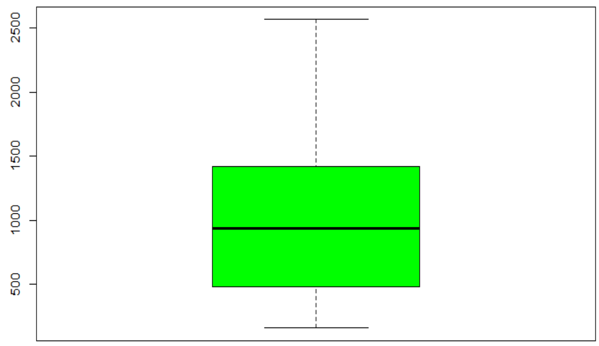

Figure 1. Box-plot for previous close price data.

Figure 2 displays a histogram plot of the dataset, providing a visual representation of the distribution of the data. Histograms divide the data into bins or intervals and display the frequency or count of data points falling into each bin. In this case, the histogram reveals important insights about the distribution of the dataset. From the histogram in Figure 2, it is evident that the data exhibits right skewness. This means that the data is skewed towards the higher values, with a longer tail on the right side of the distribution. The majority of the data points are concentrated towards the lower values, while there are fewer data points towards the higher values. Observing a right-skewed distribution in a financial stock chart provides insights into certain trends and characteristics of price movements. Right-skewness indicates that stock prices tend to exhibit an upward trend, with lower values being more common and higher values being relatively rare.

In this case, a right-skewed stock chart suggests that the stock prices have a tendency to increase over time, with occasional spikes or surges in value. This could be indicative of positive market sentiment and investor optimism surrounding the stock. Traders and investors may interpret this pattern as an opportunity for potential gains, as the stock shows a tendency to experience upward movements.

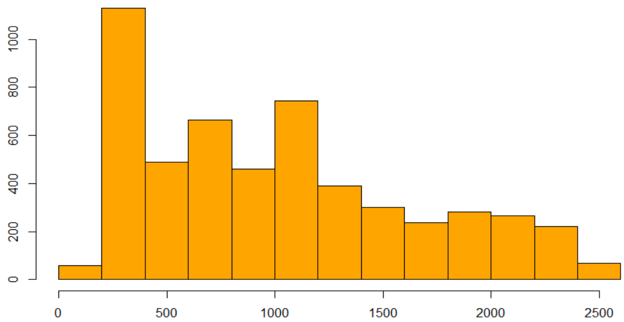

Figure 2. Histogram for the close price data.

Before selecting a distribution for the price data, a preliminary analysis was performed using the Cullen and Frey plot. The Cullen and Frey plot is a graphical tool that helps assess the skewness and kurtosis of a dataset. It is used to gain insights into the underlying distribution and determine potential candidate distributions. In the plot, the skewness of the data is represented on the x-axis, while the kurtosis is represented on the y-axis. The observed values are indicated by blue dots on the plot. The position of each dot on the plot provides information about the shape of the data [17]. Based on the Cullen and Frey plot in Figure 3, it can be inferred that the Weibull, Gamma, Log-normal, and beta distributions are potential candidates for fitting the data. This inference is made by examining the location of the blue dot in relation to the expected ranges for skewness and kurtosis associated with these distributions.

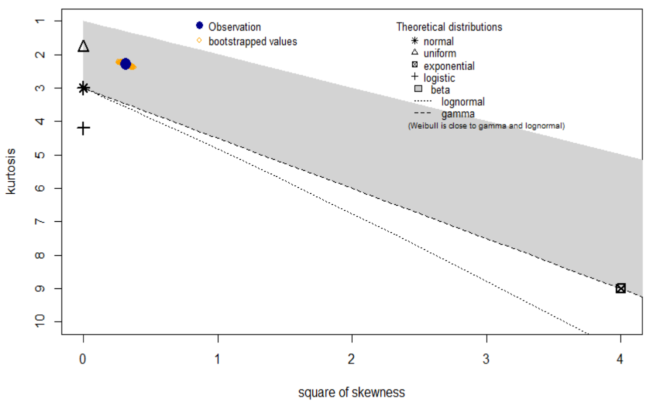

Figure 3. Cullen and Frey graph of competitive distributions for the close price data.

4. RESULTS

In this section, we assessed the suitability of alternative distributions for fitting the HDFC Bank closing price data through graphical evaluation. We employed the "fitdistrplus" package in R to examine the fit of various distributions, namely Normal, Log-normal, Weibull, and Gamma distributions. We generated histogram plots to visualize the distribution of the data. Additionally, we utilized quantile-quantile (Q-Q) plots to compare the observed quantiles of the data against the quantiles of the selected distributions. Moreover, we plotted the empirical cumulative distribution function (CDF) and the theoretical CDF of each distribution to evaluate the overall fit. Lastly, we examined probability-probability (P-P) plots to compare the observed probabilities against the theoretical probabilities derived from the fitted distributions. These graphical representations can be found in Figures 4-7.

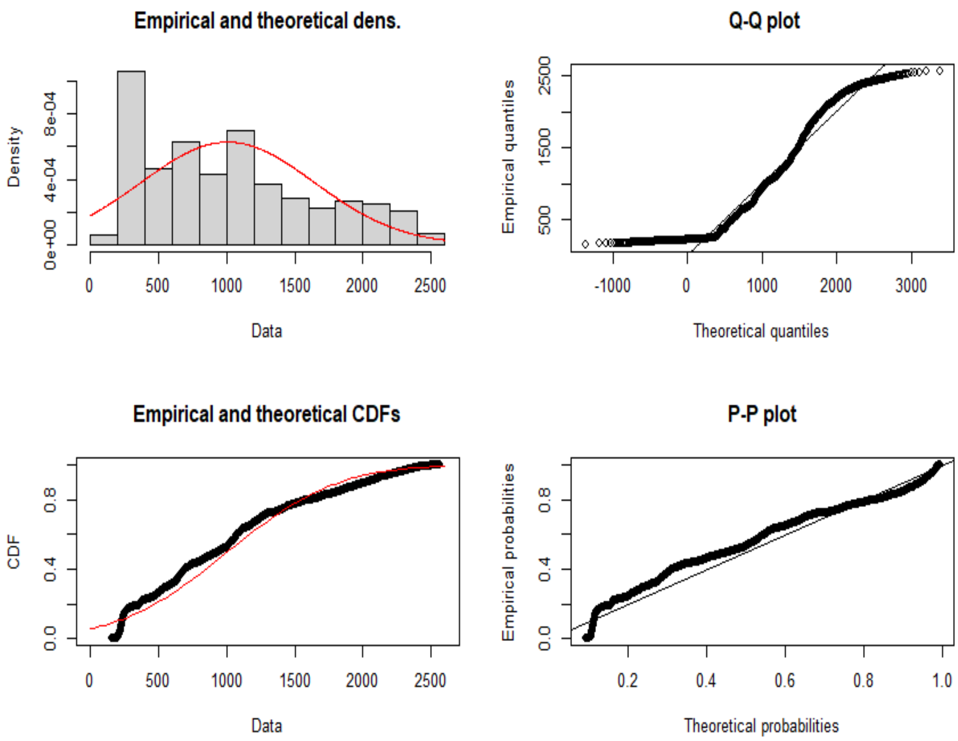

Figure 4. Graphs of goodness of fit of the Normal distribution for the close price data

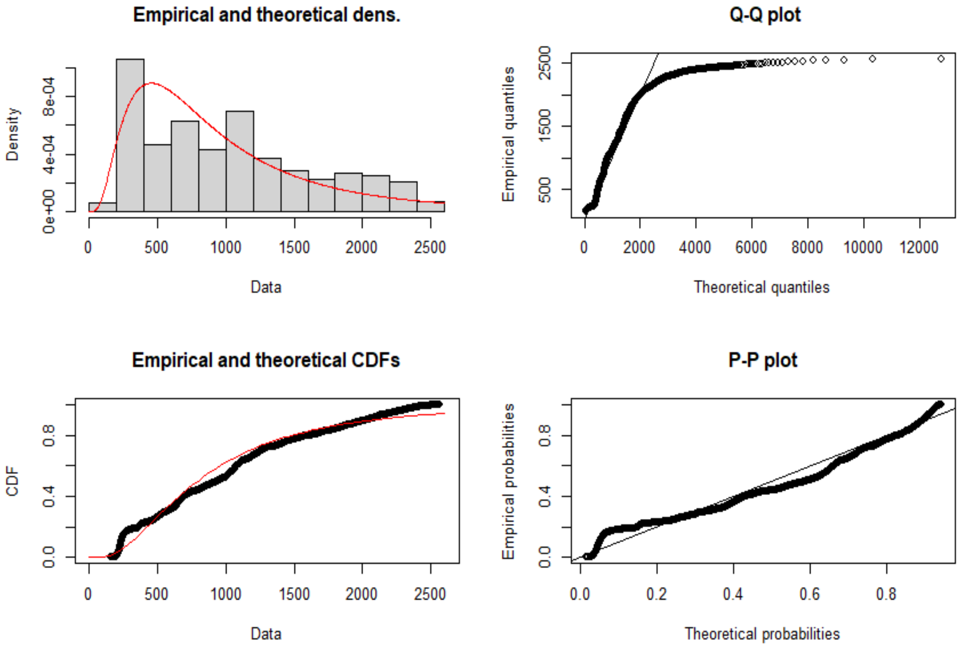

Figure 5. Graphs of goodness of fit of the Log-normal distribution for the close price data

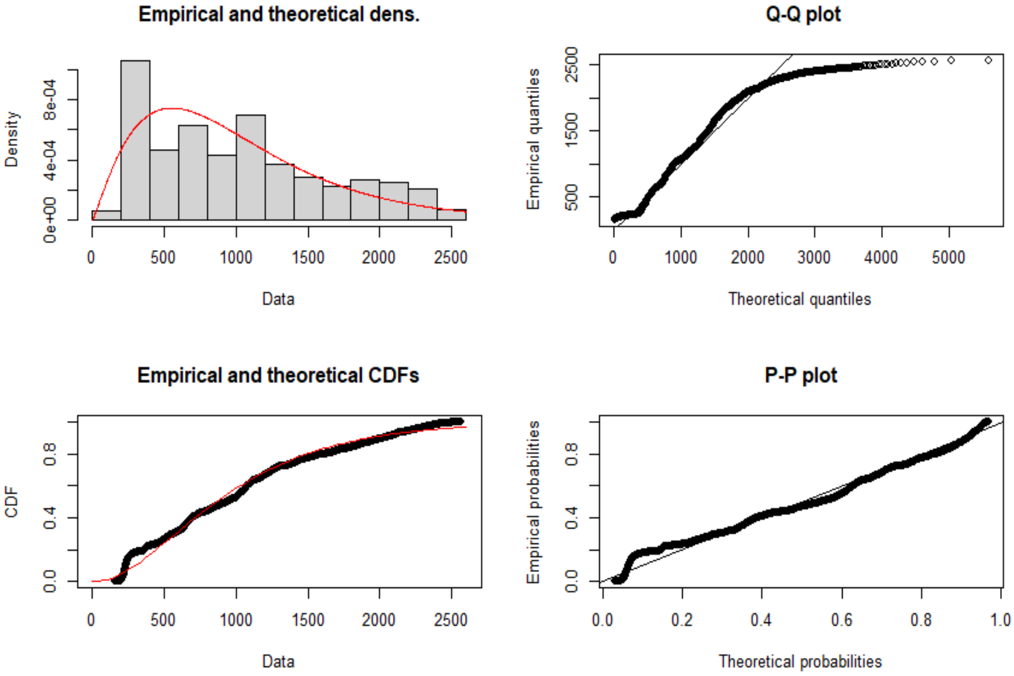

Figure 6. Graphs of goodness of fit of the Gamma distribution for the close price data

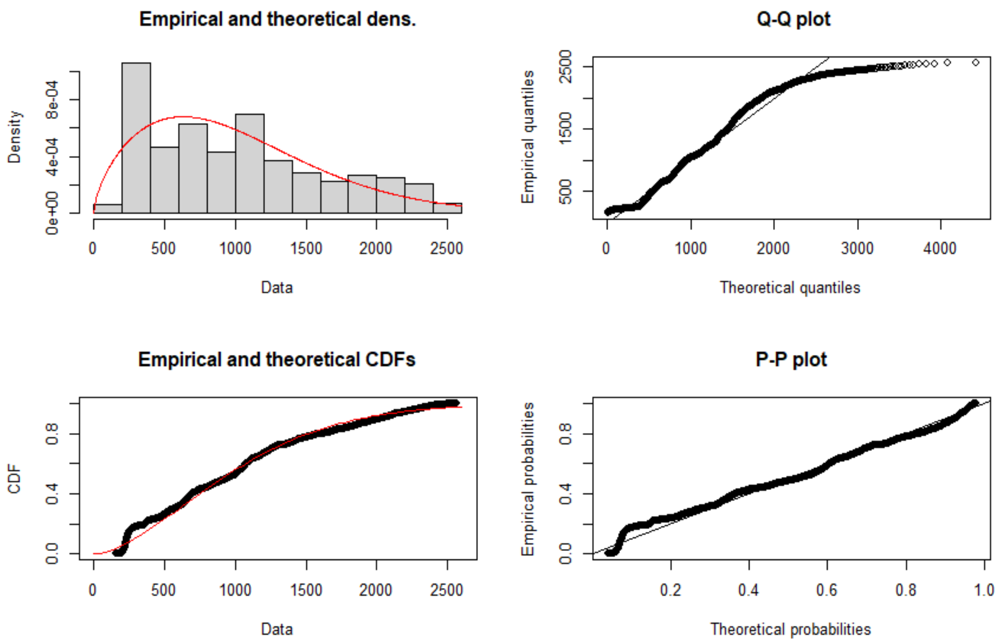

Figure 7. Graphs of goodness of fit of the Weibull distribution for the close price data

Upon examining the graphs presented in Figures 4 to 7, it can be observed that the Normal distribution does not fit the stock closing price data. Additionally, a careful examination of the P-P plot suggests that the Gamma and Weibull distributions in Figures 6 – 7 exhibit better fit to the stock closing data compared to the other distributions. Figure 4 shows the histogram of the data, where the shape of the distribution is not symmetrical and does not resemble a Normal distribution. This supports the finding that the Normal distribution is not suitable for modelling the stock closing prices. Figure 5 displays the Q-Q plot, which compares the quantiles of the observed data against the quantiles of the theoretical distribution. The deviations from the straight line indicate that the log Normal distribution is not a good fit for the data.

We also give the all distributions used in the study together in Figure 8 to compare them at the same time. Based on the graphical analysis in Figure 8, it is also evident that the Weibull and Gamma distributions provide better fits to the stock closing price data compared to the Normal distribution. These distributions can be considered as potential models for further analysis and modelling of the stock price behaviour.

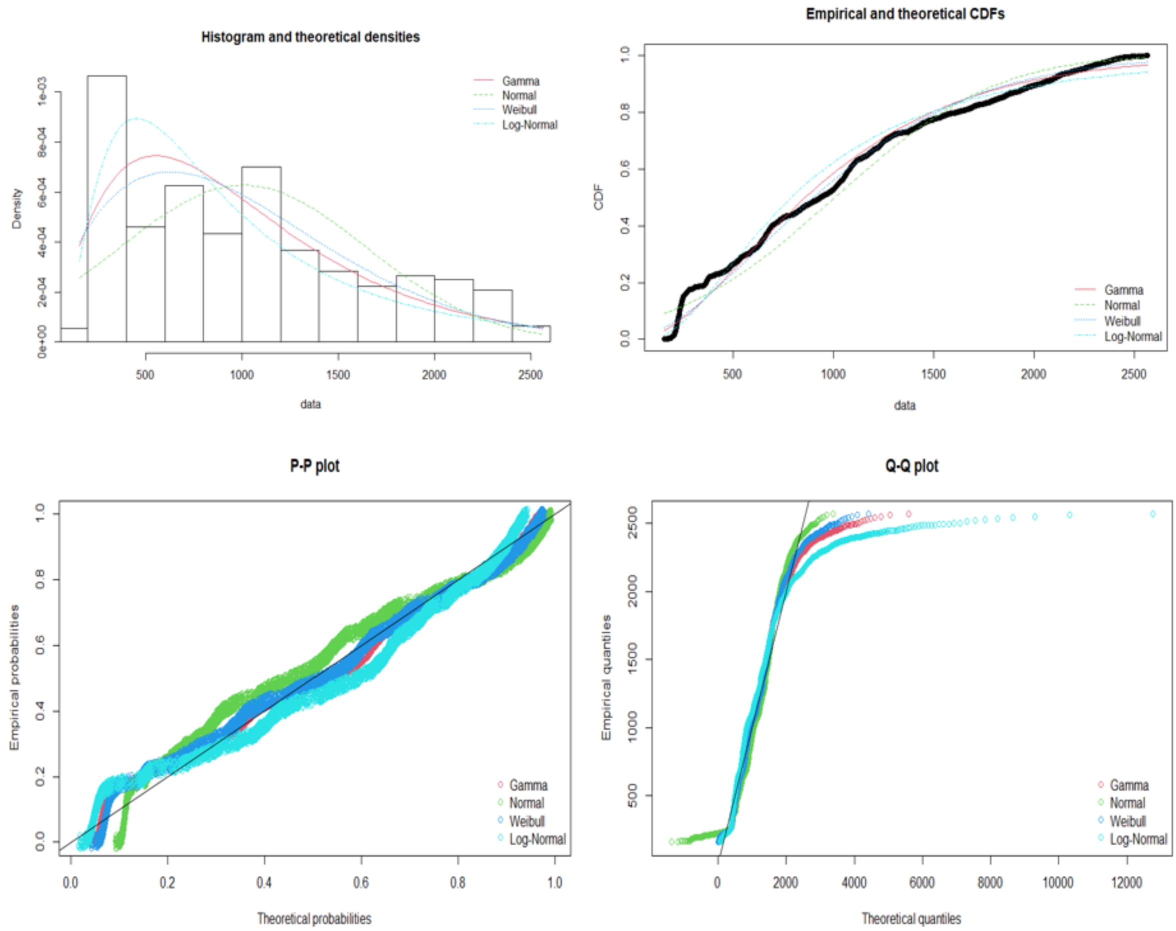

Figure 8. Plots of all competitive distributions for the close price data

To determine which alternative distribution is statistically better, Table 2 provides the AIC, BIC, and log-likelihood (log L) values. Additionally, the table includes the parameter values obtained using the maximum likelihood estimation method for the Normal, Log-normal, Exponential, Gamma, and Weibull distributions. Upon examining Table 2, it can be observed that the Weibull distribution has the smallest AIC and BIC values. This suggests that the Weibull distribution provides a better fit to the data compared to the other distributions. The log-likelihood value also supports this finding, as it is maximized for the Weibull distribution. These statistical criteria indicate that the Weibull distribution is the most suitable distribution for modelling the stock price data in terms of goodness-of-fit and model complexity. Therefore, the Weibull distribution can be considered as the preferred distribution for further analysis and modelling of the stock price behaviour.

Table 2. Parameter estimations with standard errors in parenthesis and the results of model selection criteria.

Model	a	b	logL	AIC	BIC
Normal	1007.0939 (8.7272)	635.6978 6.1708	-41777.65	83559.3	83572.45
Log-normal	6.67319 (0.0102)	0.7448 (0.0072)	-41374.18	82752.36	82765.51
Gamma	2.22138 (0.00243)	0.00220 (0.00001)	-41212.17	82428.34	82441.50
Weibull	1.631792 (0.01785)	1128.1181 (10.011)	-41189.58	82383.16	82396.32

The AIC and BIC values in Table 2 were close to each other, so non-parametric tests such as the Kolmogorov-Smirnov (KS), Cramer-von Mises (CVM), and Anderson-Darling (AD) tests were also employed to make a decision regarding the stock price data. The results of these tests are presented in Table 3. According to the results of these tests, it was determined that the Weibull distribution is the best fit for modelling the stock price data.

When applying the Weibull distribution to stock price data, the shape parameter (k=1.631792) can be useful in capturing the tail behaviour of the data. A higher value of the shape parameter indicates a heavier tail, indicating a higher likelihood of extreme price movements. On the other hand, a lower value of the shape parameter indicates a lighter tail, implying a lower likelihood of extreme events. The scale parameter (λ=1128.1181) influences the overall level or magnitude of the data. It determines the average or typical value around which the stock prices fluctuate. Adjusting the scale parameter can help align the distribution with the specific dataset being analysed.

Table 3. Results of KS, CVM and AD test for the close price data.

Model	**KS**	**CVM**	**AD**
Normal	0.0959	13.2550	98.272
Log-normal	0.0965	13.3609	99.417
Gamma	0.0826	6.1121	56.027
Weibull	0.0762	4.7887	47.812

These additional non-parametric tests provide further evidence supporting the selection of the Weibull distribution. The tests evaluate the goodness-of-fit between the empirical distribution of the stock price data and the theoretical Weibull distribution. Based on the results, it can be concluded that the Weibull distribution provides the most accurate representation of the stock price behaviour. Therefore, combining the findings from the AIC, BIC, and non-parametric tests, it can be confidently stated that the Weibull distribution is the optimal choice for modelling the stock price data in terms of both parametric and non-parametric approaches.

According to the KS test results for HDFC closing price data for the period 2023 to April 2024 (KS Statistic: 0.07198, p-value: 0.06628), because the p-value is greater than 0.05, we can conclude that the HDFC closing prices from 2023 to April 2024 fit the Weibull distribution. This suggests that the data is consistent with the Weibull distribution. Overall, the histogram and the Weibull fit curve in Figure 9 visually support the statistical conclusion that the closing prices of HDFC from 2023 to April 2024 follow a Weibull distribution.

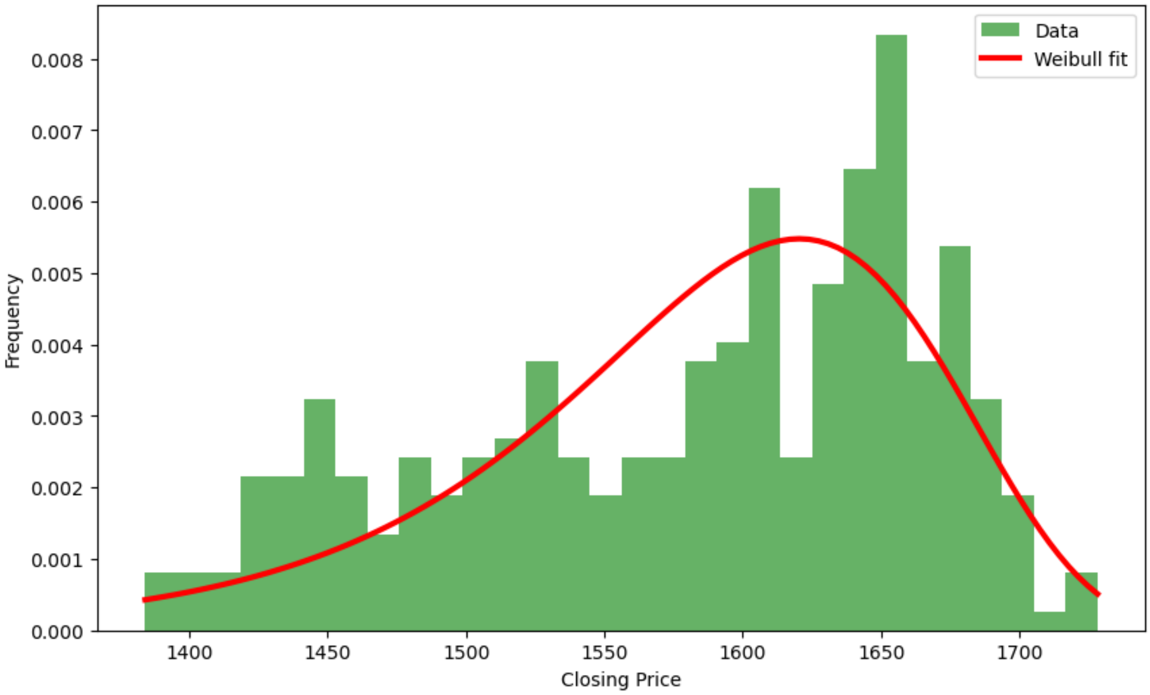

Figure 9. Weibull distribution fit to HDFC closing prices

5. CONCLUSION

Extreme value distributions play a crucial role in modelling stock price data, particularly when analysing tail behaviour and estimating tail probabilities. Extreme value distributions are designed to capture the behaviour of rare and extreme events, which occur with low probabilities but can have significant impacts on financial markets. These events include market crashes, large price fluctuations, and other extreme market conditions. In summary, extreme value distributions are valuable tools for understanding and quantifying the tail behaviour of stock price data. Their application in risk management, portfolio optimization, VaR estimation, tail risk hedging, and insurance highlights their importance in the financial industry.

The Weibull distribution is a probability distribution frequently used in modelling stock price data. It is a versatile distribution that can capture a wide range of shapes, including both heavy-tailed and light-tailed distributions. One advantage of the Weibull distribution is its flexibility in fitting a wide range of data shapes. By estimating the parameters of the Weibull distribution from historical stock price data, it is possible to model the future behaviour of stock prices, estimate tail probabilities, and assess the risk associated with extreme events. However, it is important to note that the choice of distribution should be based on the characteristics of the specific dataset and the research objectives. Other distributions, such as the Log-normal, Gamma, or generalized extreme value distributions, may also be appropriate depending on the characteristics of the stock price data being analysed.

Overall, the Weibull distribution can be a valuable tool for modelling stock price data and analysing tail behaviour, providing insights into the likelihood of extreme price movements and assisting in risk management and decision-making processes in the financial industry.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

CRediT AUTHOR STATEMENT

Ceren Ünal: Formal analysis, Visualization, Writing – Original Draft,
Gamze Özel Kadılar: Supervision, Formal analysis, Writing – Review & Editing

REFERENCES

[1] Bachelier L. Théorie de la spéculation. in annales scientifiques de l'école normale supérieure 1900; 17: 21-86, http://dx.doi.org/10.24033/asens.476.

[2] Ţiţan AG. The efficient market hypothesis: review of specialized literature and empirical research. Procedia Economics and Finance 2015; 32: 442-449.

[3] Masoliver J, Montero M, Porrà JM. A dynamical model describing stock market price distributions. Physica A: Statistical Mechanics and its Applications 2000; 233(3-4): 559-567.

[4] Cont R. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 2001; 1(2): 223–236.

[5] Mirsadeghpour Zoghi SM, Sanei M, Tohidi G, Banihashemi S, Modarresi N. The effect of underlying distribution of asset returns on efficiency in dea models. Journal of Intelligent & Fuzzy Systems 2021; 40(5): 10273-10283.

[6] Ahmad Z, Mahmoudi E, Dey S. A new family of heavy tailed distributions with an application to the heavy tailed insurance loss data. Communications in Statistics-Simulation and Computation 2022; 51(8): 4372-4395.

[7] Mandelbrot BB. The Variation of Certain Speculative Prices. Springer, New York, 1997.

[8] Yousof HM, Tashkandy Y, Emam W, Ali MM, Ibrahim M. A new reciprocal Weibull extension for modelling extreme values with risk analysis under insurance data. Mathematics 2023; 11(4): 966.

[9] Gebizlioglu OL, Şenoğlu B, Kantar YM. Comparison of certain value-at-risk estimation methods for the two-parameter Weibull loss distribution. Journal of Computational and Applied Mathematics 2011; 235(11): 3304-3314.

[10] Kowan T, Mahalapkorkiat C, Samakrat T, Sumritnorrapong P. Probability distributions for modelling of covid-19 cases and deaths in Thailand. Computer Science 2022; 17(4): 1499-1506.

[11] Campbell JY, Lo AW, MacKinlay AC. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ, 1997.

[12] Bhaumik DK, Kapur K, Gibbons RD. Testing parameters of a gamma distribution for small samples. Technometrics 2009; 51(3): 326-334.

[13] Tong YL. The Multivariate Normal Distribution. Springer Science & Business Media, 2012.

[14] Weibull W. A statistical distribution function of wide applicability. Journal of Applied Mechanics 1951; 18: 293-297.

[15] Almalki SJ, Nadarajah S. Modifications of the Weibull distribution: a review. Reliability Engineering and System Safety 2014; 124, 32-55.

[16] Yahoo Finance (2023, Jun 10). [Online] Available: https://finance.yahoo.com

[17] Cullen AC, Frey HC, Frey CH. Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs. Springer Science & Business, 1999.