Economics Letters 231 (2023) 111301

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Capturing information in extreme events

Omid M. Ardakani

Parker College of Business, Georgia Southern University, United States of America

ARTICLE INFO

Article history:
Received 3 July 2023
Received in revised form 31 July 2023
Accepted 5 August 2023
Available online 9 August 2023

JEL classification:
C13
C51
G17

Keywords:
Entropy
Generalized extreme value distribution
Generalized Pareto
Kullback–Leibler divergence
Tail risk

ABSTRACT

This study integrates information theory and extreme value theory to enhance the prediction of extreme events. Information-theoretic measures provide a foundation for model comparison in tails. The theoretical findings suggest that (1) the entropy of block maxima converges to the entropy of the generalized extreme value distribution, (2) the rate of convergence is controlled by its shape parameter, and (3) the entropy of block maxima is a monotonically decreasing function of the block size. Empirical analysis of E-mini S&P, 500 futures data evaluates the financial risk, capturing information content of extreme events using entropy and Kullback–Leibler divergence.

1. Introduction

Tail risk describes the possibility of significant financial losses during extreme events. Such events, often unpredictable and not effectively captured by conventional risk models, can impact portfolio values dramatically and trigger potentially devastating consequences (Bali et al., 2011). Illustrative examples of tail risk events include the global financial crisis and the COVID-19 pandemic, which had far-reaching implications on global financial systems. Evaluating these risks is essential for maintaining portfolio resilience, ensuring financial stability, and adhering to regulatory compliance (Danielsson, 2011).

Traditional risk measures, such as Value at Risk (VaR) and Conditional Value at Risk (CVaR), often fall short of estimating tail risk accurately. Their limitations stem from the underlying assumptions that do not align with the characteristics of financial return distributions (Fabozzi et al., 2005). Empirical evidence has shown the presence of "heavy tails" in return distributions, indicating a higher probability of extreme events (Mandelbrot and Taylor, 1967; Cont, 2001). Given these constraints, exploring innovative methodologies that can more accurately capture tail risk is crucial. One way to address tail risk is to integrate information theory (IT) with extreme value theory (EV). Entropy, a concept embedded in IT, measures uncertainty inherent in probabilities. A distribution with high entropy would signify a greater degree of uncertainty, capturing a wider range of possible outcomes. By maximizing entropy, we can choose the probability distribution that best reflects the underlying information without presumptions or unwarranted biases, thereby allowing a more thorough examination of potential outcomes (Jaynes, 1957). Meanwhile, EV focuses specifically on the tail of the distribution, providing powerful tools to quantify tail risks. EV models are particularly adept at estimating the probabilities of extreme events, even those with low frequencies and beyond what might be typically expected (Pickands, 1975). Such models have been successfully applied across various fields where predicting extreme events is required (Coles et al., 2001).

This study contributes to the literature by establishing a theoretical framework that integrates the principles of IT and EV. The Kullback–Leibler divergence, also known as relative entropy, presents another essential tool within IT that could potentially enrich the extreme value theory framework. It measures the dissimilarity between two probability distributions (Kullback and Leibler, 1951). In the context of EV, this measure serves as a metric to compare different extreme event models, offering insights into the degree of approximation between generalized extreme value or generalized Pareto distributions and the empirical distribution underlying extreme events. Theoretical findings lay out a connection between IT and EV. This connection revolves around block maxima. The block maxima method involves dividing a series of data into non-overlapping 'blocks' of equal size and then, for each block, taking the maximum value.

This paper delivers three findings. First, it establishes that the entropy of block maxima converges towards the entropy of a

E-mail address: oardakani@georgiasouthern.edu.

https://doi.org/10.1016/j.econlet.2023.111301

generalized extreme value distribution as the size of the block increases. This convergence is illustrated through simulation studies involving normal, exponential, and Pareto distributions. Second, the results highlight that given the shape parameter of the generalized extreme value theory, as the block size increases, the difference between the entropies of the block maxima and the generalized extreme value distribution tends towards a constant, emphasizing the central role of the shape parameter in extreme value theory. Lastly, a result demonstrates the monotonicity of block maxima entropy. It shows that if the entropy of block maxima is a monotonically decreasing function of block size, then as the block size increases, the entropy decreases. This is further substantiated through a simulation example. The empirical study involves daily E-mini S&P 500 futures data from September 25, 2000, to May 23, 2023. E-mini S&P 500 futures contracts are widely traded due to their liquidity, affordability, and ease of access. The findings highlight the disparity between the empirical, generalized extreme value, and generalized Pareto distributions. The entropy measures further illustrate the degree of uncertainty in these distributions.

2. Extreme events and information-theoretic measures

EV and IT offer perspectives that shed light on the characteristics of extreme events and the probabilities associated with their occurrence. This section starts by introducing the methods and applications of EV in the context of financial markets. Subsequently, it discusses the fundamental measures in IT, their definitions, properties, and potential applications to EV. Incorporating these strands together enhances our understanding of extreme events from an information-theoretic standpoint, which could inform risk management strategies and help develop robust predictive models for extreme events.

2.1. Extreme events

EV represents a statistical framework specifically designed to analyze the behavior of extreme values found in the tails of a probability distribution. It has been widely employed across various fields where quantifying the probability of rare events is essential. EV primarily comprises two methodologies. The first is block maxima, which divides data into subsets or blocks and selects the maximum (minimum) value from each. This process generates a new dataset of extremes. A generalized extreme value distribution ($\mathcal{G}$) of Jenkinson (1955), which encompasses Gumbel, Frechet, and Weibull distributions, is then fitted to this data (Longin, 2000). The second method, peaks over threshold, utilizes all data points that exceed a specific threshold. This excess data is then modeled using a generalized Pareto ($\mathcal{P}$) of Pickands (1975). The threshold is sufficiently high to justify using the $\mathcal{P}$ but low enough to ensure enough data points for analysis. Let $F(x)$ be the cumulative density function (CDF). Extreme quantiles and probabilities can be estimated by fitting a model to the empirical survival function, $1 - F(x)$, which focuses on extreme event data (Diebold et al., 2000).

Given the standardized variable $z = (x-\mu)/\sigma$, where μ is the location parameter and $\sigma > 0$ is the scale parameter, the $\mathcal{G}$ CDF can be written as

$$F(z,\xi) = \begin{cases} \exp\left(-(1+\xi z)^{-1/\xi}\right) & \text{for } \xi \neq 0 \\ \exp\left(-\exp(-z)\right) & \text{for } \xi = 0, \end{cases} \quad (1)$$

where ξ is the shape parameter governing the tail of the distribution. When $\xi = 0$, $\mathcal{G}$ simplifies to the Gumbel distribution. When $\xi > 0$, $\mathcal{G}$ transforms into the Frechet distribution. Conversely, when $\xi < 0$, it becomes equivalent to the Weibull distribution. The $\mathcal{G}$ and $\mathcal{P}$ are related through the Pickands–Balkema–de Haan theorem (Balkema and Haan, 1974; Pickands, 1975). It states that if a dataset can be modeled with a $\mathcal{G}$, then the excesses over a sufficiently high threshold of that dataset can be approximated with a $\mathcal{P}$. The $\mathcal{P}$ CDF is given by

$$F(z,\xi) = \begin{cases} 1-(1+\xi z)^{-1/\xi} & \text{for } \xi \neq 0 \\ 1-\exp(-z) & \text{for } \xi = 0. \end{cases} \quad (2)$$

$\mathcal{G}$ and $\mathcal{P}$ have become pivotal in predicting extreme events in financial markets, such as stock market crashes, catastrophic insurance claims, and severe portfolio losses. These events usually lie in the tail of the distribution and can lead to substantial financial losses (Embrechts et al., 2013). For example, the $\mathcal{G}$ can be applied to model block maxima of financial returns by partitioning the time series into non-overlapping blocks (e.g., days, weeks, or months) and selecting the maximum loss from each block. In contrast, the $\mathcal{P}$ is used in the peaks over threshold approach. The underlying principle involves fitting this distribution to the tails of the return distribution. These are the returns that could potentially result in substantial losses. The $\mathcal{P}$ is instrumental in estimating VaR and CVaR (McNeil and Frey, 2000). Accurately estimating the tail index is critical in this context. Some approaches, such as those proposed by Danielsson et al. (2001), use bootstrap methods to choose the sample fraction when estimating the tail index, providing more reliable results for risk management.

Bali (2003) raises concerns that neither the $\mathcal{G}$ nor the $\mathcal{P}$ accurately depicts extreme movements in financial markets. He suggests implementing a general extreme value distribution that utilizes the Box and Cox (1964) transformation, enhancing the precision of extreme value modeling in financial contexts.

2.2. Information-theoretic measures

IT is concerned with quantifying information. The fundamental measure developed by Shannon (1948) is entropy, which quantifies the uncertainty or randomness associated with a set of values. Applications of information theory and maximum entropy in finance can be found in Ardakani et al. (2018) and Ardakani (2022). In the context of extreme value theory, the entropy of a given distribution can be used to quantify the degree of uncertainty associated with extreme events. As entropy provides a quantitative measure of randomness in a given dataset, it can be used for examining the properties of extreme events. High entropy corresponds to high uncertainty, suggesting a higher potential for extreme events. In contrast, lower entropy indicates a more predictable dataset with fewer extreme events. Specifically, if we consider $\mathcal{G}$ and $\mathcal{P}$ used in extreme value theory, we can define the entropy H of these distributions. Entropy is usually defined as

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log(f(x))dx, \quad (3)$$

where $f(x)$ is the probability density function (PDF) of the random variable X. Therefore, the entropy of a $\mathcal{G}$ or $\mathcal{P}$ can be calculated by substituting their respective PDFs into (3). Through this connection, extreme value theory can be linked with information theory.

Another concept that can be leveraged in extreme value theory is the Kullback–Leibler (KL) divergence, also known as relative entropy. Introduced by Kullback and Leibler (1951), KL divergence measures the difference between two probability distributions. Given two PDFs $f(x)$ and $g(x)$, the KL divergence from $g(x)$ to $f(x)$ is defined as

$$K(f;g) = \int_{-\infty}^{\infty} f(x)\log\left(\frac{f(x)}{g(x)}\right)dx. \quad (4)$$

In the context of extreme value theory, KL divergence could provide a way to compare different extreme event models. For instance, it could be used to quantify how well $\mathcal{G}$ or $\mathcal{P}$ approximates the true underlying distribution of extreme events. The KL divergence measures the information loss when $g(x)$ is used to approximate $f(x)$ (Soofi, 2000).

Lemma 1 (*Properties of KL Divergence*). *KL divergence has the following properties:*

1. *Non-negativity:* $K(f;g) \geq 0$ *for all* f *and* g. *Equality holds if and only if* $f = g$ *almost everywhere.*
2. *Non-symmetry: In general,* $K(f;g) \neq K(g;f)$.
3. *Continuity: If* $f(x)$ *and* $g(x)$ *are continuous PDFs, then* $K(f;g)$ *is continuous in the parameters of* f *and* g.
4. *Additivity for Independent Random Variables: If* X *and* Y *are independent, then* $K(f_{X,Y};g_{X,Y}) = K(f_X;g_X) + K(f_Y;g_Y)$.

The connection between extreme value theory and information theory gives new possibilities for modeling and analyzing extreme events. For example, it could lead to the development of new risk management strategies that take into account the entropy and KL divergence of a given distribution to better predict and prepare for extreme events (Ardakani, 2023). Furthermore, the tools and concepts from information theory, such as mutual information, could be used to compare different extreme event models and choose the most suitable one for a given dataset (Cover and Thomas, 1991).

3. Information content of extreme events

This section links EV theory discussed in Section 2.1, and IT explained in Section 2.2.

Definition 1 (*KL Divergence for Extreme Value Distributions*). The KL divergence between two generalized extreme value distributions F_1 and F_2 with PDFs f_1 and f_2 is given by substituting f_1 and f_2 into Eq. (4). Similarly, the KL divergence between two generalized Pareto distributions can be calculated.

The KL divergence provides a natural tool to assess the fitness of a $\mathcal{G}$ or $\mathcal{P}$ to the empirical data of extreme events. A smaller KL divergence implies a better fit of the model. With the properties in Lemma 1, KL divergence can be used to provide theoretical foundations for comparisons of extreme event models. The model comparison also provides insights into the relative importance of the tail index in these models. We can use this definition to compare extreme event models. Consider empirical data points $(x_1, x_2, \ldots, x_n)$ and two extreme value models with their respective PDFs $f_1(x)$ and $f_2(x)$. If the KL divergence from the empirical data to the first model, denoted by $K(\hat{f};f_1)$, is smaller than the KL divergence from the empirical data to the second model, $K(\hat{f};f_2)$, then according to the KL divergence criterion, model 1 is considered to provide a superior fit to the data compared to model 2.

Proposition 1 (*Entropy Convergence of Block Maxima*). *Consider a sequence of independent and identically distributed random variables* $(X_1, X_2, \ldots, X_N)$ *with CDF* $F(x)$. *Define* $M_n = \max\{X_1, X_2, \ldots, X_n\}$ *as the block maxima. If the normalized maxima, denoted by* $Z_n = (M_n - b_n)/a_n$, *where* $a_n > 0$ *and* $b_n \in \mathbb{R}$, *converge in distribution to a non-degenerate distribution* $\mathcal{G}(z)$, *where* $\mathcal{G}(z)$ *is a generalized extreme value distribution, then the entropy of the block maxima* $H(M_n)$ *will converge to the entropy of generalized extreme value distribution* $H(\mathcal{G})$. *That is,*

$$\lim_{n\to\infty} H(M_n) = H(\mathcal{G}). \tag{5}$$

Proof. The PDF of the block maxima, denoted $f_{M_n}(x)$, can be expressed using the CDF $F(x)$ and its corresponding PDF $f(x)$, as

$$f_{M_n}(x) = nF(x)^{n-1}f(x).$$

By the Fisher–Tippett–Gnedenko theorem (Fisher and Tippett, 1928; Gnedenko, 1943; Brookes, 1955), there exist sequences of constants $a_n > 0$ and b_n such that the block maxima M_n, when normalized by a_n and b_n, converge in distribution to $\mathcal{G}$. That is,

$$\lim_{n\to\infty} P(Z_n \leq z) = \mathcal{G}(z).$$

Given the convergence in distribution, the PDF of the normalized block maxima, denoted $g_n(z)$, can be derived from $f_{M_n}(x)$ and approaches the PDF of the generalized extreme value distribution, denoted $g(z)$. This is due to the change of variable $x = a_n z + b_n$. The $H(M_n)$ can be calculated by

$$H(M_n) = -\int_{-\infty}^{\infty} f_{M_n}(x)\log(f_{M_n}(x))dx,$$

and $H(\mathcal{G})$ is calculated by

$$H(\mathcal{G}) = -\int_{-\infty}^{\infty} g(z)\log(g(z))dz.$$

To prove the convergence of the entropy, consider

$$\lim_{n\to\infty} H(M_n) = \lim_{n\to\infty}\left[-\int_{-\infty}^{\infty} g_n(z)\log(g_n(z))dz\right].$$

By applying the limit inside the integral (which can be justified by Lebesgue's Dominated Convergence Theorem), the limit of the entropy becomes

$$\begin{aligned}\lim_{n\to\infty} H(M_n) &= -\int_{-\infty}^{\infty} \lim_{n\to\infty}\left[g_n(z)\log(g_n(z))\right]dz \\ &= -\int_{-\infty}^{\infty} g(z)\log(g(z))dz = H(\mathcal{G}).\end{aligned}$$

Therefore, the entropy of the block maxima converges to the entropy of the generalized extreme value distribution as the block size increases.

This proposition provides a connection between extreme value theory and information theory by examining the convergence of entropy, and it may also be beneficial for anomaly detection.

Example 1 (*Entropy Convergence of Block Maxima*). To numerically demonstrate Proposition 1, a sequence of i.i.d. random variables are simulated from normal, exponential, and Pareto distributions ($N = 10{,}000$). The block maxima M_n are calculated for each simulation, and a generalized extreme value distribution is fitted to M_n. This fitting procedure provides us with the scale parameter a_n and the location parameter b_n. The block maxima are then normalized using these parameters, giving us the sequence Z_n. Next, a random sample is simulated from the generalized extreme value distribution with a_n and b_n. Lastly, the entropy of the block maxima $H(M_n)$ and the entropy of the simulated generalized extreme value distribution $H(\mathcal{G})$ are computed. Table 1 exhibits the convergence of $H(M_n)$ towards $H(\mathcal{G})$ as the size of the block n increases for normal, exponential, and Pareto distributions. Specifically, it can be observed that for all three distributions, the entropy of the block maxima increases as the block size n increases, indicating a higher degree of uncertainty. The entropy values become increasingly close to $H(\mathcal{G})$ as n increases. For instance, in the exponential distribution case, $H(M_n)$ for $n = 1000$ is 6.908, while $H(\mathcal{G}) = 9.210$. As n increases, $H(M_n)$ increases, becoming 8.004, 8.513, 8.849, and 9.205 for $n = 3000, 5000, 7000, 10,000$, respectively. Similar observations can be made for the normal and Pareto distributions.

Table 1
Entropy of block maxima ($H(\mathcal{G}) = 9.210$).

Distribution	$n = 1,000$	$n = 3,000$	$n = 5,000$	$n = 7,000$	$n = 10,000$
Normal	6.908	8.006	8.517	8.854	9.210
Exponential	6.908	8.004	8.513	8.849	9.205
Pareto	6.906	8.003	8.513	8.847	9.200

According to Proposition 1, $H(M_n)$ converges to $H(\mathcal{G})$ as $n \to \infty$. This simulation gives an empirical demonstration of this convergence.

Corollary 1. *If shape parameter ξ remains constant as $n \to \infty$, then the difference between the entropy of the block maxima $H(M_n)$ and the entropy of the generalized extreme value distribution $H(\mathcal{G})$ will tend to a constant c. That is,*

$$\lim_{n\to\infty} [H(M_n) - H(\mathcal{G})] = c. \tag{6}$$

The corollary suggests that the rate at which the entropy of the block maxima converges to the entropy of the generalized extreme value distribution is controlled by its shape parameter ξ. This is important for two reasons. First, the result further emphasizes the central role of the shape parameter ξ in extreme value theory. Specifically, ξ is known to control the tail heaviness of the generalized extreme value distribution, which determines the probabilities of extreme events. Therefore, this result shows that ξ governs the tail behavior and determines the rate at which the entropy of the block maxima (i.e., the uncertainty associated with the maximum value) converges to the entropy of the generalized extreme value distribution. Second, the corollary measures the discrepancy between the block maxima and the generalized extreme value distribution in terms of entropy, which quantifies the information content. A constant difference implies that the additional uncertainty introduced using the block maxima instead of the generalized extreme value distribution does not increase indefinitely as $n \to \infty$. This provides an upper bound on the additional uncertainty that the block maxima will introduce, which could be helpful in assessing the performance of extreme value models based on the block maxima method.

Proposition 2 (*Monotonicity of Block Maxima Entropy*)**.** *If the entropy of the block maxima $H(M_n)$ is a monotonically decreasing function of the block size n, then*

$$n_1 < n_2 \Rightarrow H(M_{n1}) \geq H(M_{n2}). \tag{7}$$

Proof. By the definition of a monotonically decreasing function, for any two block sizes n_1 and n_2 such that $n_1 < n_2$, the function value at n_1 will be greater than or equal to the function value at n_2. Therefore, under the condition that $H(M_n)$ is a monotonically decreasing function of n, for any two block sizes n_1 and n_2 such that $n_1 < n_2$, it will indeed hold that $H(M_{n1}) \geq H(M_{n2})$.

This could be useful for understanding the effect of the block size on the uncertainty in estimating extreme events.

Example 2 (*Monotonicity of Block Maxima Entropy*)**.** Consider a sequence of i.i.d. random variable from a standard normal distribution $N(0, 1)$ for different block sizes, $n = 100, 500, 1000, 5000$. For each simulation, the block maxima $M_n = \max\{X_1, X_2, \ldots, X_n\}$ are computed, and a generalized extreme value distribution is fitted to M_n. This provides us with the shape parameter ξ, the scale parameter a_n, and the location parameter b_n. Using these parameters, the block maxima are normalized, giving us the sequence Z_n. Fig. 1 illustrates the monotonicity of the entropy of the block maxima $H(M_n)$ as the block size n increases. The plot on the left visualizes the entropy of the block maxima against the block size. The entropy of the block maxima $H(M_n)$ declines with n, illustrating the monotonicity. For instance, the $H(M_n)$ for $n = 100$ is 6.907, while for $n = 500$, it increases to 5.298. As n increases to 1000 and 5000, $H(M_n)$ decreases to 4.605 and 2.995, respectively.

Table 2
Information-theoretic measures.

$K(\hat{f}; f_{\mathcal{G}})$	$K(\hat{f}; f_{\mathcal{P}})$	$H(\mathcal{G})$	$H(M_n)$
1.746	1.756	1.123	2.473

Block size n equals 500.

4. Evaluating financial risk in futures

This empirical analysis employs historical futures data to understand extreme events and their associated risks. The dataset used for this study comprises daily E-mini S&P 500 futures data procured from Bloomberg. This dataset spans from September 25, 2000, to May 23, 2023, and includes adjusted closing values. The E-mini S&P 500 futures are a derivatives product based on the S&P 500 index. These futures contracts are commonly traded due to their high liquidity, affordability, and accessibility. For analysis, we calculate the log returns, the natural logarithm of the ratio of consecutive closing prices. To confirm stationarity, the augmented Dickey–Fuller unit root test is applied. The test yields a significantly small p-value (less than 0.01), ensuring stationarity.

Table 2 presents KL divergence and entropy measures. The KL divergence measures the difference between our empirical distribution and two fitted distributions, $\mathcal{G}$ and $\mathcal{P}$. The smaller the KL divergence, the closer our fitted distribution is to the empirical one. For $\mathcal{G}$ and $\mathcal{P}$, the KL divergences are 1.746 and 1.756, respectively, suggesting a slightly better fit for the generalized extreme value distribution. The table also presents the entropy measures, quantifying the level of uncertainty in these distributions. Here, the entropy measures for the block maxima $H(M_n)$ and the generalized extreme value distribution $H(\mathcal{G})$ are reported. The entropy of the block maxima $H(M_n)$ represents the uncertainty associated with the maximum value of a given block size n. As the block size increases, $H(M_n)$ approaches $H(G)$. $H(G)$ is smaller than that of the block maxima $H(M_n)$ for $n = 500$. As the block size increases, $H(M_n)$ would decrease. In other words, with larger blocks, the maximum values' randomness would be very close to that modeled by the generalized extreme value distribution. This provides evidence for the suitability of the generalized extreme value distribution for modeling extreme events and highlights the important relationship between extreme value theory and information theory.

Table 3 presents the risk measures, VaR, and CVaR, at different confidence levels (99%, 95%, and 90%) for the fitted generalized extreme value distribution based on the E-mini S&P 500 futures. Moving to a lower confidence level (from 99% to 90%), we observe an increasing trend in VaR. This suggests we are accounting for increasingly severe potential losses as our confidence level decreases. At a 99% confidence level, the VaR is 1.695, meaning we do not expect to lose more than this value in 99% of the scenarios. As the confidence level drops to 95% and 90%, the VaR increases to 1.751 and 1.819, respectively. The CVaR follows a similar pattern. It increases from 1.676 at the 99% confidence level to 1.717 and

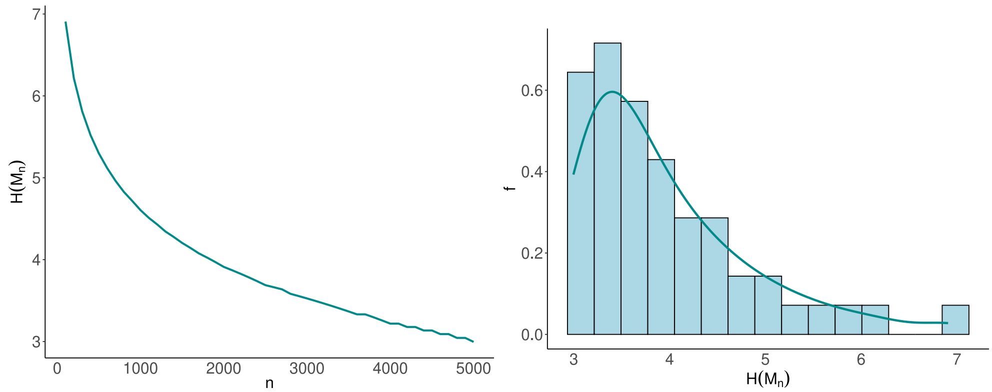

Fig. 1. Entropy of block maxima for a normally distributed sequence.

Table 3
Risk measures.

	99%	95%	90%
VaR	1.695	1.751	1.819
CVaR	1.676	1.717	1.752

1.752 at the 95% and 90% confidence levels, respectively. CVaR represents the average of potential losses exceeding the VaR, which measures the expected loss in the worst-case scenarios. The increasing trend of CVaR as we decrease the confidence level suggests that the expected loss given a loss beyond VaR, becomes more severe.

5. Concluding remarks

This study establishes a theoretical framework that bridges the gap between the principles of information theory and extreme value theory. Results show that the entropy of block maxima converges to the entropy of the generalized extreme value distribution and that the distribution's shape parameter modulates its convergence rate. The empirical analysis employs E-mini S&P 500 futures data to illustrate the theoretical results. The findings provide insights into enhancing risk assessments and open new research opportunities in extreme value modeling.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request

References

Ardakani, O.M., 2022. Option pricing with maximum entropy densities: The inclusion of higher-order moments. J. Futures Mark. 1–16.
Ardakani, O.M., 2023. Coherent measure of portfolio risk. Finance Res. Lett. 57, 104222.
Ardakani, O.M., Ebrahimi, N., Soofi, E.S., 2018. Ranking forecasts by stochastic error distance, information and reliability measures. Internat. Statist. Rev. 86 (3), 442–468.
Bali, T.G., 2003. The generalized extreme value distribution. Econom. Lett. 79 (3), 423–427.
Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. J. Financ. Econ. 99, 427–446.
Balkema, A.A., Haan, L.d., 1974. Residual life time at great age. Ann. Probab. 2 (5), 792–804.
Box, G.E., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2), 211–243.
Brookes, B.C., 1955. Statistical theory of extreme values and some practical applications. by EJ Gumbel. Pp. viii, 51. 40 cents. 1954.(National Bureau of Standards Applied Mathematics Series, 33; Washington, DC).. Math. Gaz. 39 (330), 341–342.
Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.
Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quant. Finance 1, 223–236.
Cover, T.M., Thomas, J.A., 1991. Information Theory and Statistics: Elements of Information Theory. Wiley.
Danielsson, J., 2011. Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab. Wiley.
Danielsson, J., Haan, L.d., Peng, L., Vries, C.d., 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. J. Multivariate Anal. 76 (2), 226–248.
Diebold, F.X., Schuermann, T., Stroughair, J.D., 2000. Pitfalls and opportunities in the use of extreme value theory in risk management. J. Risk Finance 1 (2), 30–35.
Embrechts, P., Kluppelberg, C., Mikosch, T., 2013. Modelling Extremal Events: For Insurance and Finance. Springer.
Fabozzi, F.J., Rachev, S.T., Menn, C., 2005. Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing. Wiley.
Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 24. pp. 180–190.
Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Mat. 44 (3), 423–453.
Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106, 620–630.
Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q. J. R. Meteorol. Soc. 81 (348), 158–171.
Kullback, S., Leibler, R., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79 86.
Longin, F.M., 2000. From value at risk to stress testing: The extreme value approach. J. Bank. Financ. 24 (7), 1097–1130.
Mandelbrot, B., Taylor, H.M., 1967. On the distribution of stock price differences. Oper. Res. 15, 985–1188.
McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. J. Empir. Financ. 7 (3–4), 271–300.
Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 119–131.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423.
Soofi, E.S., 2000. Principal information theoretic approaches. J. Amer. Statist. Assoc. 95, 1349–1353.