Mathematics

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

AIMS Mathematics, 9(1): 2345–2368.
DOI: 10.3934/math.2024116
Received: 17 September 2023
Revised: 06 December 2023
Accepted: 14 December 2023
Published: 22 December 2023

Research article

Asset selection based on estimating stress-strength probabilities: The case of returns following three-parameter generalized extreme value distributions

Felipe S. Quintino[1], Melquisadec Oliveira[1], Pushpa N. Rathie[1], Luan C. S. M. Ozelim[2,*] and Tiago A. da Fonseca[3]

[1] Department of Statistics, University of Brasília, Brasília, DF 70.910-900, Brazil

[2] Department of Civil and Environmental Engineering, University of Brasília, Brasília, DF 70.910-900, Brazil

[3] Gama Engineering College, University of Brasília, Brasília, DF 72.444-240, Brazil

* **Correspondence:** Email: luanoz@gmail.com.

Abstract: Analyzing the statistical behavior of the assets' returns has shown to be an interesting approach to perform asset selection. In this work, we explore a stress-strength reliability approach to perform asset selection based on probabilities of the type $P(X < Y)$ when both X and Y follow a generalized extreme value (GEV) distribution with three parameters. At first, we derive new analytical and closed form relations in terms of the extreme value $\mathbb{H}$-function, which have been obtained under fewer parameter restrictions compared to similar results in the literature. To show the performance of our results, we include a Monte-Carlo simulation study and we investigate the application of the reliability measure $P(X < Y)$ in selecting financial assets with returns characterized by the distributions X and Y. Therefore, rather than the conventional approach of comparing the expected values of X and Y based on modern portfolio theory, we delve into the metric $P(X < Y)$ as an alternative parameter for assessing better returns.

Keywords: stress-strength reliability; $\mathbb{H}$-functions; generalized extreme value distribution
Mathematics Subject Classification: 91-10, 62N05, 33C60, 33F05

1. Introdution

The work of Markowitz [1] stands as one of the pioneering works on portfolio theory [2] and details Markowitz's major contributions to the rising modern portfolio theory. Besides providing criterious description of Markowitz's model of portfolio choice [3], the work in [1] traced future research directions explored by other scientists like the suggestion to replace variance with semi-variance as

a measure of risk, recommendations on the maximization of the expected logarithmic utility of return, and the outlining of a market model developed in depth by Sharpe [4].

However, Markowitz's portfolio selection framework relies on the premise that one can measure asset's return variance as emphasized by the statement that "examples of rapidly increasing variances are of mostly academic interest" [1]. This restriction brings concerns related to the framework suitability in a scenario where heavy-tailed distributions better model assets return by allowing fast varying volatilities resulting from extreme events. We can find evidences in literature that financial data is better modelled by α-stable processes (heavy-tailed alternative to Brownian motion [5]) or by heavy-tailed time series models [6, 7]. Although in this work we rely on the general hypothesis that logarithmic returns in financial data follow an α-stable process with parameter $0 < \alpha < 2$ [8] (which implies that variance of return is undefined), without loss of generality, we use results from extreme value theory (EVT) regarding generalized extreme value (GEV) distribution [9] as an alternative to α-stable distributions. This approach can be considered valid since the GEV distribution has fat-tailed behaviour and can be used as a proxy of various fat-tailed distributions.

From an economic point of view, it is well known that extreme share returns on stock markets can have important implications for financial risk management and several studies have successfully applied GEV to model financial data [10]. For example, Gettinby et al. [11] characterized the distribution of extreme returns for a UK share index over the years 1975 to 2000. They considered the suitability of several distributions, being the weekly maxima and minima of daily return best modelled by the GEV and the Generalised Logistic distributions. For the UK case, Generalised Logistic was a better choice overall. On the other hand, GEV has presented a similar modelling capability as well as presenting some important properties due to EVT. Also, Hussain & Li [12] studied the distribution of the extreme daily returns of the Shanghai Stock Exchange (SSE) Composite Index. They modelled the SSE Composite index returns based on the data from 1991 to 2013, which indicated that the Generalized Logistic distribution is a better fit for the minima series and that the GEV distribution is a better fit for the maxima series of the returns for the Chinese stock market.

EVT is a branch of probability and statistics that deals with the modeling of extreme events that are related to maximums and minimums of independent random samples. Applications of this theory are found in finance [13], natural catastrophes, and equipment failures, among others. The books [6,14–16] provide extensive coverage that allows for a detailed study of EVT.

Furthermore, EVT provides a theoretical basis and framework to deal with extreme deviations from the mean of distribution functions (DFs) by restricting the behavior of the DFs in the tails. It focuses on the study of the possible limiting distributions and their properties for the normalized maximum.

Specifically, let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed (*i.i.d.*) random variables (RVs) with common distribution function F and set $M_n = \max\{X_1, \cdots, X_n\}$. The theory is concerned with properties of F and of the possible non-degenerate distribution functions G satisfying

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \quad \forall x \in C(G), \tag{1.1}$$

for sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ $(n = 1, 2, \cdots)$, suitably chosen, where $C(G)$ denotes the set of continuity points of G.

The possible distribution functions G satisfying (1.1) have been known for some time [17] and have been extensively studied by several authors from then on. They are also known as *max-stable laws* (or

max-stable distributions) and can only be of three well-known types: Fréchet, Weibull or Gumbel.

Goncu et al. [18] used the EVT to model the extreme return behaviour of the Istanbul Stock Exchange (ISE), Turkey. They considered Gumbel, Fréchet and Weibull distributions for modelling extreme returns over different investment horizons. Their results indicate that when the Value at Risk (VaR) is computed with the proposed distributions, backtesting results indicate that the EVT provides superior risk management in all the sub-intervals considered compared to the VaR estimation under the assumption of a normal distribution.

For statistical applications, the max-stable distributions can be summarized in a single distribution function called *generalized extreme value (GEV) distribution*. Essentially, the GEV distribution has the cumulative distribution function (CDF) given by

$$G(x) = \exp\left\{-(1+\gamma x)^{\frac{-1}{\gamma}}\right\}, \quad 1+\gamma x > 0, \quad \gamma \in \mathbb{R}, \tag{1.2}$$

where γ is the shape parameter.

Our interest in this work is the property of stress-strength probability which, in general terms, consists of the study of the probability of failure of a system or component based on the comparison of the applied stress to the strength of the system. Let stress Y and strength X be independent continuous RVs with probability density function (PDF) f_Y and CDF F_X, respectively. The stress-strength probability (or reliability) is defined as

$$R = P(X < Y) = \int_{-\infty}^{\infty} F_X(x) f_Y(x) dx. \tag{1.3}$$

There are several applications of this theory such as in engineering and manufacturing, aerospace and defense, automotive industry, energy sector, healthcare, and electronics, among others. See [19] for more details.

The stress-strength reliability framework is versatile and finds various applications in economics. Besides this contribution regarding financial data, previous works researched economic inequality [19, 20]. Regarding financial data, the authors explored stress-strength reliability framework in [21–24].

The stress-strength probability for the extreme Fréchet, Weibull, and Gumbel distributions has been widely studied in the literature. Nadarajah [25] considered the class of extreme value distributions and derived the corresponding forms for the reliability R in terms of special functions. Confidence limits for R involving Weibull models were presented in [26]. Kundu & Raqab [27] proposed a modified maximum likelihood estimator of R and obtained the asymptotic distribution of the modified maximum likelihood estimators, which was used to construct the confidence interval of R. The previous results of R for Weibull distribution were generalized by Nojosa & Rathie [28], where R was expressed in terms of H-functions. Bayesian estimation of R for Fréchet and Weibull distributions has also been explored [29, 30].

The goal of this paper is to present an asset selection approach based on the probability $R = P(X < Y)$ when both X and Y present the distributions of returns of two assets. In particular, we seek to derive an expression of R when X and Y have three-parameter GEV distributions and to propose an estimation procedure of R by not using transformations in the data and with as fewer parameter restrictions as possible.

The paper is organized as follows: In Section 2, we define the H-function, $\mathbb{H}$-function and the three-parameter GEV distribution. Section 3 deals with the derivation of R when X and Y are independent GEV RVs. The maximum likelihood estimation for R is presented in Section 4. In Section 5, we present Monte-Carlo simulations for the estimation of R and also deal with two real situations involving log-returns of stock prices and different-length carbon fibers. The last section deals with the conclusions and the Appendix presents the correlation matrices of the data set modeled in Section 5.

2. Preliminaries

In this section, we present definitions and results on which our contributions are based.

2.1. Special functions

Recently, Rathie et al. [31] introduced the extreme-value $\mathbb{H}$-function as:

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5, a_6) := \int_0^\infty y^{a_6} \exp\{-a_1 y - (a_2 y^{a_3} + a_4)^{a_5}\} dy, \tag{2.1}$$

where $\Re(a_1), \Re(a_2), \Re(a_4) \in \mathbb{R}_+, a_3, a_5 \in \mathbb{C}$, not both $\Re(a_1)$ and $\Re(a_2)$ can be equal to zero at the same time, $\Re(a_6) > -1$ when $a_1 \neq 0$ or $a_1 = 0$ and $\text{sign}(a_3) = \text{sign}(a_5)$, $\Re(a_6) < -1$ when $a_1 = 0$ and $\text{sign}(a_3) \neq \text{sign}(a_5)$. In this paper, $\mathbb{R}$, $\mathbb{C}$ and $\Re$ denote the real numbers, complex numbers and the real part of a complex number, respectively.

In this work, we are interested in the case $a_6 = 0$. Thus, we omit such a parameter from the representation and denote only:

$$\mathbb{H}(a_1, a_2, a_3, a_4, a_5) := \int_0^\infty \exp\{-a_1 y - (a_2 y^{a_3} + a_4)^{a_5}\} dy. \tag{2.2}$$

In the following sections, we prove that all stress-strength probabilities involving three-parameter GEV distribution with shape parameters of equal sign can be written as $\mathbb{H}$-functions.

Note that (2.1) generalizes some important cases of the H-function (cf. [32]) defined by

$$H_{p,q}^{m,n}\left[z\middle| \begin{array}{c} (a_1, A_1), \cdots, (a_p, A_p) \\ (b_1, B_1), \cdots, (b_q, B_q) \end{array}\right] = \frac{1}{2\pi i} \int_L \frac{\prod_{k=1}^m \Gamma(b_j + B_j s) \prod_{j=1}^n \Gamma(1 - a_j - A_j s)}{\prod_{k=m+1}^q \Gamma(1 - b_j - B_j s) \prod_{j=n+1}^p \Gamma(a_j + A_j s)} z^{-s} ds, \tag{2.3}$$

where $0 \leq m \leq q$, $0 \leq n \leq p$ (not both m and n simultaneously zero), $A_j > 0$ $(j = 1, \cdots, p)$, $B_k > 0$ $(k = 1, \cdots, q)$, a_j and b_k are complex numbers such that no poles of $\Gamma(b_k + B_k s)$ $(k = 1, \cdots, m)$ coincide with poles of $\Gamma(1 - a_j - A_j s)$ $(j = 1, \cdots, n)$. L is a suitable contour $w - i\infty$ to $w + i\infty$, $w \in \mathbb{R}$, separating the poles of the two types mentioned above. For more details, see [32].

An important special case of this function is obtained by taking $a_4 = 0$, which represents an upper (or lower) bound for its value depending on the sign of a_5. This case is, therefore, an extreme value of the function and can be written in terms of the H-function as [31]:

$$\begin{aligned}
\mathbb{H}(a_1, a_2, a_3, 0, a_5, a_6) &= \int_0^\infty y^{a_6} \exp\{-a_1 y - a_2^{a_5} y^{a_3 a_5}\} dy \\
&= \frac{1}{a_2^{(1+a_6)/a_3} a_3 a_5} H_{1,1}^{1,1}\left[a_1 a_2^{-1/a_3} \middle| \begin{matrix} (1 - \frac{(1+a_6)}{a_3 a_5}, \frac{1}{a_3 a_5}) \\ (0, 1) \end{matrix}\right] \\
&= \frac{1}{a_1^{a_6+1}} H_{1,1}^{1,1}\left[\left(\frac{a_2}{a_1^{a_3}}\right)^{a_5} \middle| \begin{matrix} (-a_6, a_3 a_5) \\ (0, 1) \end{matrix}\right], \qquad (2.4)
\end{aligned}$$

when $\text{sign}(a_3) = \text{sign}(a_5)$ and:

$$\begin{aligned}
\mathbb{H}(a_1, a_2, a_3, 0, a_5, a_6) &= \frac{1}{a_2^{(1+a_6)/a_3} |a_3 a_5|} H_{0,2}^{2,0}\left[a_1 a_2^{-1/a_3} \middle| \begin{matrix} - \\ (0, 1), (\frac{(1+a_6)}{a_3 a_5}, \frac{1}{|a_3 a_5|}) \end{matrix}\right] \\
&= \frac{1}{a_1^{a_6+1}} H_{0,2}^{2,0}\left[\left(\frac{a_2}{a_1^{a_3}}\right)^{a_5} \middle| \begin{matrix} - \\ (0, 1), (1 + a_6, |a_3 a_5|) \end{matrix}\right], \qquad (2.5)
\end{aligned}$$

otherwise.

2.2. *Three-parameter GEV distribution*

The three-parameter GEV distribution is obtained by taking CDF of the same type of the standard GEV G defined in (1.2). That means, $G(x; \mu, \sigma, \gamma) = G\left(\frac{x-\mu}{\sigma}\right)$. We denote by $X \sim GEV(\mu, \sigma, \gamma), \mu, \gamma \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$, an RV with CDF given by

$$G(x; \mu, \sigma, \gamma) = \exp\left\{-\left[1 + \frac{\gamma}{\sigma}(x - \mu)\right]^{-\frac{1}{\gamma}}\right\}, \quad 1 + \frac{\gamma}{\sigma}(x - \mu) > 0, \qquad (2.6)$$

where μ is the location parameter, σ is the scale parameter and γ is the shape parameter. The corresponding probability density function (PDF) is given by

$$g(x; \mu, \sigma, \gamma) = G(x_i, \gamma_i, \mu, \sigma)\frac{1}{\sigma}\left[1 + \frac{\gamma}{\sigma}(x - \mu)\right]^{-\frac{1}{\gamma}-1}, \quad 1 + \frac{\gamma}{\sigma}(x - \mu) > 0. \qquad (2.7)$$

Figure 1 shows the behavior of g for some parameter choices. Note that the location parameter shifts the curve, the scale controls dispersion, and the density changes according to the sign of the shape parameter.

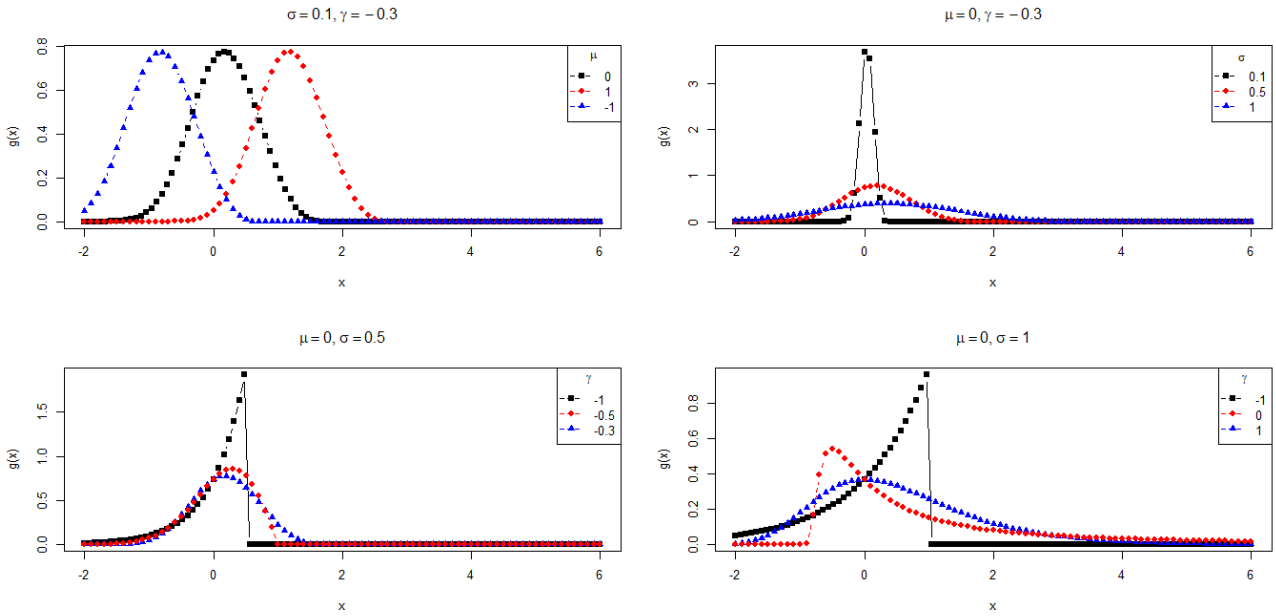


Figure 1. Plot for the PDF g.

3. Main results

In this section, the reliability of two independent three-parameter GEV RVs is derived in terms of the $\mathbb{H}$-function. In addition, with suitable parameter restrictions, representations of R as an H-function and an explicit form are obtained. We consider the case of two independent GEV distributions with different shape parameters (but of the same sign). Cases of opposite signs of shape parameters are not normally of interest, as it would indicate that random variables with incompatible support are being compared. Therefore, these cases are not treated in the present paper.

Theorem 3.1. *Let Y and X be independent RVs, respectively, with distribution $GEV(\mu_1, \sigma_1, \gamma_1)$ and $GEV(\mu_2, \sigma_2, \gamma_2)$, $\mu \in \mathbb{R}$, $\sigma_j \in \mathbb{R}_+$, $\gamma_j \in \mathbb{R}$ $(\gamma_j \neq 0)$, $j = 1, 2$. Then*

- *When $\gamma_j > 0$, $j = 1, 2$:*

$$R = P(X < Y) = \mathbb{H}\left(1, \frac{\gamma_2\sigma_1}{\sigma_2\gamma_1}, -\gamma_1, 1 + \frac{\gamma_2}{\sigma_2}\left(\mu_1 - \mu_2 - \frac{\sigma_1}{\gamma_1}\right), -\frac{1}{\gamma_2}\right), \tag{3.1}$$

provided that $\mu_1 - \frac{\sigma_1}{\gamma_1} \geq \mu_2 - \frac{\sigma_2}{\gamma_2}$. When $\mu_1 - \frac{\sigma_1}{\gamma_1} \leq \mu_2 - \frac{\sigma_2}{\gamma_2}$:

$$R = P(X < Y) = 1 - \mathbb{H}\left(1, \frac{\gamma_1\sigma_2}{\sigma_1\gamma_2}, -\gamma_2, 1 + \frac{\gamma_1}{\sigma_1}\left(\mu_2 - \mu_1 - \frac{\sigma_2}{\gamma_2}\right), -\frac{1}{\gamma_1}\right). \tag{3.2}$$

- *When $\gamma_j < 0$, $j = 1, 2$:*

$$R = P(X < Y) = \mathbb{H}\left(1, \frac{\gamma_2\sigma_1}{\sigma_2\gamma_1}, -\gamma_1, 1 + \frac{\gamma_2}{\sigma_2}\left(\mu_1 - \mu_2 - \frac{\sigma_1}{\gamma_1}\right), -\frac{1}{\gamma_2}\right), \tag{3.3}$$

provided that $\mu_1 - \frac{\sigma_1}{\gamma_1} \leq \mu_2 - \frac{\sigma_2}{\gamma_2}$. When $\mu_1 - \frac{\sigma_1}{\gamma_1} \geq \mu_2 - \frac{\sigma_2}{\gamma_2}$:

$$R = P(X < Y) = 1 - \mathbb{H}\left(1, \frac{\gamma_1\sigma_2}{\sigma_1\gamma_2}, -\gamma_2, 1 + \frac{\gamma_1}{\sigma_1}\left(\mu_2 - \mu_1 - \frac{\sigma_2}{\gamma_2}\right), -\frac{1}{\gamma_1}\right). \tag{3.4}$$

In particular, if $\mu_1 - \frac{\sigma_1}{\gamma_1} = \mu_2 - \frac{\sigma_2}{\gamma_2}$, *we have*

$$R = \frac{\gamma_2}{\gamma_1}\left(\frac{\gamma_2\sigma_1}{\gamma_1\sigma_2}\right)^{1/\gamma_1} H_{1,1}^{1,1}\left[\left(\frac{\gamma_2\sigma_1}{\gamma_1\sigma_2}\right)^{1/\gamma_1} \Bigg| \begin{matrix} (\frac{\gamma_1-\gamma_2}{\gamma_1}, \frac{\gamma_2}{\gamma_1}) \\ (0,1) \end{matrix}\right]. \tag{3.5}$$

Proof. Set $\mu_j \in \mathbb{R}$, $\sigma_j, \gamma_j \in \mathbb{R}_+$ $(j = 1, 2)$. Then

$$\begin{aligned} R &= P(X < Y) = \int_{-\infty}^{\infty} G(x;\mu_2,\sigma_2,\gamma_2)g(x;\mu_1,\sigma_1,\gamma_1)dx \\ &= \int_M^{+\infty} \exp\left\{-[1+\frac{\gamma_2}{\sigma_2}(x-\mu_2)]^{-\frac{1}{\gamma_2}} - [1+\frac{\gamma_1}{\sigma_1}(x-\mu_1)]^{-\frac{1}{\gamma_1}}\right\}[1+\frac{\gamma_1}{\sigma_1}(x-\mu_1)]^{-\frac{1}{\gamma_1}-1}\frac{dx}{\sigma_1}, \end{aligned} \tag{3.6}$$

where $M = \max\left\{\mu_1 - \frac{\sigma_1}{\gamma_1}, \mu_2 - \frac{\sigma_2}{\gamma_2}\right\}$. Substituting $y = \left[1 + \frac{\gamma_1}{\sigma_1}(x-\mu_1)\right]^{-\frac{1}{\gamma_1}}$ and taking $M = \mu_1 - \frac{\sigma_1}{\gamma_1}$, we can rewrite (3.6) as

$$R = \int_0^{+\infty} \exp\left\{-y - \left[1 + \frac{\gamma_2}{\sigma_2}\left(\mu_1 - \mu_2 - \frac{\sigma_1}{\gamma_1}\right) + \frac{\gamma_2}{\sigma_2}\frac{\sigma_1}{\gamma_1}y^{-\gamma_1}\right]^{-1/\gamma_2}\right\}dy. \tag{3.7}$$

Hence, (3.1) follows from (2.1) and (3.7). For the case where $\gamma_j > 0$, $j = 1, 2$ and $\mu_1 - \frac{\sigma_1}{\gamma_1} \leq \mu_2 - \frac{\sigma_2}{\gamma_2}$, it suffices to notice that $P(X < Y) = 1 - P(Y < X)$ and the result in (3.1) is applied with interchanged sub-indices. For the cases where $\gamma_j < 0$, $j = 1, 2$, the same rationale can be applied, just noticing that in such cases x mostly takes negative values. The case where $\gamma_j = 0$, $j = 1, 2$, can be obtained as a limiting procedure and shall be explicitly explored later on the present paper. In addition, applying (2.4) with $\mu_1 - \frac{\sigma_1}{\gamma_1} = \mu_2 - \frac{\sigma_2}{\gamma_2}$, we obtain (3.5). □

Remark 3.2. *In a practical scenario, the estimates* $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\gamma}_1, \hat{\mu}_2, \hat{\sigma}_2, \hat{\gamma}_2)$ *should be obtained. Then, if* $sign(\hat{\gamma}_1) = sign(\hat{\gamma}_2)$, *the conditions* $\mu_1 - \frac{\sigma_1}{\gamma_1} \geq \mu_2 - \frac{\sigma_2}{\gamma_2}$ *or* $\mu_1 - \frac{\sigma_1}{\gamma_1} \leq \mu_2 - \frac{\sigma_2}{\gamma_2}$ *must be verified and the corresponding R expression should be used.*

Next, we consider some special cases of two independent GEV random variables. We have the following immediate consequence of Theorem 3.1:

Corollary 3.3. *Let* $Y \sim GEV(\mu_1, \sigma_1, \gamma_1)$ *and* $X \sim GEV(\mu_2, \sigma_2, \gamma_2)$ *be independent RVs, with* $\gamma_1 = \gamma_2 = \gamma \in \mathbb{R}, \gamma \neq 0$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_+$ *and* $\mu_1 - \mu_2 = \frac{\sigma_1}{\gamma} - \frac{\sigma_2}{\gamma}$. *Then, we have*

$$R = P(X < Y) = \left(\frac{\sigma_1^{1/\gamma}}{\sigma_1^{1/\gamma} + \sigma_2^{1/\gamma}}\right). \tag{3.8}$$

Lastly, we consider the cases of two independent GEV with $\gamma_1 = \gamma_2 = 0$.

Theorem 3.4. *Let Y and X be independent RVs, respectively, with distribution* $GEV(\mu_1, \sigma_1, 0)$ *and* $GEV(\mu_2, \sigma_2, 0)$, $\mu_j \in \mathbb{R}$, $\sigma_j \in \mathbb{R}_+$, $j = 1, 2$. *Then*

$$\begin{aligned} R &= \exp\left\{\frac{\mu_1}{\sigma_1}\right\}\mathbb{H}\left(\exp\left\{\frac{\mu_1}{\sigma_1}\right\}, \exp\left\{\frac{\mu_2}{\sigma_2}\right\}, \frac{\sigma_1}{\sigma_2}, 0, 1\right) \\ &= \exp\left\{\frac{\mu_1-\mu_2}{\sigma_1}\right\}\frac{\sigma_2}{\sigma_1}H_{1,1}^{1,1}\left[\exp\left\{\frac{\mu_1-\mu_2}{\sigma_1}\right\} \Bigg| \begin{matrix} (\frac{\sigma_1-\sigma_2}{\sigma_1}, \frac{\sigma_2}{\sigma_1}) \\ (0,1) \end{matrix}\right]. \end{aligned} \tag{3.9}$$

Proof. Set $\mu_j \in \mathbb{R}$ and $\sigma_j \in \mathbb{R}_+$ $(j = 1, 2)$. Then

$$\begin{aligned} R &= P(X < Y) = \int_{-\infty}^{\infty} G(x; \mu_2, \sigma_2, 0) g(x; \mu_1, \sigma_1, 0) dx \\ &= \int_{-\infty}^{\infty} \exp\left\{-\exp\left\{-\frac{x-\mu_2}{\sigma_2}\right\} - \exp\left\{-\frac{x-\mu_1}{\sigma_1}\right\}\right\} \exp\left\{-\frac{x-\mu_1}{\sigma_1}\right\} \frac{dx}{\sigma_1}. \end{aligned} \tag{3.10}$$

Substituting $y = \exp\{-x/\sigma_1\}$, we can rewrite (3.10) as

$$R = \exp\left\{\frac{\mu_1}{\sigma_1}\right\} \int_0^{+\infty} \exp\left\{-\exp\left\{\frac{\mu_1}{\sigma_1}\right\} y - \exp\left\{\frac{\mu_2}{\sigma_2}\right\} y^{\sigma_1/\sigma_2}\right\} dy. \tag{3.11}$$

Hence, (3.9) follows from (2.1) and (3.11).

□

We have the immediate consequence of Theorem 3.4.

Corollary 3.5. *Let $Y \sim GEV(\mu_1, \sigma_1, 0)$ and $X \sim GEV(\mu_2, \sigma_2, 0)$ be independent RVs, with $\sigma_1 = \sigma_2 = \sigma \in \mathbb{R}_+$. Then, we have*

$$R = P(X < Y) = \frac{\exp\{\mu_1/\sigma\}}{\exp\{\mu_1/\sigma\} + \exp\{\mu_2/\sigma\}}. \tag{3.12}$$

The results presented in Theorems 3.1 and 3.4 are more general than that presented in the literature. The $\mathbb{H}$-function allows us to write the probability R with as little parameter restrictions as possible. Table 1 lists related studies and their parameter restrictions.

Table 1. Extreme distributions and related studies of stress-strength probability.

$sign(\gamma)$	Distribution	Reference	Parameter restriction
0	Gumbel	[25]	$\sigma_1 = \sigma_2$ or $\sigma_1 = 2\sigma_2$ or $\sigma_2/\sigma_1 > 1$
1	Fréchet	[25]	$\mu_1 = \mu_2$ and ($\gamma_1 = \gamma_2$ or $\gamma_2 = 2\gamma_1$ or $\gamma_2/\gamma_1 = p/q$)*
		[30]	$\mu_1 = \mu_2 = 0$
		[29]	$\mu_1 = \mu_2 = 0$ and $\gamma_1 = \gamma_2$
		[33]	$\mu_1 = \mu_2 = 0$
-1	Weibull (min)	[25]	$\mu_1 = \mu_2$ and ($\gamma_1 = \gamma_2$ or $\gamma_2 = 2\gamma_1$ or $\gamma_2/\gamma_1 = p/q$)*
		[26]	$\mu_1 = \mu_2 = 0$ and $\gamma_1 = \gamma_2 = \gamma$
		[27]	$\mu_1 = \mu_2 = \mu$ and $\gamma_1 = \gamma_2 = \gamma$
		[28]	$\mu_1 = \mu_2 = 0$

*p and q are coprime integers.

Remark 3.6. *The particular case of GEV while $sign(\gamma) = -1$ (cf. [14]) is called reversed Weibull. The Weibull distribution studied by the authors cited in Table 1 is obtained as the limit of a normalized minimum of i.i.d. RVs. That is, the Weibull distribution is obtained by*

$$\lim_{n\to\infty} P\left(\frac{\min\{X_1,\cdots,X_n\}-b_n}{a_n}\le x\right) = \lim_{n\to\infty}(1-(1-F(a_nx+b_n))^n)$$
$$= 1-\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{\gamma}\right\},\quad x\ge\mu, \tag{3.13}$$

where $X_1,\cdots,X_n$ are i.i.d. RVs of F and a_n and b_n are suitable sequences of constants (see Theorem 2.1.5 in [14]).

3.1. Multicomponent system reliability

Let $X_1,\cdots,X_n$ be i.i.d. RVs with distribution $GEV(\mu_2,\sigma_2,\gamma_2)$ and Y be an independent RV with distribution $GEV(\mu_1,\sigma_1,\gamma_1)$. Set $M_n=\max\{X_1,\cdots,X_n\}$. Then, $P(M_n\le u)=G^n(u;\mu_2,\sigma_2,\gamma_2)$ and we have

$$P(X_1<Y,\cdots,X_n<Y)=P(M_n\le Y)=\int_{-\infty}^{\infty}G^n(u;\mu_2,\sigma_2,\gamma_2)g(u;\mu_1,\sigma_1,\gamma_1)du=:I_n. \tag{3.14}$$

Closed expressions for (3.14) are presented in the following result. Its proof follows the same steps of Theorems 3.1 and 3.4 and it will be omitted.

Theorem 3.7. *Let $X_1,\cdots,X_n$ be i.i.d. RVs with distribution $GEV(\mu_2,\sigma_2,\gamma_2)$ and Y be an independent RV with distribution $GEV(\mu_1,\sigma_1,\gamma_1)$. Then*

- *When $\gamma_j>0,\ j=1,2$:*

$$P(X_1<Y,\cdots,X_n<Y)=\mathbb{H}\left(1,\frac{\gamma_2\sigma_1}{\sigma_2\gamma_1}n^{-\gamma_2},-\gamma_1,\left[1+\frac{\gamma_2}{\sigma_2}\left(\mu_1-\mu_2-\frac{\sigma_1}{\gamma_1}\right)\right]n^{-\gamma_2},-\frac{1}{\gamma_2}\right), \tag{3.15}$$

provided that $\mu_1-\frac{\sigma_1}{\gamma_1}\ge\mu_2-\frac{\sigma_2}{\gamma_2}$. When $\mu_1-\frac{\sigma_1}{\gamma_1}\le\mu_2-\frac{\sigma_2}{\gamma_2}$:

$$P(X_1<Y,\cdots,X_n<Y)=1-\mathbb{H}\left(1,\frac{\gamma_1\sigma_2}{\sigma_1\gamma_2}n^{-\gamma_1},-\gamma_2,\left[1+\frac{\gamma_1}{\sigma_1}\left(\mu_2-\mu_1-\frac{\sigma_2}{\gamma_2}\right)\right]n^{-\gamma_1},-\frac{1}{\gamma_1}\right). \tag{3.16}$$

- *When $\gamma_j<0,\ j=1,2$:*

$$P(X_1<Y,\cdots,X_n<Y)=\mathbb{H}\left(1,\frac{\gamma_2\sigma_1}{\sigma_2\gamma_1}n^{-\gamma_2},-\gamma_1,\left[1+\frac{\gamma_2}{\sigma_2}\left(\mu_1-\mu_2-\frac{\sigma_1}{\gamma_1}\right)\right]n^{-\gamma_2},-\frac{1}{\gamma_2}\right), \tag{3.17}$$

provided that $\mu_1-\frac{\sigma_1}{\gamma_1}\le\mu_2-\frac{\sigma_2}{\gamma_2}$. When $\mu_1-\frac{\sigma_1}{\gamma_1}\ge\mu_2-\frac{\sigma_2}{\gamma_2}$:

$$P(X_1<Y,\cdots,X_n<Y)=1-\mathbb{H}\left(1,\frac{\gamma_1\sigma_2}{\sigma_1\gamma_2}n^{-\gamma_1},-\gamma_2,\left[1+\frac{\gamma_1}{\sigma_1}\left(\mu_2-\mu_1-\frac{\sigma_2}{\gamma_2}\right)\right]n^{-\gamma_1},-\frac{1}{\gamma_1}\right). \tag{3.18}$$

- *When $\gamma_1=\gamma_2=0$:*

$$P(X_1<Y,\cdots,X_n<Y)=\exp\left\{\frac{\mu_1}{\sigma_1}\right\}\mathbb{H}\left(\exp\left\{\frac{\mu_1}{\sigma_1}\right\},\exp\left\{\frac{\mu_2}{\sigma_2}\right\}n,\frac{\sigma_1}{\sigma_2},0,1\right). \tag{3.19}$$

Remark 3.8. *In a broader k-out-of-n multicomponent reliability context, consider independent RVs $Y, X_1, \cdots, X_k$ with $Y \sim GEV(\mu_1, \sigma_1, \gamma_1)$ and $X_j \sim GEV(\mu_2, \sigma_2, \gamma_2)$, for $j = 1, \cdots, k$ (X_j's are i.i.d.). The reliability for this kind of model is given by*

$$\begin{aligned} R_{s,k} &= P(\textit{at least s out of } (X_1, \cdots, X_k) \textit{ exceed } Y) \\ &= \sum_{j=s}^{k} \binom{k}{j} \int_{-\infty}^{\infty} (1 - G(u; \mu_2, \sigma_2, \gamma_2))^j \, (G(u; \mu_2, \sigma_2, \gamma_2))^{k-j} \, g(u; \mu_1, \sigma_1, \gamma_1) du. \end{aligned}$$

Using a binomial expansion, we obtain

$$R_{s,k} = \sum_{j=s}^{k} \sum_{r=0}^{j} \binom{k}{j} \binom{j}{r} (-1)^{j-r} \int_{-\infty}^{\infty} (G(u; \mu_2, \sigma_2, \gamma_2))^{k-r} \, g(u; \mu_1, \sigma_1, \gamma_1) du. \tag{3.20}$$

Note that the integral terms in (3.20) *are particular cases of* (3.14) *provided that $n = k - r$. Therefore,*

$$R_{s,k} = \sum_{j=s}^{k} \sum_{r=0}^{j} \binom{k}{j} \binom{j}{r} (-1)^{j-r} I_{k-r}.$$

4. Estimation

This section deals with parameter estimation for $R = P(X < Y)$ given two independent GEV RVs. The literature presents some maximum likelihood estimators for R considering explicit forms of R obtained by strong parameter restrictions on extreme value distributions (such as [27, 29, 30]). Those approaches require the estimation of the parameters to be done jointly in the two samples. In our case, we release any requirements about having the same parameters between different samples, since we deal with expressions of R in terms of functions $\mathbb{H}$.

Consider the PDF $g(\cdot; \mu, \sigma, \gamma)$ defined in (2.7). Take $\mathbf{X} = (X_1, \cdots, X_n)$ as a sample of n observations. The likelihood function for the $GEV(\cdot; \mu, \sigma, \gamma)$ is given by:

$$L(\mu, \sigma, \gamma; \mathbf{X}) = \prod_{i=1}^{n} g(X_i; \mu, \sigma, \gamma) \mathbb{1}_{[1+\gamma(X_i-\mu)/\sigma>0]}, \tag{4.1}$$

where $\mathbb{1}_A$ denotes the indicator function on the set A. Note that $\prod_{i=1}^{n} \mathbb{1}_{[1+\gamma(X_i-\mu)/\sigma>0]} > 0$ if and only if $x_i \in supp\ g(\cdot; \mu, \sigma, \gamma)$ for all $i = 1, \cdots, n$. Here, *supp g* denotes the support of the function g. Then, if $\gamma \neq 0$, we are not able to obtain the MLE explicitly, so an additional numeric procedure is required in the likelihood maximization (see [6] for a more detailed discussion).

Remark 4.1. *Set $\mathbf{X} = (X_1, \cdots, X_n)$, a random sample of $GEV(\mu_2, \sigma_2, \gamma_2)$, and $\mathbf{Y} = (Y_1, \cdots, Y_m)$, a random sample of $GEV(\mu_1, \sigma_1, \gamma_1)$, with $\gamma_j > 0$, $j = 1, 2$ and $\mu_1 - \sigma_1/\gamma_1 \geq \mu_2 - \sigma_2/\gamma_2$ (or $\gamma_j < 0$, $j = 1, 2$ and $\mu_1 - \sigma_1/\gamma_1 \leq \mu_2 - \sigma_2/\gamma_2$). Let $\hat{\mu}_i, \hat{\sigma}_i, \hat{\gamma}_i$ ($i = 1, 2$) be the estimates of $\mu_i, \sigma_i, \gamma_i$. We are able to estimate R by the invariance property of MLE, as follows:*

$$\hat{R} = \mathbb{H}\left(1, \frac{\hat{\gamma}_2 \hat{\sigma}_1}{\hat{\sigma}_2 \hat{\gamma}_1}, -\hat{\gamma}_1, 1 + \frac{\hat{\gamma}_2}{\hat{\sigma}_2}\left(\hat{\mu}_1 - \hat{\mu}_2 - \frac{\hat{\sigma}_1}{\hat{\gamma}_1}\right), -\frac{1}{\hat{\gamma}_2}\right). \tag{4.2}$$

Alternatively, whenever $\gamma_j > 0$, $j = 1, 2$ and $\mu_1 - \sigma_1/\gamma_1 \leq \mu_2 - \sigma_2/\gamma_2$ (or $\gamma_j < 0$, $j = 1, 2$ and $\mu_1 - \sigma_1/\gamma_1 \geq \mu_2 - \sigma_2/\gamma_2$), the same invariance property can be applied, leading to:

$$\hat{R} = 1 - \mathbb{H}\left(1, \frac{\hat{\gamma}_1\hat{\sigma}_2}{\hat{\sigma}_1\hat{\gamma}_2}, -\hat{\gamma}_2, 1 + \frac{\hat{\gamma}_1}{\hat{\sigma}_1}\left(\hat{\mu}_2 - \hat{\mu}_1 - \frac{\hat{\sigma}_2}{\hat{\gamma}_2}\right), -\frac{1}{\hat{\gamma}_1}\right). \tag{4.3}$$

This is due to the Theorems 3.1 and 3.4 that describe R in terms of the function $\mathbb{H}$ (which is an integral, hence a continuous and measurable function).

Whenever a single set of realizations of the random variables involved is available, the MLE approach above is of utmost importance. This is the case, for example, of asset selection, when a single time series of observed returns is available for each asset.

On the other hand, to illustrate the suitability of the analytical closed-form expressions hereby derived, a direct simulation approach can be carried out. In such case, several samples of size n can be drawn from each random variable, which are then used to estimate the empirical value of R and can be repeated several times. Both approaches will be explored in the next section.

5. Application

5.1. Simulation study

To evaluate the correctness of the closed-form expression for R given in Theorem 3.1, we generate N Monte-Carlo samples, each of which is size n, of the random variables $GEV(\mu_2, \sigma_2, \gamma_2)$ and $GEV(\mu_1, \sigma_1, \gamma_1)$. In these cases, the values of $\mu_2, \sigma_2, \gamma_2, \mu_1, \sigma_1, \gamma_1$ are pre-specified.

The GEV distribution with negative shape parameter is treated in Tables 2 and 3 where we analyze the estimates $\hat{R}$, bias and root mean squared error (RMSE). Table 4 deals with positive shape parameter.

For the simulation, for each line in the Tables 2–4, the following procedure was carried out:

(1) for each Monte-Carlo sample, the estimate $\hat{R}$ is computed empirically, simply as $n^{-1}\sum_i I(x_i < y_i)$, where $I(.)$ is an indicator function, which is 1 for true arguments and 0 otherwise;

(2) $\hat{R}_{MC}$ is evaluated by taking the sample mean of the Monte-Carlo samples $\hat{R}$;

(3) the bias is computed as the difference between the value obtained by Theorem 3.1 and $\hat{R}_{MC}$. The same follows for the RMSE, which also considers the true value as the analytically obtained one.

As expected, Tables 2 and 3 illustrate that the analytical results obtained match the empirical ones. This is a clear evidence of the correctness of the new expressions hereby derived. Besides, it is clear that increasing the sample size n leads to more precise estimations of R, indicating reduced bias and greater consistency. In Table 4 we observe the same good behavior of the estimator characterized by low bias and RMSE.

Table 2. Negative-shape mean, bias and RMSE of $\hat{R}_{MC}$ (N = 100 and n = 100).

μ_2	σ_2	γ_2	μ_1	σ_1	γ_1	R	$\hat{R}_{MC}$	Bias	RMSE
2.0	1.5	-1.0	0.0	0.5	-0.3	0.1147	0.1151	-0.0004	0.0330
0.0	1.5	-1.0	0.0	0.5	-0.3	0.4350	0.4280	0.0070	0.0502
0.4	1.5	-1.0	0.5	0.5	-0.3	0.4650	0.4591	0.0059	0.0525
2.0	1.0	-1.0	0.0	0.7	-0.3	0.0798	0.0796	0.0002	0.0241
0.0	1.0	-1.0	0.0	0.7	-0.3	0.5298	0.5277	0.0021	0.0466
0.4	1.0	-1.0	0.5	0.7	-0.3	0.5686	0.5669	0.0017	0.0492
2.0	1.5	-1.0	0.0	0.9	-0.3	0.1414	0.1387	0.0027	0.0389
0.0	1.5	-1.0	0.0	0.9	-0.3	0.5092	0.5129	-0.0037	0.0462
0.4	1.5	-1.0	0.5	0.9	-0.3	0.5371	0.5362	0.0009	0.0461
2.0	1.5	-1.0	0.0	0.5	-1.0	0.1015	0.1025	-0.0010	0.0313
0.0	1.5	-1.0	0.0	0.5	-1.0	0.3851	0.3873	-0.0022	0.0463
0.4	1.5	-1.0	0.5	0.5	-1.0	0.4116	0.4159	-0.0043	0.0511
2.0	1.0	-1.0	0.0	0.7	-1.0	0.0590	0.0573	0.0017	0.0270
0.0	1.0	-1.0	0.0	0.7	-1.0	0.4358	0.4275	0.0083	0.0530
0.4	1.0	-1.0	0.5	0.7	-1.0	0.4816	0.4807	0.0009	0.0474
2.0	1.5	-1.0	0.0	0.9	-1.0	0.1104	0.1120	-0.0016	0.0277
0.0	1.5	-1.0	0.0	0.9	-1.0	0.4190	0.4279	-0.0089	0.0474
0.4	1.5	-1.0	0.5	0.9	-1.0	0.4478	0.4500	-0.0022	0.0506
2.0	1.5	-1.5	0.0	0.5	-1.5	0.1237	0.1234	0.0003	0.0302
0.0	1.5	-1.5	0.0	0.5	-1.5	0.3715	0.3701	0.0014	0.0488
0.4	1.5	-1.5	0.5	0.5	-1.5	0.3989	0.4034	-0.0045	0.0472
2.0	1.0	-1.5	0.0	0.7	-1.5	0.0822	0.0797	0.0025	0.0276
0.0	1.0	-1.5	0.0	0.7	-1.5	0.4160	0.4134	0.0026	0.0403
0.4	1.0	-1.5	0.5	0.7	-1.5	0.4739	0.4794	-0.0055	0.0548
2.0	1.5	-1.5	0.0	0.9	-1.5	0.1271	0.1303	-0.0032	0.0325
0.0	1.5	-1.5	0.0	0.9	-1.5	0.3999	0.4068	-0.0069	0.0457
0.4	1.5	-1.5	0.5	0.9	-1.5	0.4329	0.4344	-0.0015	0.0452

Table 3. Negative-shape mean, bias and RMSE of $\hat{R}_{MC}$ (N = 1000 and n = 1000).

μ_2	σ_2	γ_2	μ_1	σ_1	γ_1	R	$\hat{R}_{MC}$	Bias	RMSE
2.0	1.5	-1.0	0.0	0.5	-0.3	0.1147	0.1147	-0.0000	0.0106
0.0	1.5	-1.0	0.0	0.5	-0.3	0.4350	0.4355	-0.0005	0.0151
0.4	1.5	-1.0	0.5	0.5	-0.3	0.4650	0.4645	0.0005	0.0156
2.0	1.0	-1.0	0.0	0.7	-0.3	0.0798	0.0798	0.0001	0.0084
0.0	1.0	-1.0	0.0	0.7	-0.3	0.5298	0.5299	-0.0001	0.0159
0.4	1.0	-1.0	0.5	0.7	-0.3	0.5686	0.5693	-0.0007	0.0151
2.0	1.5	-1.0	0.0	0.9	-0.3	0.1414	0.1414	-0.0000	0.0112
0.0	1.5	-1.0	0.0	0.9	-0.3	0.5092	0.5096	-0.0004	0.0162
0.4	1.5	-1.0	0.5	0.9	-0.3	0.5371	0.5374	-0.0003	0.0158
2.0	1.5	-1.0	0.0	0.5	-1.0	0.1015	0.1018	-0.0003	0.0098
0.0	1.5	-1.0	0.0	0.5	-1.0	0.3851	0.3852	-0.0001	0.0152
0.4	1.5	-1.0	0.5	0.5	-1.0	0.4116	0.4119	-0.0003	0.0152
2.0	1.0	-1.0	0.0	0.7	-1.0	0.0590	0.0587	0.0003	0.0074
0.0	1.0	-1.0	0.0	0.7	-1.0	0.4358	0.4366	-0.0008	0.0154
0.4	1.0	-1.0	0.5	0.7	-1.0	0.4816	0.4814	0.0002	0.0150
2.0	1.5	-1.0	0.0	0.9	-1.0	0.1104	0.1106	-0.0002	0.0101
0.0	1.5	-1.0	0.0	0.9	-1.0	0.4190	0.4196	-0.0006	0.0157
0.4	1.5	-1.0	0.5	0.9	-1.0	0.4478	0.4471	0.0008	0.0150
2.0	1.5	-1.5	0.0	0.5	-1.5	0.1237	0.1243	-0.0006	0.0103
0.0	1.5	-1.5	0.0	0.5	-1.5	0.3715	0.3722	-0.0007	0.0158
0.4	1.5	-1.5	0.5	0.5	-1.5	0.3989	0.3974	0.0014	0.0154
2.0	1.0	-1.5	0.0	0.7	-1.5	0.0822	0.0822	0.0000	0.0086
0.0	1.0	-1.5	0.0	0.7	-1.5	0.4160	0.4166	-0.0005	0.0160
0.4	1.0	-1.5	0.5	0.7	-1.5	0.4739	0.4736	0.0004	0.0162
2.0	1.5	-1.5	0.0	0.9	-1.5	0.1271	0.1265	0.0007	0.0105
0.0	1.5	-1.5	0.0	0.9	-1.5	0.3999	0.3998	0.0001	0.0155
0.4	1.5	-1.5	0.5	0.9	-1.5	0.4329	0.4323	0.0006	0.0151

Table 4. Positive-shape mean, bias and RMSE of $\hat{R}_{MC}$ ($N = 1000$ and $n = 100$).

μ_2	σ_2	γ_2	μ_1	σ_1	γ_1	R	$\hat{R}_{MC}$	Bias	RMSE
2.0	1.5	1.0	0.0	0.5	0.3	0.0617	0.0624	-0.0007	0.0244
0.0	1.5	1.0	0.0	0.5	0.3	0.4288	0.4306	-0.0018	0.0497
0.4	1.5	1.0	0.5	0.5	0.3	0.4491	0.4478	0.0012	0.0514
2.0	1.0	1.0	0.0	0.7	0.3	0.0906	0.0891	0.0015	0.0284
0.0	1.0	1.0	0.0	0.7	0.3	0.4443	0.4463	-0.0020	0.0507
0.4	1.0	1.0	0.5	0.7	0.3	0.4717	0.4727	-0.0010	0.0499
2.0	1.5	1.0	0.0	0.9	0.3	0.1299	0.1300	-0.0001	0.0332
0.0	1.5	1.0	0.0	0.9	0.3	0.4419	0.4423	-0.0004	0.0509
0.4	1.5	1.0	0.5	0.9	0.3	0.4611	0.4602	0.0008	0.0497
2.0	1.5	1.0	0.0	0.5	1.0	0.1469	0.1474	-0.0004	0.0359
0.0	1.5	1.0	0.0	0.5	1.0	0.4764	0.4774	-0.0010	0.0486
0.4	1.5	1.0	0.5	0.5	1.0	0.4947	0.4940	0.0007	0.0504
2.0	1.0	1.0	0.0	0.7	1.0	0.1846	0.1847	-0.0001	0.0382
0.0	1.0	1.0	0.0	0.7	1.0	0.4980	0.4985	-0.0006	0.0520
0.4	1.0	1.0	0.5	0.7	1.0	0.5240	0.5235	0.0004	0.0480
2.0	1.5	1.0	0.0	0.9	1.0	0.2144	0.2164	-0.0019	0.0431
0.0	1.5	1.0	0.0	0.9	1.0	0.4951	0.4925	0.0026	0.0497
0.4	1.5	1.0	0.5	0.9	1.0	0.5128	0.5124	0.0004	0.0480
2.0	1.5	1.5	0.0	0.5	1.5	0.1868	0.1883	-0.0015	0.0393
0.0	1.5	1.5	0.0	0.5	1.5	0.4906	0.4894	0.0013	0.0523
0.4	1.5	1.5	0.5	0.5	1.5	0.5076	0.5072	0.0004	0.0495
2.0	1.0	1.5	0.0	0.7	1.5	0.2237	0.2243	-0.0006	0.0414
0.0	1.0	1.5	0.0	0.7	1.5	0.5056	0.5052	0.0004	0.0501
0.4	1.0	1.5	0.5	0.7	1.5	0.5318	0.5310	0.0008	0.0496
2.0	1.5	1.5	0.0	0.9	1.5	0.2461	0.2451	0.0011	0.0444
0.0	1.5	1.5	0.0	0.9	1.5	0.5046	0.5042	0.0004	0.0491
0.4	1.5	1.5	0.5	0.9	1.5	0.5220	0.5226	-0.0006	0.0490

5.2. Brazil stock market

In order to evaluate the proposed framework, we modeled stock prices log-returns as GEV and compare them in a reliability sense. For this, we assume that the returns are independent. To meet independence requirements, we take stock log-returns that are not correlated and correlations are measured using Pearson, Kendal and Spearman methods.

Denote X_1, X_2, X_3 and X_4 the stock price log-returns of BBAS3, ITUB4, VALE3 and VIIA3, respectively. The data sets are retrieved directly through the software R [34] by the command

```
quantmod::getSymbols("BBAS3.SA", src = "yahoo", auto.assign = FALSE,
from = '2022-01-01', to = '2023-04-30', return.class = 'xts').
```

The data sets have information in Brazilian currency (R$, BRL).

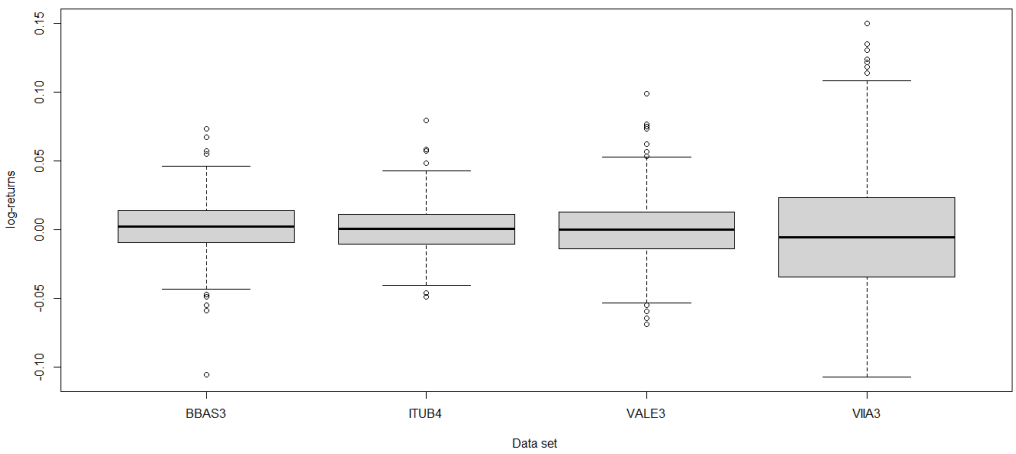
Summary statistics for the data sets X_1, X_2, X_3 and X_4 are presented in Table 5. A boxplot is

presented in Figure 2 showing the symmetry around zero of the log-returns and that X_4 has more dispersion than the others.

Table 5. Summary statistics for the stock price log-returns X_1, X_2, X_3 and X_4.

RV	Data Set	min	1st Qu.	Median	Mean	3rd Qu.	Max	n
X_1	BBAS3	-0.1057	-0.0097	0.0019	0.0012	0.0136	0.0736	330
X_2	ITUB4	-0.0492	-0.0105	0.0004	0.0006	0.0109	0.0794	330
X_3	VALE3	-0.0689	-0.0140	0.0001	-0.0002	0.0128	0.0989	330
X_4	VIIA3	-0.1075	-0.0344	-0.006	-0.0030	0.0231	0.1504	330

Figure 2. Boxplot for the stock price log-returns of the data sets BBAS3, ITUB4, VALE3 and VIIA3.



Maximum likelihood (ML) estimates and Kolmogorov-Smirnov (KS) p-values are given in Table 6. Figure 3 shows histograms and the fit of the GEV model to X_1, X_2, X_3 and X_4. For each data set, the empirical CDF (ECDF) is compared to the theoretical model in Figure 4.

Although the p-value of the Kolmogorov-Smirnov test is small for the BBAS3 (X_1), the graphical analysis does not invalidate the good fit of the distribution to the data.

Table 6. ML estimates, log-likelihood (ll_{max}) and KS p-values for the GEV models.

RV	Data set	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$-ll_{max}$	KS p-value
X_1	BBAS3	-0.0063	0.0219	-0.2535	-803.9168	0.0147
X_2	ITUB4	-0.0064	0.0165	-0.1545	-870.9392	0.4299
X_3	VALE3	-0.0095	0.0222	-0.1631	-774.0895	0.2331
X_4	VIIA3	-0.0217	0.0396	-0.1170	-567.6979	0.6996

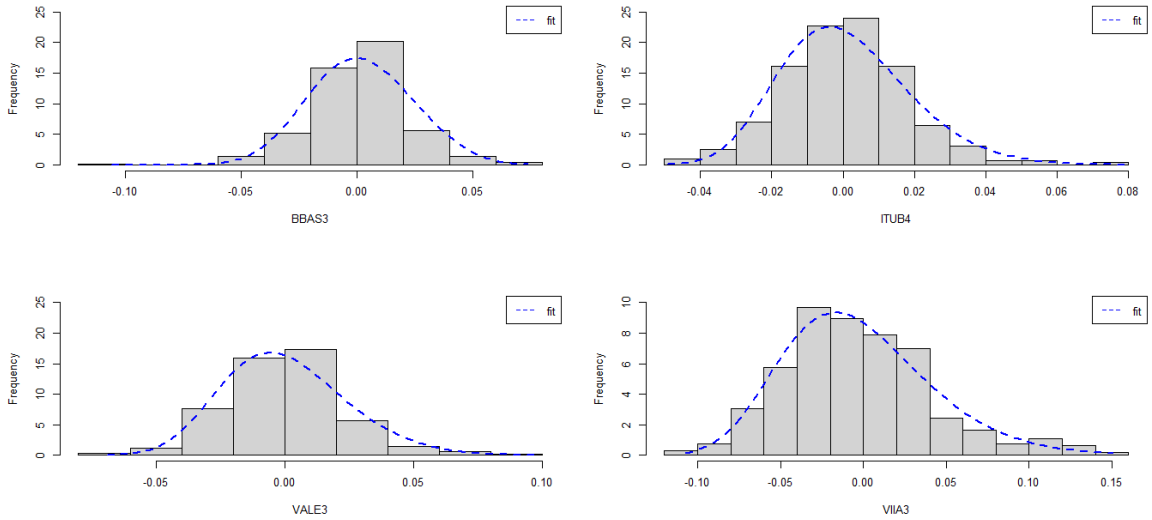


Figure 3. Histograms and fitted GEV densities for the stock prices log-returns.

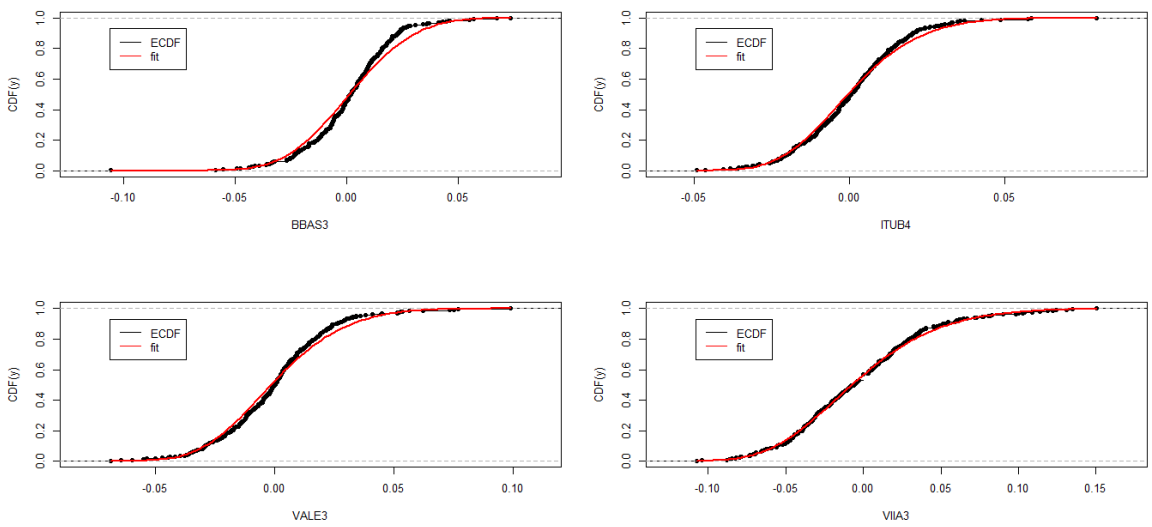


Figure 4. Fitted ECDF for GEV models.

Aiming at estimating probabilities of the type $R = P(X < Y)$ via Theorems 3.1–3.4, we need X and Y to be independent RVs. In this sense, we analyzed the dependency structures of $X_1, \cdots, X_4$ using the Pearson, Kendal and Spearman correlation matrices, and the results are presented in the Appendix. The pairs chosen to be compared are $X_3 - X_2$, $X_3 - X_4$ and $X_3 - X_1$.

Reliability measures of the type $R = P(X < Y)$ appear in a decision process of an investor. In summary, when X and Y represent profit RVs and $R < 1/2$, it is advisable that the investor chooses the variable X. If $R > 1/2$, the opposite occurs. The case $R = 1/2$ is inconclusive. Thus, knowing how to evaluate R accurately is important to support the decision process. In this sense, Table 7 presents the estimates of $P(X_3 < X_1)$, $P(X_3 < X_2)$ and $P(X_3 < X_4)$ and the 95% bootstrap confidence

intervals (CI). For all subsequent Tables, $\hat{R}$ is the estimate obtained by first fitting the dataset to a GEV distribution and using the results from Remark 4.1 (invariance property of MLE of the parameters and the analytical expression obtained for R). Also, $\hat{R}_{NP}$ is a non-parametric estimation of R, which considers a similar approach as in the simulation study (uses an indicator function to compare samples from the two distributions). Finally, $\hat{R}_{Boot}$ is the mean value of the bootstrap estimation of R using the non-parametric approach and the CI reported is for such mean.

Table 7. Stress-strength probability estimates and bootstrap CI.

$R = P(X < Y)$	$\hat{R}$	$\hat{R}_{NP}$	$\hat{R}_{Boot}$	95% CI
$P(X_3 < X_1)$	0.5283	0.5242	0.5277	(0.4883; 0.5676)
$P(X_3 < X_2)$	0.5174	0.5242	0.5181	(0.4766; 0.5569)
$P(X_3 < X_4)$	0.4506	0.4364	0.4500	(0.4055; 0.4944)

The estimates of R indicate that, within the analyzed time period, VALE3 would be preferred over BBAS3 and ITUB4, while VIIA3 would be preferred over VALE3. On the other hand, the bootstrap CI estimates indicate that only the case $P(X_3 < X_4)$ was conclusive since 0.5 does not belong to the estimated confidence interval.

It is important to compare different distributions as candidate models for the log-returns modelling. Considering previous results in the literature [11], we compared the performance of the GEV and generalized logistic distributions as models for daily returns, as presented in Table 8.

Table 8. Log-likelihood comparison between different candidate random variables.

RV	Data set	GEV	Generalized Logistic
X_1	BBAS3	-803.92	- 833.43
X_2	ITUB4	-870.94	- 882.01
X_3	VALE3	-774.09	- 787.81
X_4	VIIA3	-567.70	- 569.60

It is possible to see that both GEV and generalized logistic provided quite similar modelling capabilities (about the same log-likelihood values). Since the two distributions considered have three parameters, there is no need to consider information criteria.

5.2.1. Considering weekly maximum log-returns

Instead of considering the log-returns for the closing prices, as in the previous analyses, one may study how the weekly maximums (or minimums) behave. This has a direct economic interpretation: a proxy for greater profits (or greater losses, i.e., shortfalls and values-at-risk) and has been explored in the literature [11, 12, 18].

Following previous studies [12], it is of interest to explore the modelling scenario where the maximum returns are considered GEV random variables. Table 9 show the GEV ML estimates for such case.

Table 9. ML estimates, log-likelihood (ll_{max}) and KS p-values for weekly maximum log-returns and their GEV modelling.

RV	Data set	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	$-ll_{max}$	KS p-value
X_1	BBAS3	-0.0128	0.0393	-0.1157	-116.3016	0.8076
X_2	ITUB4	-0.091	0.0397	-0.4319	-126.5530	0.8292
X_3	VALE3	-0.0197	0.0426	-0.2003	-114.0183	0.9365
X_4	VIIA3	-0.0464	0.0750	-0.1339	-72.9854	0.5378

Table 9 indicates that the GEV random variable is adequate for every case considered, which is expected as the EVT predicts such convergence. Now that the parameters have been obtained, Table 10 presents the reliability calculations.

Table 10. Stress-strength probability estimates and bootstrap CI for weekly maximums.

$R = P(X < Y)$	$\hat{R}$	$\hat{R}_{NP}$	$\hat{R}_{Boot}$	95% CI
$P(X_3 < X_1)$	0.5453	0.5522	0.5413	(0.4483; 0.6334)
$P(X_3 < X_2)$	0.5409	0.5522	0.5483	(0.4421; 0.6358)
$P(X_3 < X_4)$	0.4364	0.4328	0.4350	(0.3379; 0.5350)

Since $R = 0.5$ is within the confidence intervals, the metric becomes inconclusive. It is possible to notice, on the other hand, that if $\hat{R}$ is considered as a metric by itself, it would precisely reflect the high volatility of VIIA3, which suffered severe instability and losses during the time window analyzed.

Besides the direct application to financial assets selection, some engineering applications can also benefit from the new relations hereby defined. One application is illustrated in the next subsection.

5.3. Carbon fibers

Stress-strength reliability can be also applied to the modelling and comparison of carbon fibers of lengths 10 and 20 mm. The data represent the strength data measured in GPa (gigapascal), for single carbon fibers tested under tension. The data are frequently used in the literature and are also presented below (e.g., [28]).

Carbon fibers of length 20 mm:

$$X = (1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.977, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.140, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.570, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.880, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585, 3.585).$$

Carbon fibers of length 10 mm:

$$Y = (1.901, 2.132, 2.203, 2.228, 2.257, 2.350, 2.361, 2.396, 2.397, 2.445, 2.454, 2.474, 2.518, 2.522, 2.525, 2.532, 2.575, 2.614, 2.616, 2.618, 2.624, 2.659, 2.675, 2.738, 2.740, 2.856, 2.917, 2.928, 2.937, 2.937, 2.977, 2.996, 3.030, 3.125, 3.139, 3.145, 3.220, 3.223, 3.235, 3.243, 3.264, 3.272,$$

3.294, 3.332, 3.346, 3.377, 3.408, 3.435, 3.493, 3.501, 3.537, 3.554, 3.562, 3.628, 3.852, 3.871, 3.886, 3.971, 4.024, 4.027, 4.225, 4.395, 5.020).

Table 11 and Figure 5 show the summary statistics of X and Y in which it is possible to observe that Y (carbon fibers of length 10 mm) tend to have greater strength values than X (carbon fibers of length 20 mm).

Table 11. Summary statistics for the carbon fibers of length 20 mm (X) and 10 mm (Y).

RV	Data set	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n
X	Carbon fibers of 20 mm	1.312	2.098	2.478	2.451	2.773	3.585	69
Y	Carbon fibers of 10 mm	1.901	2.554	2.996	3.059	3.421	5.020	63

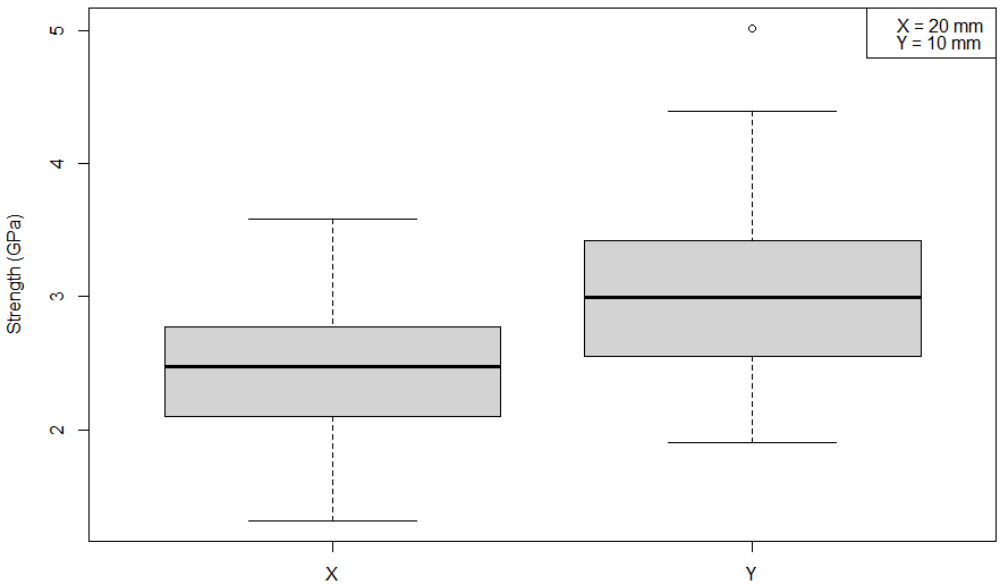


Figure 5. Boxplot for the carbon fibers of length 20 mm (X) and 10 mm (Y).

ML estimates and KS p-values for the GEV model are presented in Table 12 and the good fit can be observed in Figure 6.

Table 12. ML estimates and Kolmogorov-Smirnov (KS) p-values for the GEV model.

Data set	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$	KS p-value
X	2.2781	0.4956	-0.2851	0.9978
Y	2.7904	0.5245	-0.0747	0.8216

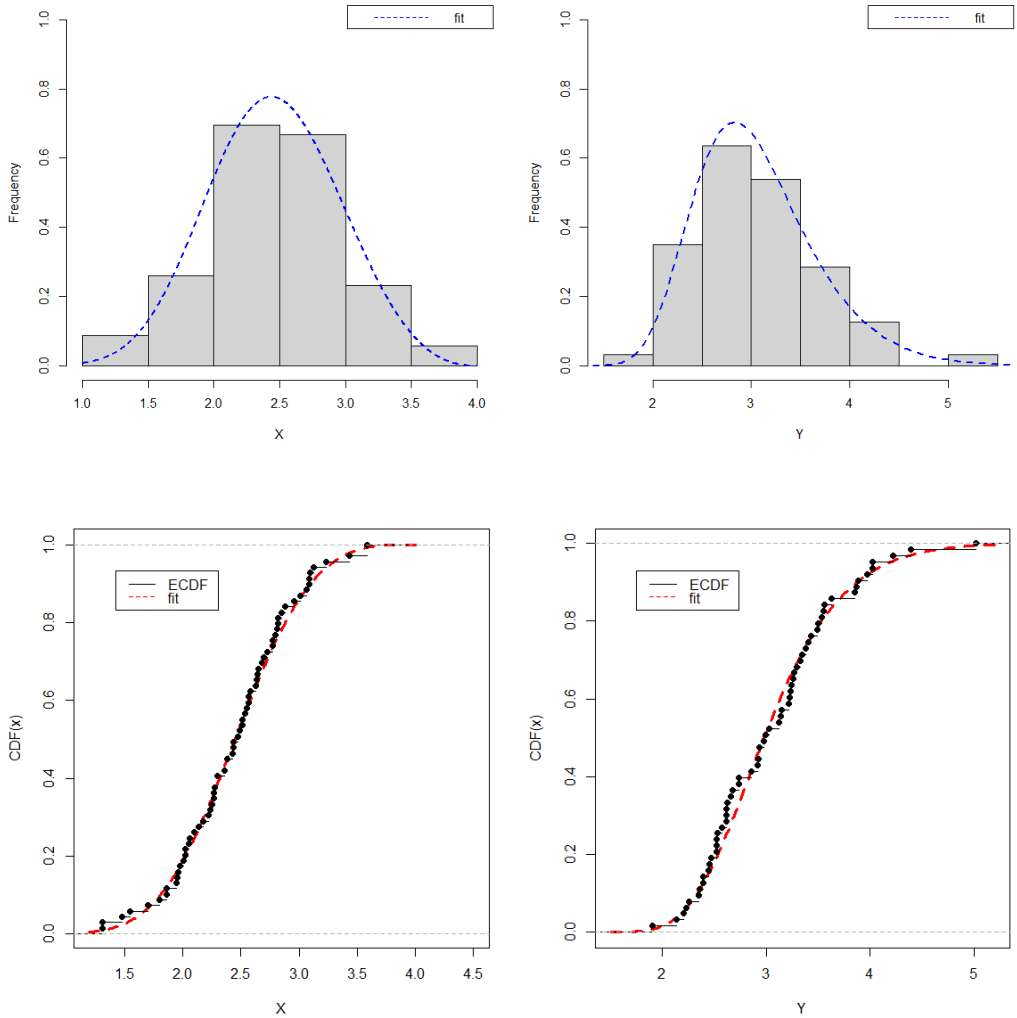


Figure 6. Histogram, ECDF and fitted GEV model for carbon fibers.

The estimates of stress-strength reliability and bootstrap 95% CI are $\hat{R} = 0.774$ and $(0.699, 0.851)$, respectively. It is easy to conclude that since $P(X < Y) > 1/2$ and 0.5 is not within the CI, that $X < Y$ statistically, i.e., the fibers with length of 10 mm have statistically greater strength values than the ones of length of 20 mm.

6. Conclusions

Despite its significant limitations, Markowitz's modern portfolio theory is still relied upon by many practitioners because of its user-friendly simplicity. This way, studying alternative approaches that are also of straightforward comprehension is of utmost importance.

In this paper, we studied the stress-strength reliability $R = P(X < Y)$ when both X and Y follow three-parameter GEV distributions. In summary, when X and Y represent return RVs and $R < 1/2$, it is advisable that the investor chooses the variable X. If $R > 1/2$, the opposite occurs. The case $R = 1/2$ is inconclusive. Thus, exact expressions for R have been obtained in terms of the extreme-

value $\mathbb{H}$-function with minimal parameter restrictions. With additional restrictions, it was shown that R can be calculated in terms of H-functions and even in terms of standard functions (such as exponential functions).

Monte-Carlo simulations attested to the performance of the analytical closed-form expressions hereby derived. By applying our methodology to real-world financial data, we could orient a stock selection procedure by calculating $P(X < Y)$ when both X and Y represent stock returns. Besides, an engineering application was also described, where carbon fibers tested under tension were modelled in a stress-strength reliability sense.

Appendix

In this appendix, we present the correlation matrices of the data sets (log-returns) modeled in Section 5. Thus, Tables 13–15 give the correlations of Pearson, Spearman and Kendal, respectively.

Table 13. Pearson correlation matrix.

	X_1	X_2	X_3	X_4
X_1	1.00	0.61	**0.10**	0.29
X_2		1.00	**0.15**	0.31
X_3			1.00	**0.03**
X_4				1.00

Table 14. Spearman rank correlation matrix.

	X_1	X_2	X_3	X_4
X_1	1.00	0.63	**0.15**	0.31
X_2		1.00	**0.22**	0.34
X_3			1.00	**0.06**
X_4				1.00

Table 15. Kendall rank correlation matrix.

	X_1	X_2	X_3	X_4
X_1	1.00	0.45	**0.10**	0.21
X_2		1.00	**0.15**	0.24
X_3			1.00	**0.04**
X_4				1.00

Use of AI tools declaration

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

Acknowledgments

The authors acknowledge the support provided by the University of Brasilia (UnB). Additionally, M.O. acknowledges the Coordination for the Improvement of Higher Education Personnel (CAPES) for awarding him a scholarship for a Master of Science program.

Conflict of interest

The authors declare no conflicts of interest.

References

1. H. Markowitz, *Portfolio selection: Efficient diversification of investments*, John Wiley & Sons, 1959.
2. M. Rubinstein, Markowitz's "portfolio selection": A fifty-year retrospective, *J. Financ.*, **57** (2002), 1041–1045. https://doi.org/10.1111/1540-6261.00453
3. H. Markowitz, Portfolio selection, *J. Financ.*, **7** (1952), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
4. W. Sharpe, A simplified model for portfolio analysis, *Manag. Sci.*, **9** (1963), 277–293. https://doi.org/10.1287/mnsc.9.2.277
5. L. Bachelier, *Theorie de la speculation*, Doctor Thesis, The Random Character Of Stock Market Prices, Annales Scientifiques Ecole Normale Sperieure III-17, 1900. https://doi.org/10.24033/asens.476
6. P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling extremal events: For insurance and finance*, Springer Science & Business Media, 2013.
7. N. N. Taleb, *Statistical consequences of fat tails, technical incerto*, STEM Academic Press, 2020.
8. B. Mandelbrot, The variation of some other speculative prices, *J. Bus.*, **40** (1967), 393–413. Available from: `http://www.jstor.org/stable/2351623.`
9. A. Jenkinson, The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Q. J. Roy. Meteor. Soc.*, **81** (1955), 158–171.
10. P. Cirillo, N. N. Taleb, Expected shortfall estimation for apparently infinite-mean models of operational risk, *Quant. Financ.*, **16** (2016), 1485–1494. https://doi.org/10.1080/14697688.2016.1162908
11. G. D. Gettinby, C. D. Sinclair, D. M. Power, R. A. Brown, An analysis of the distribution of extreme share returns in the UK from 1975 to 2000, *J. Bus. Finan. Account.*, **31** (2004), 607–646. https://doi.org/10.1111/j.0306-686X.2004.00551.x
12. S. I. Hussain, S. Li, Modeling the distribution of extreme returns in the Chinese stock market, *J. Int. Financ. Mark. I.*, **34** (2015), 263–276. https://doi.org/10.1016/j.intfin.2014.11.007
13. R. Gencay, F. Selçuk, Extreme value theory and Value-at-Risk: Relative performance in emerging markets, *Int. J. Forecasting*, **20** (2004), 287–303. https://doi.org/10.1016/j.ijforecast.2003.09.005

14. J. Galambos, *The asymptotic theory of extreme order statistics*, R. E. Krieger Publishing Company, 1987.
15. L. Haan, A. Ferreira, *Extreme value theory: An introduction*, Springer, 2006. https://doi.org/10.1007/0-387-34471-3
16. S. Resnick, *Extreme values, regular variation, and point processes*, Springer Science & Business Media, 2008.
17. R. Fisher, L. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Math. Proc. Cambridge*, **24** (1928), 180–190. https://doi.org/10.1017/S0305004100015681
18. A. Goncu, A. K. Akgul., O. Imamoglu, M. Tiryakioglu, An analysis of the extreme returns distribution: The case of the Istanbul Stock Exchange, *Appl. Financ. Econ.*, **22** (2012), 723–732. https://doi.org/10.1080/09603107.2011.624081
19. S. Kotz, Y. Lumelskii, M. Pensky, *The stress-strength model and its generalizations: Theory and applications*, World Scientific, 2003. https://doi.org/10.1142/9789812564511
20. F. Domma, S. Giordano, A stress-strength model with dependent variables to measure household financial fragility, *Stat. Method. Appl.*, **21** (2012), 375–389. https://doi.org/10.1007/s10260-012-0192-5
21. L. C. S. M. Ozelim, C. E. G. Otiniano, P. N. Rathie, On the linear combination of n logistic random variables and reliability analysis, *South East Asian J. Math. Math. Sci.*, **12** (2016), 19–34.
22. P. N. Rathie, L. C. S. M. Ozelim, Exact and approximate expressions for the reliability of stable Lévy random variables with applications to stock market modelling, *J. Comput. Appl. Math.*, **321** (2017), 314–322. https://doi.org/10.1016/j.cam.2017.02.043
23. L. C. S. M. Ozelim, P. N. Rathie, Linear combination and reliability of generalized logistic random variables, *Eur. J. Pure Appl. Math.*, **12** (2019), 722–733. https://doi.org/10.29020/nybg.ejpam.v12i3.3444
24. P. N. Rathie, L. C. S. M. Ozelim, B. B. Andrade, Portfolio management of copula-dependent assets based on $P(Y < X)$ reliability models: Revisiting Frank copula and Dagum distributions, *Stats*, **4** (2021), 1027–1050. https://doi.org/10.3390/stats4040059
25. S. Nadarajah, Reliability for extreme value distributions, *Math. Comput. Model.*, **37** (2003), 915–922. https://doi.org/10.1016/S0895-7177(03)00107-9
26. K. Krishnamoorthy, Y. Lin, Confidence limits for stress-strength reliability involving Weibull models, *J. Stat. Plan. Infer.*, **140** (2010), 1754–1764. https://doi.org/10.1016/j.jspi.2009.12.028
27. D. Kundu, M. Raqab, Estimation of $R = P(Y < X)$ for three-parameter Weibull distribution, *Stat. Probabil. Lett.*, **79** (2009), 1839–1846. https://doi.org/10.1016/j.spl.2009.05.026
28. R. Nojosa, P. N. Rathie, Stress-strength reliability models involving generalized gamma and Weibull distributions, *Int. J. Qual. Reliab. Ma.*, **37** (2020), 538–551. https://doi.org/10.1108/IJQRM-06-2019-0190
29. K. Abbas, Y. Tang, Objective Bayesian analysis of the Frechet stress-strength model, *Stat. Probabil. Lett.*, **84** (2014), 169–175. https://doi.org/10.1016/j.spl.2013.09.014

30. X. Jia, S. Nadarajah, B. Guo, Bayes estimation of $P(Y < X)$ for the Weibull distribution with arbitrary parameters, *Appl. Math. Model.*, **47** (2017), 249–259. https://doi.org/10.1016/j.apm.2017.03.020

31. P. N. Rathie, L. C. S. M. Ozelim, F. Quintino, T. Fonseca, On the extreme value H-function, *Stats*, **6** (2023), 802–811. Available from: `https://www.mdpi.com/2571-905X/6/3/51.`

32. A. Mathai, R. Saxena, H. Haubold, *The H-function: Theory and applications*, Springer Science & Business Media, 2009.

33. P. N. Rathie, S. Freitas, R. Nojosa, A. Mendes, T. Silva, Stress-strength reliability models involving H-function distributions, *J. Ramanu. Soc. Math. Math. Sci.*, **9** (2022), 217–234.

34. R. C. Team, R: A language and environment for statistical computing, *R Found. Stat. Comput.*, 2022. Available from: `https://www.R-project.org/.`

AIMS Press