

Modeling Expected Shortfall Using Tail Entropy

Daniel Traian Pele ^{1,*}, Emese Lazar ² and Miruna Mazurencu-Marinescu-Pele ¹

¹ Department of Statistics and Econometrics, Faculty of Cybernetics, Statistics and Economic Informatics, The Bucharest University of Economic Studies, Piata Romana, nr.6, Sector 1, 010371 Bucharest, Romania; miruna@ase.ro

² Henley Business School, University of Reading, ICMA Centre, Whiteknights, Reading RG6 6BA, UK; e.lazar@icmacentre.ac.uk

* Correspondence: danpele@ase.ro; Tel.: +40-21-319-19-01 (ext. 383)

Received: 10 October 2019; Accepted: 5 December 2019; Published: 7 December 2019

Abstract: Given the recent replacement of value-at-risk as the regulatory standard measure of risk with expected shortfall (ES) undertaken by the Basel Committee on Banking Supervision, it is imperative that ES gives correct estimates for the value of expected levels of losses in crisis situations. However, the measurement of ES is affected by a lack of observations in the tail of the distribution. While kernel-based smoothing techniques can be used to partially circumvent this problem, in this paper we propose a simple nonparametric tail measure of risk based on information entropy and compare its backtesting performance with that of other standard ES models.

JEL Classification Codes: C14; C22; G10

Keywords: expected shortfall; measure of risk; information entropy; tail risk

1. Introduction

Past banking regulation has concerned and current banking regulation concerns the measurement of risk needed for capital calculation, which can be computed based on various models. Additionally, banks and financial institutions need to estimate their own exposure to different relevant risk factors in order to understand the overall riskiness of their activities and to help them prepare for undesirable situations. While value-at-risk (VaR) is the most popular risk measure and has been widely used in the past, it is not a subadditive measure (Artzner et al. [1,2]) and it is a misleading measure for portfolio optimizers [3]. Expected shortfall (ES), which has been popularized by Artzner et al. [1,2], having the advantage of being a coherent measure [4], has increased in importance since the Basel Committee on Banking Supervision ([5,6]) began to base its regulatory capital calculations on ES estimates at the 2.5% significance level.

The standard way to estimate ES at level α is to simply compute the weighted average of the historical returns in the α -tail of the returns' distribution, where the weights are the probabilities associated with the returns estimated from historical data (nonparametric method). However, this often gives an estimate that is inaccurate (see McNeil et al. [7]), data-sensitive (due to the low number of observations in the tail), and often characterized by an undesirably large variability.

A comprehensive presentation of the modern techniques used in risk management, especially with regard to estimating value-at-risk, expected shortfall, and spectral risk measures, can be found in Guégan et al. [8].

Several alternatives exist to computing ES, and a comprehensive review can be found in Nadaraj et al. [9]. Starting with semiparametric methods, a strand of this approach is based on extreme value theory, as in McNeil and Frey [10], who apply the theoretical framework of Embrechts et al. [11] and employ the generalized Pareto distribution combined with an ARMA-GARCH(1,1) process. In a later paper, Embrechts et al. [12] propose an ES estimation method based on the Hill estimator. Another

semiparametric ES estimation method, using expectiles, has been proposed by Taylor [13] based on an asymmetric regression methodology of Aigner et al. [14] and the CaViaR model of Engle [15]. A smoothing technique, detailed in Schnabel and Eilers [16] and based on the algorithm of Schall [17], can be employed to improve the performance of the ES estimate. Daouia et al. [18] consider expectile-based ES estimation for extreme percentiles and study their asymptotic behavior.

There are many parametric methods used to estimate ES. Nadaraj et al. [9,19] enlist more than 100 approaches, including methods that make different distributional assumptions (Gaussian, Johnson family, Student's t , stable distributions, mixture distributions, generalized Pareto and elliptical distributions, etc.)—see also Broda and Paoletta [20]. Many methods employ different econometric formulations (like GARCH, as in a recent paper by Krause and Paoletta [21], or ARMA, or even Bayesian approaches); these methods are mostly applied to historical data, but they can be applied within a Monte Carlo simulation setup as well. All these parametric approaches make an assumption of a particular model, so they bear significant model risk.

Chen [22] argues that semiparametric or parametric tail-risk estimation models struggle to specify an adequate model for the tail distribution (due to a lack of data), meaning nonparametric models are preferred. In terms of nonparametric ES estimation (which has the advantage of not employing a particular model and hence not being affected by misspecification risk), in addition to the classical historical ES estimation (which is considered the gold standard), the estimator of Yamai and Yoshihara [3], the estimator of Inui and Kijima [23], the ES estimator of Chen [22], and the methodology of Jadhav et al. [24], which has been shown to perform better than an historical estimator, have all been proposed. Kernel methods based on a kernel-smoothed estimator, as in Scaillet [25], can be used as well, but as shown by Chen [22], smoothing negatively affects the accuracy of ES estimates as it introduces a bias and does not reduce the variance of the estimates. To correct this, Alemany et al. [26] have shown that a double transformed kernel estimator can, in some cases, be a good alternative. Cai and Wang [27] suggest a nonparametric ES estimator using weighted double kernels and study its finite sample performance. The trimmed kernel method of Hill [28] is a bias-corrected estimator with a small error only. Another alternative is provided by the method of Richardson [29] and the algorithm of Fan et al. [30], which reduce the bias but not the error of the estimation (see Inui and Kijima [23]). A two-stage estimation method based on extreme value theory has been proposed by Martins-Filho et al. [31]. In this paper we propose a nonparametric estimator based on tail entropy (TE) and compare it with alternatives.

As entropy is a measure of uncertainty, which is in some ways similar to volatility, it is widely used in physics and the social sciences. In general, the higher the uncertainty of a system, the larger its entropy. The central theme of this paper is risk, which is directly linked to the tail of the distribution of returns (as opposed to its full distribution), so it makes sense to compute entropy based solely on returns that determine risk (that is, the tail of the distribution), which then gives tail entropy, as discussed in detail in this paper. Roughly speaking, tail entropy is defined as the weighted sum of log probabilities in the tail, where the weights are the conditional tail probabilities themselves.

Furthermore, it has to be noted that there is a strong similarity between definitions of ES and tail entropy. While ES computes the weighted average of tail returns, tail entropy calculates the weighted average of the log probabilities of the returns (again, in the tail of the distribution), but the two calculations use the same weights (probabilities). As such, tail entropy is an estimate of tail risk, which is free of the main problem with ES, namely, its direct dependence on the actual observations in the tail, hence the motivation of our paper. Our aim was to study how tail entropy can be used as an efficient measure of risk and how it compares with alternative estimates of ES.

It is natural to compare the theoretical and empirical properties of entropy with those of variance or volatility, with both being measures of uncertainty. While volatility is often considered a measure of risk in finance, here we consider it a measure of uncertainty. To differentiate between the two types of measures, i.e., measures of risk and measures of uncertainty, measures that are symmetric by nature are measures of uncertainty, and (asymmetric) tail measures are considered measures of risk. Thus, tail entropy is a measure of risk. Philippatos and Wilson in 1972 [32], and, later, Ebrahimi et al. [33] suggested that entropy compares favorably to volatility/variance as a measure of uncertainty.

The same has been shown by Dionisio et al. [34], with the authors arguing that while variance measures the concentration around the mean, entropy measures the dispersion of the density irrespective of the location of the concentration (also see Ebrahimi et al. [35] and Allen et al. [36]). Moreover, the entropy of a distribution function is strongly related to its tails, as shown by Pele et al. [37], which becomes an important feature for distributions with heavy tails or with an infinite second-order moment for which an estimator of variance is obsolete. Liu et al. [38] compare an entropy-based measure with the classical coefficient of correlation and conclude that their measure has certain superior characteristics as a descriptor of the relationship between time series when compared with the correlation measure.

The study of the relationship between entropy and financial markets goes back more than a decade. Entropy has been used as a measure of stock market efficiency in Zunino et al. [39], while Risso [40] links it to stock market crashes, arguing that the probability of having a crash increases as entropy decreases. Oh et al. [41] use entropy as a measure of the relative efficiency of the foreign exchange (FX) markets. Based on generalized entropy theory, Wang et al. [42] analyze the interactions among agents of stock markets. Entropy has also been used as a tool for the predictability of stock market returns. Maasoumi and Racine [43] show that entropy can capture nonlinear dependence between financial returns. An effective early warning indicator for crisis situations of banking systems has been built by Billio et al. [44]; this indicator considers different definitions of entropy.

Entropy has also been used in option pricing (Stutzer [45] and Stutzer and Kitamura [46]). An application of entropy-based risk measures for decision-making can be found in Yang and Qiu [47]. Applications of Tsallis entropy in risk management can be found in Gradojevic and Gencay [48], Gencay and Gradojevic [49], and Gradojevic and Caric [50]. Bowden [51] introduces directional entropy and uses it to improve the performance of VaR in capturing regime changes. Portfolio optimization based on maximum entropy has been discussed by Geman et al. [52].

The main objectives of this paper were to introduce a measure for expected shortfalls based on tail entropy, to study this measure's properties, and to compare it with alternative measures of ES. The main advantage of the measure we propose is that it is less sensitive to the actual values of observations in the tail of the distribution (compared to historical ES), making it a more stable measure of tail risk. We use several ES backtests to verify the accuracy of the proposed measure. The rest of the paper is organized as follows. Section 2 derives the theoretical results specifying the steps in estimating tail entropy as a measure of market risk and details our methodology, Section 3 presents an empirical application, and Section 4 concludes.

2. Methodology

For a series of asset prices denoted by S_t , in the following we compute and denote log-returns by X with realizations $(X_t)_{t \in \mathbb{N}}$, i.e.,

$$X_{t+1} = \log(S_{t+1}/S_t). \quad (1)$$

2.1. Measures of Risk: VaR and ES

Given the returns series X , with F denoting its cumulative density (or distribution) function, for an $\alpha \in (0,1)$ the VaR at level α is the smallest number with cumulative density at least as big as α , i.e.,

$$VaR_\alpha(X) = -\inf\{x \mid F(x) \geq \alpha\}. \quad (2)$$

In probabilistic terms we can write that

$$P(X < -VaR_\alpha(X)) = \alpha. \quad (3)$$

In addition, VaR is the negative of the quantile, i.e.,

$$VaR_\alpha(X) = -F_\alpha^{-1}. \tag{4}$$

The ES at significance level α is the average of the returns below the VaR at level α , as given by

$$ES_\alpha(X) = -E\left[x \mid x \leq -VaR_\alpha(X)\right] = \frac{1}{\alpha} \int_0^\alpha VaR_\gamma(X) d\gamma. \tag{5}$$

For a continuous variable with density function f , the above definition for ES can be rewritten as

$$ES_\alpha(X) = -\frac{1}{\alpha} \left(\int_{-\infty}^{-x_\alpha} xf(x) dx + x_\alpha (\alpha - P[X \leq x_\alpha]) \right), \tag{6}$$

where $x_\alpha = \inf\{x \in R \mid P(X \leq x) \geq \alpha\}$ is the left α -quantile.

2.2. Entropy

Entropy is a measure of uncertainty, which is in some ways similar to volatility. Various definitions exist (e.g., Shannon entropy, Tsallis entropy, and Kullback Cross entropy etc.) based on the informational content of a discrete or continuous random variable (see Zhou et al. [53] for a comprehensive review of entropy measures used in finance). Shannon information entropy is the most commonly used definition of entropy; it quantifies the expected value of the information contained in a discrete distribution.

Definition 1 (Shannon information entropy). For X , a discrete random variable with probability distribution $X : \begin{pmatrix} x_1 \dots x_n \\ p_1 \dots p_n \end{pmatrix}$, with $p_i = P(X = x_i)$, $0 \leq p_i \leq 1$, and $\sum_i p_i = 1$, the Shannon information entropy is defined as

$$H(X) = -\sum_i p_i \log_2 p_i. \tag{7}$$

For a discrete distribution, the entropy reaches its maximum value of $H(X) = \log_2 n$ for the uniform distribution when all the p_i values are the same (i.e., when there is a high level of uncertainty). Similarly, the entropy will reach its minimum value of 0 for a distribution with zero uncertainty (i.e., one of the probabilities p_i is 1 and the rest are all 0).

For X , a continuous random variable with probability density function $f(x)$, the differential entropy is given by

$$H(f) = -\int_A f(x) \log_2 f(x) dx, \quad A = \text{supp}(X). \tag{8}$$

However, unlike with Shannon entropy, differential entropy does not possess certain desirable properties like invariance to linear transformations and non-negativity (Lorentz [54] and Pele [55]). A measure of entropy similar to Shannon entropy can be defined via a transformation called quantization, as defined below [54].

Definition 2 (sampled function). For $f : I = [a, b] \rightarrow \mathbf{R}$, a real valued continuous function, $n \in \mathbf{N}^*$, a fixed number, and $x_i = a + (i + 1/2)h$, $i = 0, \dots, n-1$, where $h = (b - a) / n$, the sampled function for f is given by (point sampling)

$$S_n(f)(i) = f(x_i), \quad \text{for } i = 0, \dots, n-1. \tag{9}$$

For $f : I = [a, b] \rightarrow \mathbf{R}$, which is essentially bounded, the sampled function is (mean sampling)

$$S_n(f)(i) = h^{-1} \int_{x_i-h/2}^{x_i+h/2} f(x)dx \text{ for } i = 0, \dots, n-1. \tag{10}$$

Definition 3 (quantization). The quantization of a function creates a simple function that approximates the original one. Given $q > 0$, a quantum, the following function defines a quantization of f .

$$Q_q(f)(x) = (i+1/2)q, \text{ if } f(x) \in [iq, (i+1)q). \tag{11}$$

Definition 4 (entropy of a function at quantization level q). Let f be a measurable and essentially bounded real valued function defined on $[a, b]$ and let $q > 0$. Let $I_i = [iq, (i+1)q)$ and $B_i = f^{-1}(I_i)$. Then, the entropy of f at quantization level q is given by

$$H_q(f) = -\sum_i \mu(B_i) \log_2(\mu(B_i)), \tag{12}$$

where μ denotes the Lebesgue measure.

Lorentz’s theorem given below calculates the entropy of a continuous function on a compact interval. Eventually, it helps define the entropy of a probability distribution function for a continuous random variable on a compact interval, regardless of the sampling and quantization.

Theorem (Lorentz [54]). Let f be continuous for point sampling, measurable, and essentially bounded for mean sampling. The sampling spacing is $1/n$. Let $S_n(f)$ be the corresponding sampling; fix $q > 0$ and let $Q_q S_n$ be the quantization of the samples with resolution q as in Definition 3. The number of occurrences of $(i+1/2)q$ in $Q_q S_n$ is $c_n(i) = \text{card}\{(i+1/2)q \in Q_q S_n\}$ and the relative probability of the occurrence

of the value i is denoted by $p_n(i) = \frac{c_n(i)}{\sum_j c_n(j)} = \frac{c_n(i)}{n}$. Then, we have

$$\lim_{n \rightarrow \infty} -\sum_i p_n(i) \log_2 p_n(i) = H_q(f). \tag{13}$$

In the following we assume that we are dealing with a continuous random variable X whose support is the set A with the distribution function $F : \mathbf{R} \rightarrow [0, 1]$, $F(x) = P(X < x)$. If the distribution function is absolutely continuous, then there exists a non-negative, Lebesgue integrable function f which is a probability density function (PDF), i.e.,

(i) $F(b) - F(a) = P(a < X < b) = \int_a^b f(x)dx$;

(ii) $\int_A f(x)dx = 1$;

(iii) $F(x) = \int_{-\infty}^x f(t)dt$.

Assuming that the probability density function f is essentially bounded, we can define the entropy of the probability density function $H_q(f)$ at the quantization level $q > 0$.

Definition 5 (entropy of a probability density function at quantization level q). For a probability density function $f : \mathbf{R} \rightarrow [0, \infty)$ which is essentially bounded, let $I_i = [iq, (i+1)q)$ disjoint giving $\bigcup_i I_i = [0, 1]$ and $B_i = f^{-1}(I_i)$ not necessarily disjoint, with $\bigcup_i B_i = [0, 1]$. Then, the entropy of f at quantization level q is $H_q(f) = -\sum_i \mu(B_i) \log_2(\mu(B_i))$, where μ is the Lebesgue measure.

Of note is that in general, for a probability density function defined on set A not necessarily of finite measure, it is possible to consider this function's restriction on a compact interval, i.e.,

$$f|_{[a,b]} : [a,b] \rightarrow \text{Im}(f|_{[a,b]}), f|_{[a,b]}(x) = f(x).$$

Then, $f|_{[a,b]} : [a,b] \rightarrow \text{Im}(f|_{[a,b]})$ satisfies the conditions of Lorentz's theorem, meaning the entropy can be defined.

The above framework can be applied to estimation of the entropy of a probability density function of a continuous random variable X .

Next, we present the estimation algorithm of the entropy of a probability density function. Let X_0, \dots, X_{n-1} be a sample of an *i.i.d.* variable with probability density function f . The Algorithm 1 (see below) estimates the entropy of a probability density function (see Pele et al. [37]).

Algorithm 1. Estimation of the entropy of a probability density function.

- I. Estimate the probability density function, obtaining values $\hat{f}_n(X_i)$ for $i = 0, \dots, n-1$;
 - II. Sample from the probability density function using the sampled function $S_n(\hat{f}_n)(i) = f_n(X_i)$ for $i = 0, \dots, n-1$;
 - III. Define a quantum $q > 0$; then $Q_q S_n(\hat{f}_n)(j) = (i+1/2)q$, if $\hat{f}_n(X_j) \in [iq, (i+1)q)$;
 - IV. Compute the probabilities $p_n(i) = \frac{c_n(i)}{\sum_j c_n(j)} = \frac{c_n(i)}{n} = \frac{\text{card}\{\hat{f}_n(X_j) \in [iq, (i+1)q)\}}{n}$;
 - V. Estimate the entropy of the probability density function, i.e., $H_q(\hat{f}_n) = -\sum_i p_n(i) \log_2 p_n(i)$
-

A similar approach can be found in Miśkiewicz [56], where two possible algorithms for estimating entropy of a continuous probability density function are discussed.

The entropy of the probability density function reaches its maximum value for the uniform distribution. A dimensionless measure of uncertainty, normalized entropy, can be defined as the ratio between the entropy of the probability density function and the entropy of the uniform distribution, i.e.,

$$NH_q(\hat{f}_n) = \frac{-\sum_i p_n(i) \log_2 p_n(i)}{\log_2 n} \in [0, 1]. \tag{14}$$

In the following we refer to the entropy of the probability density function as the normalized entropy of the probability density function, i.e., $H_q(f) \equiv NH_q(f) \in [0, 1]$.

2.3. Tail Entropy

Similarly to the definition of entropy, we can define tail entropy as being when entropy is computed using observations in the tail only. Shannon tail entropy is defined below.

Definition 6 (Shannon tail entropy). For X , a discrete random variable with probability distribution $X : \begin{pmatrix} x_1 \dots x_n \\ p_1 \dots p_n \end{pmatrix}$, with $p_i = P(X = x_i)$, $0 \leq p_i \leq 1$, and $\sum_i p_i = 1$, and α -level value-at-risk $VaR_\alpha(X)$ for $0 < \alpha < 1$, the Shannon tail information entropy at level α is defined by

$$H_\alpha(X) = - \sum_{x_i < VaR_\alpha(X)} p_i^* \log_2 p_i^* \tag{15}$$

The tail-adjusted probabilities are denoted by $p_i^* = p_i / \alpha$, and this normalization is required to make sure that the probabilities in the tail add up to 1. For a discrete distribution, the TE will reach its maximum value for a distribution which has uniform distribution in the α -tail, with all tail observations having the same probability. Similarly, for a given α , the tail entropy will reach its minimum for a distribution with zero uncertainty in the tail (an observation in the tail has a probability of $p_i = \alpha$ and the rest of the tail observations have a probability of 0).

For X , the discrete random variable given above, we can compute the expected shortfall based on the variable's probability distribution at level α ; this is denoted by ESPD and given by

$$ESPD_\alpha(X) = \sum_{x_i < VaR_\alpha(X)} p_i^* x_i \tag{16}$$

The similarity between Formulas (15) and (16) is obvious. There is a minus sign in Formula (15) to make sure the tail entropy is a positive measure. More importantly, while the entropy is a weighted sum of log probabilities, the ES is the weighted sum of observations in the tail. To visualize the difference between the two formulae, in Figure 1 we compare the left tail of a histogram of a variable distributed $N(0,1)$ (normalized with the tail probability and using bins of width 1) with the left tail of a histogram built using the entropy measure. In Formula (15) we can denote $\log_2 p_i^*$ by x_i , giving $p_i^* = 2^{-x_i}$, and this probability is presented in Figure 1 in the entropy histogram. It can be seen that the probabilities calculated using entropy are higher (except for when using the value -1) than the probabilities in the normal histogram.

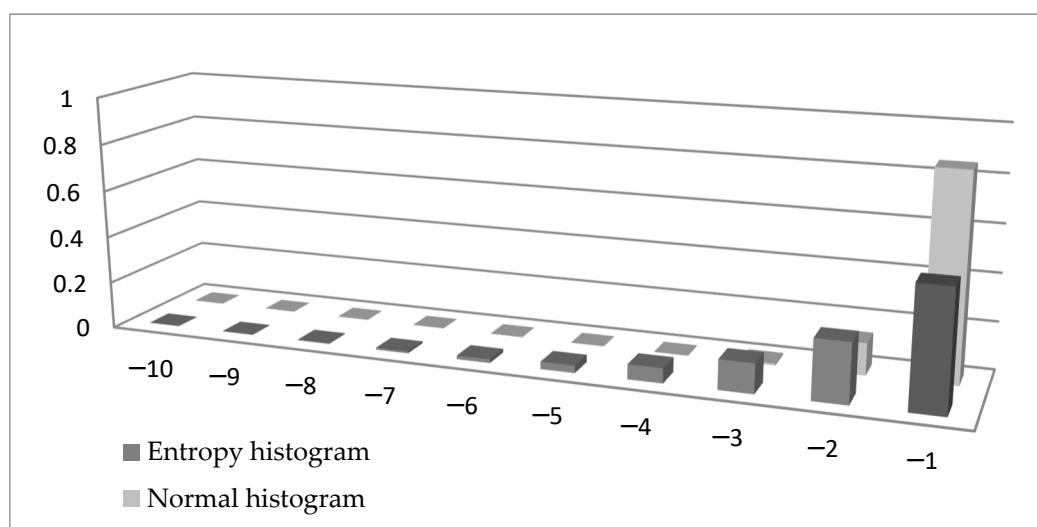


Figure 1. Comparison between a normal histogram and a histogram built using entropy.

For a continuous random variable with density function $f(x)$, the differential tail entropy is given by

$$H_\alpha(f) = - \int_A f(x) \log_2 f(x) dx, \quad A = \text{supp}(X) \cap \{x < -\text{VaR}_\alpha(x)\}. \quad (17)$$

Similarly to differential entropy, tail differential entropy does not possess the properties of invariance to linear transformations and non-negativity. Using quantization, a measure of entropy similar to Shannon tail entropy can be defined as follows.

Definition 7 (tail entropy of a probability density function at quantization level q). For the

probability density function $f : [0,1] \rightarrow \text{Im}(f)$, let $I_i = [iq, (i+1)q)$ disjoint giving $\bigcup_{i=0}^n I_i = [0,1]$ and

$B_i = f^{-1}(I_i)$ not necessarily disjoint, with $\bigcup_{i=0}^n B_i = [0,1]$. Then, the tail entropy of f at level α with quantum q (using the notation μ^* for the Lebesgue measure divided by α and $[\cdot]$ for the integer part of a number) is

$$H_{\alpha,q}(f) = - \sum_{i=0}^{[\frac{\alpha}{q}]} \mu^*(B_i) \log_2(\mu^*(B_i)). \quad (18)$$

Of note is that as in Definition 5, for a probability density function defined on set A which is not necessarily of finite measure, we can consider its restriction on a compact interval, i.e.

$$f|_{[a,b]} : [a,b] \rightarrow \text{Im}(f|_{[a,b]}), \quad f|_{[a,b]}(x) = f(x). \quad (19)$$

The tail entropy of the returns' distribution can be interpreted as the expected shortfall of the returns, if in the tail of the distribution the probability of losses can be expressed as an exponential function of the size of the losses. The tail entropy is a measure of uncertainty in the tail, with low values of tail entropy being associated with a lower risk in the tail.

This framework can be used to estimate the tail entropy of a distribution function of a continuous random variable X , defined on the support set of X , with values on $[0, 1]$. The distribution function can be estimated using the histogram estimator of a probability density or kernel density estimation methods as described in the previous section. We proceed by presenting an estimation algorithm of the tail entropy of a distribution function.

Let X_0, \dots, X_{n-1} be a sample of an *i.i.d.* variable with probability density function f . The Algorithm 2 (see below) estimates the tail entropy of a distribution function.

Algorithm 2. Estimation of the tail entropy of a distribution function

- I. Estimate the probability density function, obtaining values $\hat{f}_n(X_i)$ for $i = 0, \dots, n-1$;
 - II. Sample from the probability density function using the sampled function $S_n(\hat{f}_n)(i) = f_n(X_i)$ for $i = 0, \dots, n-1$;
 - III. Define a quantum $q > 0$; then $Q_q S_n(\hat{f}_n)(j) = (i+1/2)q$, if $\hat{f}_n(X_j) \in [iq, (i+1)q)$;
 - IV. Compute the probabilities $p_n^*(i) = \frac{c_n(i)}{\alpha \sum_j c_n(j)} = \frac{c_n(i)}{\alpha n} = \frac{\text{card}\{\hat{f}_n(X_j) \in [iq, (i+1)q)\}}{\alpha n}$;
-

$$V. \quad \text{Estimate the tail entropy of } F \text{ as } H_{\alpha,q}(\hat{f}_n) = -\sum_{i=0}^m p_n^*(i) \log_2 p_n^*(i) \text{ where } m = \left\lceil \frac{\alpha}{q} \right\rceil.$$

The tail entropy of a distribution function reaches its maximum value for a distribution which has uniform distribution in the α -tail.

Property (maximum value of the tail entropy of a probability density function): *The α -tail entropy of a distribution function $F : [0, 1] \rightarrow [0, 1]$ reaches its maximum for a distribution with $F(x) = x$ for $x \leq \alpha$, and this maximum value is $\log_2 m$.*

Proof. This is similar to the proof of the maximum Shannon entropy. Indeed, when all probabilities are equal for the observations in the α -tail ($= 1/m$), the tail probability can be written as:

$$H_{\alpha,q}(\hat{f}_n) = -\sum_{i=0}^m p_n^*(i) \log_2 p_n^*(i) = -\sum_{i=0}^m q \log_2 q = -\log_2 q = \log_2 m. \tag{20}$$

□

As such, a dimensionless measure of uncertainty, the normalized tail entropy, can be defined as the ratio between the entropy of the probability density function and its maximum value, i.e.,

$$NH_{\alpha,q}(\hat{f}_n) = \frac{-\sum_{i=0}^m p_n^*(i) \log_2 p_n^*(i)}{\log_2 m} \in [0, 1]. \tag{21}$$

We refer below to the tail entropy of the probability density function as the normalized tail entropy of the probability density function: $H_{\alpha,q}(\hat{f}) \equiv NH_{\alpha,q}(f) \in [0, 1]$.

In practical applications, the estimated tail entropy may be severely biased for small samples (Liu et al. [57]). In order to correct for small sample bias we propose a bootstrapping method, following DeDeo et al. [58]. Thus, we generate B independent samples of volume k from a multinomial distribution with probabilities $\hat{\mathbf{p}} = (p_n^*(0), \dots, p_n^*(m))$, and for each sample we estimate the normalized tail entropy $H_{\alpha,q}^b(\hat{f})$, $b = 1 \dots B$. The unbiased estimator of the normalized tail entropy has the form

$$H_{\alpha,q}^{bootstrap}(\hat{f}) = 2H_{\alpha,q}(f) - \frac{1}{B} \sum_{b=1}^B H_{\alpha,q}^b(\hat{f}). \tag{22}$$

The estimated entropy can be influenced by the quantum value $q > 0$; as there is no canonical quantum/level of quantization, apparently all results depend on an arbitrarily chosen quantum/quantization level. However, the Lorentz theorem ([54]) shows that the entropy estimator is consistent regardless of the choice of q ; in practical applications, such as stock market data, we have used $q = 0.2$ (see Section 3). For $q = 0.2$, the number of bins used to compute the tail entropy is $1/q = 5$; a smaller value of the quantum q will increase the number of bins and also the likelihood of having zero probability bins. For example, when the window used for estimation has 1000 observations (roughly 4 years of daily observations) and $\alpha = 5\%$, there are 50 observations in the left tail of the distribution; by decreasing the quantum level and increasing the number of bins, there is a high chance of having bins with zero probability where the term $p \log p$ is undefined.

Unlike expected shortfall, which is not well-defined when the expectation fails to exist (for instance for the Pareto family when the parameter $a \leq 1$), tail entropy does exist for the Pareto

distribution for any tail parameter value a . In general, under the hypothesis of the Lorentz theorem ([54]), tail entropy exists for any probability density function with bounded support. Even if the probability density function is not bounded, we can consider its restriction on a compact interval in order to fulfil the conditions of the Lorentz theorem. Indeed, tail entropy is sensitive to the choice of the level of the quantum, but in the next section we show how it can be transformed (via normalizing it and using a linear adjustment) into an expected-shortfall-type measure which will be less sensitive to the choice of quantum.

2.3.1. Tail Entropy Expected Shortfall

Next, we turn our attention to the link between tail entropy and measures of risk and uncertainty, seeking to propose a tail-entropy-enhanced ES estimator. First, we perform a Monte Carlo experiment, estimating and comparing tail entropy and historical expected shortfall for several simulated Pareto-like distributions; we report our results for different tail probabilities. Here we use the random variable $Y = -X \sim \text{Pareto}(a)$, which has the following characteristics:

(i) Cumulative distribution function: $F(x) = P(X < x) = 1 - \left(\frac{c}{x}\right)^a, x > c > 0.$

(ii) Probability density function: $f(x) = \frac{ac^a}{x^{a+1}}, x > c.$

(iii) $P(X > x) = \left(\frac{c}{x}\right)^a = Cx^{-a}.$

Figure 2 presents the probability density function of the random variable Y for different values of the a parameter. The a parameter (the tail index) is a measure of tail probability: higher values of a correspond to lower probability in the tail.

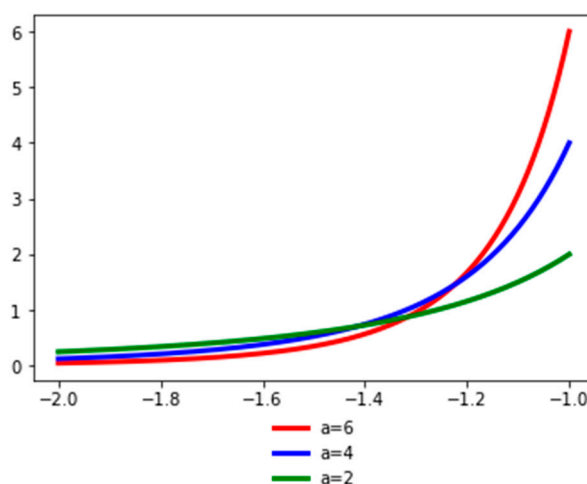


Figure 2. Probability density function for Y . See the Github page [SFM Sim Pareto like](#).

In order to assess the relationship between the tail entropy and the historical expected shortfall (ESH), we simulate random variables $Y_k = -X_k \sim \text{Pareto}(a_k)$, with $a_k = 2 + 0.1k, k = 1 \dots 180$. Figure 3 presents the relationship between TE and tail index a using scatterplots for different levels, i.e., 1% and 2.5%, of α . As the a parameter of the simulated distribution decreases, the tail entropy of the distribution decreases, too; as expected, high tail entropy values are associated with heavy-tailed distributions. Figure 4 presents the relationship between TE and historical ES, as functions of the a parameter of the simulated distribution, for 1% and 2.5% levels of α .

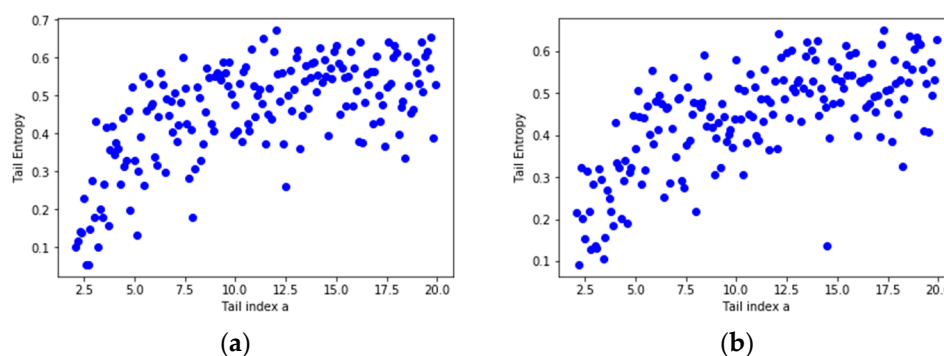


Figure 3. Tail Entropy and tail index a of the simulated distributions: (a) $\alpha = 1\%$ and (b) $\alpha = 2.5\%$. See the Github page [SFM ES Pareto](#).

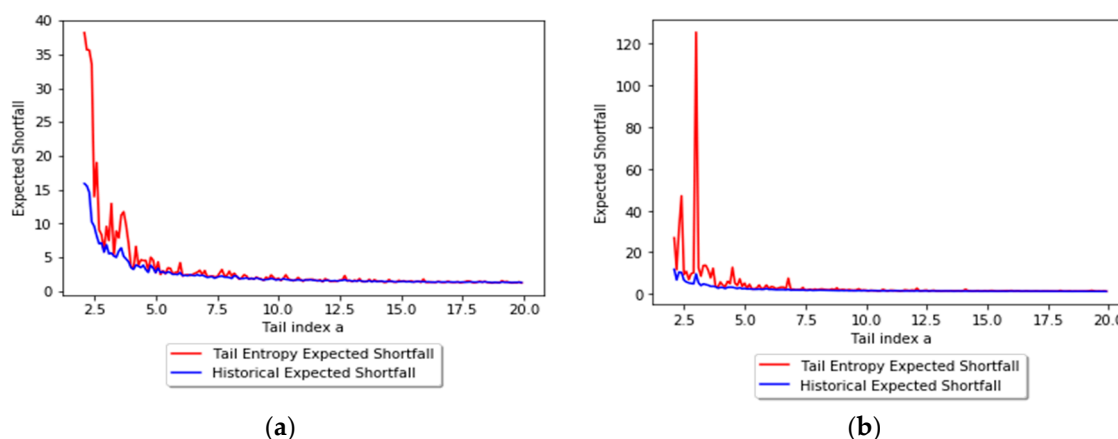


Figure 4. Tail entropy and historical expected shortfall of the simulated distributions: (a) $\alpha = 1\%$ and (b) $\alpha = 2.5\%$. See the Github page [SFM ES Pareto](#).

Additionally, we regress the historical ES estimates of the simulated returns above on tail entropy; the results are reported in Table 1. We estimate the historical ES using the formula

$$ESH_{\alpha}(X) = -E\left[x \mid x \leq -VaR_{\alpha}(X)\right]. \tag{23}$$

Table 1. Relationship between historical ES (ESH) and tail entropy (TE) for simulated distributions.

Dependent Variable: $ESH_{\alpha,t}$		
Significance Level	$\alpha = 1\%$	$\alpha = 2.5\%$
$H_{\alpha,q;t}$	-12.95 *** [0.94]	-8.32 *** [0.64]
R^2_{adj}	0.51	0.48

Note: standard errors are given in brackets. *** denotes significance at 1%.

It can be seen that there is a strong linear negative relationship between the tail entropy and ESH, with higher values of tail entropy corresponding to lower values of expected shortfall. The validity of this regression approach depends on the tail behavior of the underlying distribution; without this explicit assumption, one would need regularly varying tails and a particular tail behavior in order to make the regression approach sound. Indeed, for some distributions the relationship between historical ES and tail entropy might not be exactly linear, but we show below how assuming a linear relationship helps us define a new measure of expected shortfall.

A multiple regression with several tail entropies of varying tail levels was also estimated in order to explicitly assess the relationship between historical ES and TE for a Pareto-like simulated distribution, as in the above examples.

Thus, we estimated the following regression model, where the historical ES (1% and 2.5%) is regressed against the tail entropy for significance levels 1%, 2.5%, and 5%, i.e.,

$$ESH_{\alpha;t} = \gamma_0 + \gamma_1 H_{0.01,q;t} + \gamma_2 H_{0.025,q;t} + \gamma_3 H_{0.05,q;t} + \eta_t \tag{24}$$

The results of the regression as calculated using Formula (23) are reported in Table 2. By including varying tail levels for entropy, the performance of the model is significantly improved, with R^2_{adj} increasing for $\alpha = 1\%$ from 51% (see Table 1, left panel) to 55%, and for $\alpha = 2.5\%$ from 48% (see Table 1, right panel) to 60%.

A second simulation using α -stable distributions can help understand this behavior. We chose α -stable distributions because they have several interesting properties: they allow for heavy tails and any linear combination of independent stable variables follows a stable distribution for up to a scale and location parameter (Nolan [59]); additionally, the Gaussian distribution is a particular case of a stable distribution.

Table 2. The results of the multivariate regression model obtained using Formula (23).

Dependent Variable: $ESH_{\alpha;t}$		
Significance Level	$\alpha = 1\%$	$\alpha = 2.5\%$
Intercept	5.75 *** [0.13]	4.32 *** [0.08]
$H_{0.01,q;t}$	-3.69 *** [0.37]	-2.48 *** [0.23]
$H_{0.025,q;t}$	-3.64 *** [0.38]	-2.57 *** [0.24]
$H_{0.05,q;t}$	-3.07 *** [0.37]	-2.20 *** [0.23]
R^2_{adj}	0.55	0.60

Note: standard errors are given in brackets. *** denotes significance at 1%.

We have that a random variable X which follows an α -stable distribution $S(a,b,c,d;1)$ if its characteristic function is

$$\varphi(t) = \mathbf{E}[e^{itX}] = \begin{cases} \exp(-c^a |t|^a [1 - ib \tan(\frac{\pi a}{2}) \text{sign}(t)] + idt), & a \neq 1 \\ \exp(-c |t| [1 + ibt \frac{2}{\pi} \text{sign}(t)(\ln(|t|)) + idt]), & a = 1 \end{cases} \tag{25}$$

In the above expression $a \in (0,2]$ is the tail index (we obtain $a = 2$ for a normal distribution and lower values correspond to heavier tails), $b \in [-1,1]$ is the skewness parameter, $c \in (0,\infty)$ is the scale parameter, and $d \in \mathbf{R}$ is the location parameter. We simulate 1000 observations from an α -stable distribution $S(a_k,b,c,d;1)$ for $a_k = 1.1 + 0.01k, k = 1 \dots 90$ and study the relationship between tail index and tail entropy.

The tail entropy reaches its maximum value for a distribution that has uniform distribution in the tail, and as the a parameter decreases, the tail entropy of the stable distribution decreases, too. As expected, low tail entropy values are associated with heavier-tailed distributions. Figure 5 presents the relationship between TE and tail index, while Table 3 present the regression results between TE and tail index.

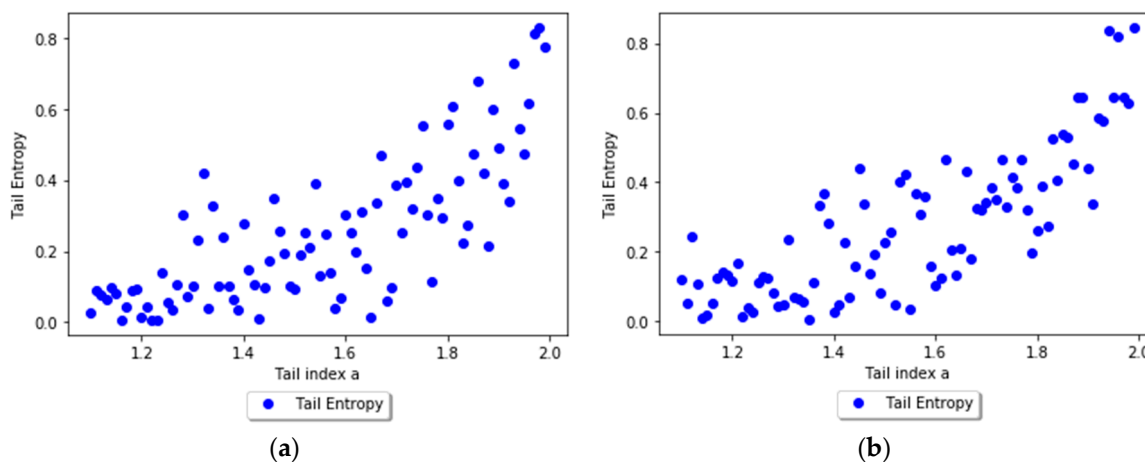


Figure 5. Tail Entropy and tail index of the simulated distributions: (a) $\alpha = 1\%$ and (b) $\alpha = 2.5\%$. See the Github page [SFM ES Stable](#).

Table 3. Relationship between tail index and TE of simulated distributions.

Dependent Variable: $TE_{\alpha,t}$		
Significance Level	$\alpha = 1\%$	$\alpha = 2.5\%$
a	0.51 *** [0.05]	0.63 *** [0.05]
R^2_{adj}	0.50	0.65

Note: standard errors are given in brackets. *** denotes significance at 1%.

Based on the results above (and taking into consideration that the tail entropy is between 0 and 1), we perform a linear adjustment on the tail entropy to obtain the tail entropy expected shortfall, i.e.,

$$ESTE_{\alpha,q}(f) = - \left[b_0 + \left(\frac{b_m - b_0}{2} \right) H_{\alpha,q}(f) \right], \tag{26}$$

where b_0 is the mid-point of the first bin B_0 and b_m is the mid-point of the last bin in the α -tail of the distribution B_m .

The justification for Formula (25) lies in the following points.

- If the entropy reaches its maximum ($H_{\alpha,q}(f) = 1$), which corresponds to the highest uncertainty level, then we have $ESTE_{\alpha,q}(f) = - \left[b_0 + \left(\frac{b_m - b_0}{2} \right) \right] = - \left(\frac{b_m + b_0}{2} \right)$, i.e., the tail entropy expected shortfall is the average of losses below the value-at-risk.
- If the entropy reaches its minimum ($H_{\alpha,q}(f) = 0$), which corresponds to the lowest uncertainty level, then we have $ESTE_{\alpha,q}(f) = -b_0$, i.e., the tail entropy expected shortfall is the minimum value of the distribution.

2.3.2. Tail Entropy Expected Shortfall as a Spectral Risk Measure

According to Guégan et al. [8], a spectral risk measure can be defined as the function $\rho: \mathcal{L} \rightarrow \mathbf{R}$, where \mathcal{L} is a functional space and

$$\rho(X) = - \int_0^1 \phi(p) F^{-1}(x)(p) dp \tag{27}$$

In Formula (26) ϕ is a positive or null, non-increasing, right-continuous, integrable function defined on $[0, 1]$ such as $\int_0^1 \phi(p)dp = 1$.

Following Formula (25), the tail entropy expected shortfall can be written as a linear function of the tail entropy (even though for some distributions this is only an approximation), i.e.,

$$ESTE_{\alpha,q}(f) = -\left[X_{\min} + ES_{\alpha}(X)H_{\alpha,q}(f) \right] = -\left[X_{\min} + \beta \int_0^1 \phi(p)F^{-1}(x)(p)dp \right], \quad (28)$$

where $\beta = -\frac{1}{(4 \ln 2) \log_2 m} = -\frac{1}{4 \ln m}$ and $\phi(p) = -(4 \ln 2)p \log_2 p = -4p \ln p$.

Thus, according to Formula (27), the tail entropy expected shortfall is a spectral risk measure and satisfies the conditions of coherence (see Artzner [1,2]), any spectral risk measure being law-invariant and comonotonic additive (see [8]).

The tail entropy expected shortfall can have interpretability in terms of monetary units, being a linear combination of the expected shortfall and the adimensional normalized entropy which takes values between $[0,1]$, as can be seen from Formula (27).

2.3.3. Backtesting ES

In order to forecast daily ES, non-parametric and parametric models are used based on a rolling window approach. Thus, we compare the ES forecasting ability of the following six models.

1. Historical ES forecasts: $\widehat{ESH}_{\alpha;k+w+1}(X) = -E\left[x_t \mid x_t \leq -\widehat{VaR}_{\alpha;k+w+1}^{Historical} \right]$, where $\widehat{VaR}_{\alpha;k+w+1}^{Historical}$ is the α -quantile of the daily log-returns distribution;
2. Tail entropy ES forecasts: $\widehat{ESTE}_{\alpha,q;k+w+1}(f) = -\left[b_0 + \left(\frac{b_m - b_0}{2} \right) H_{\alpha,q}(\hat{f}) \right]$;
3. Normal GARCH(1,1) ES forecasts: $\widehat{ES}_{\alpha;k+w+1}^{N-GARCH(1,1)}(X) = -E\left[x_t \mid x_t \leq -\widehat{VaR}_{\alpha;k+w+1}^{N-GARCH(1,1)} \right]$,
 where $\widehat{VaR}_{\alpha;k+w+1}^{N-GARCH(1,1)} = -(z_{\alpha} \hat{\sigma}_{k+w+1} + \mu_{k+w+1})$, z_{α} being the Gaussian quantile;
4. Student's t GARCH(1,1) ES forecasts: $\widehat{ES}_{\alpha;k+w+1}^{t-GARCH(1,1)}(X) = -E\left[x_t \mid x_t \leq -\widehat{VaR}_{\alpha;k+w+1}^{t-GARCH(1,1)} \right]$,
 where $\widehat{VaR}_{\alpha;k+w+1}^{t-GARCH(1,1)} = -(t_{\alpha;df} \hat{\sigma}_{k+w+1} + \mu_{k+w+1})$, $t_{\alpha;df}$ being the Student's t quantile with df being the estimated degrees of freedom;
5. Gaussian ES forecasts: $\widehat{ES}_{\alpha;k+w+1}^{Gauss}(X) = -E\left[x_t \mid x_t \leq -\widehat{VaR}_{\alpha;k+w+1}^{Gauss} \right]$, where $-\widehat{VaR}_{\alpha;k+w+1}^{Gauss}$ is the α -quantile of the estimated Gaussian distribution of log-returns;
6. Student's t ES forecasts: $\widehat{ES}_{\alpha;k+w+1}^t(X) = -E\left[x_t \mid x_t \leq -\widehat{VaR}_{\alpha;k+w+1}^t \right]$, where $-\widehat{VaR}_{\alpha;k+w+1}^t$ is the α -quantile of the estimated Student's t distribution of log-returns.

In the above formulae, every risk measure is estimated using a rolling window of length w , with $t \in \{k+1, \dots, k+w\}$, $k \in \{0, \dots, T-w+1\}$, and T the number of daily log-returns. To see the ES forecasting performance of the above models, we run several backtests to compare the ES estimates. These are the unconditional exception frequency (UC) test and the conditional independence (CC) test of Du and Escanciano [60], and the Z_2 exception frequency and magnitude test of Acerbi and Szekely [61]. These are described in Appendixes A and B.

3. Empirical Analysis

In order to illustrate the application of tail entropy in financial risk management, we consider S&P500 daily log-returns (sourced from Bloomberg). The time period considered is January 1 1980 to December 12 2018 (9835 observations). The estimators of expected shortfall described in Section 2 are computed on a rolling basis using a length of $w = 1000$ days for the estimation windows; these forecasts are later subjected to ES backtests. Tail entropy is estimated using a quantum $q = 0.2$ (see the theorem in Section 2). Figure 6 depicts the dynamics of tail-entropy-based ES versus historical ES estimated at 1% and 2.5% significance levels. It can be noted that the tail-entropy-based ES gives higher estimates of risk compared to the historical ES for both significance levels, as this model assigns a higher probability to large losses when compared to the historical ES model. However, it is not straightforward to comment on the correctness of these risk estimates as the level of true risk is not observable, meaning no direct comparisons can be made for true risk. Statistical backtests, performed in the next section, can be used to verify the correctness of VaR and ES forecasts.

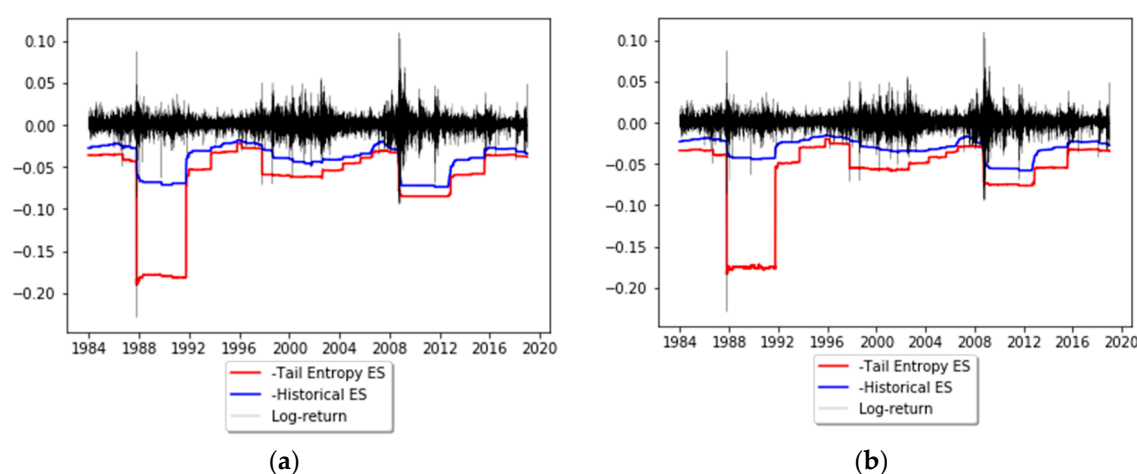


Figure 6. Dynamics of tail-entropy-based ES versus historical ES: (a) $\alpha = 1\%$ and (b) $\alpha = 2.5\%$. See the Github page [SFM TE ES](#).

3.1. Backtesting ES

Backtesting of the ES forecasts based on the models given in Section 2.3 was performed using a window length of 1000 trading days and for $\alpha = 1\%$ and 2.5% . Table 4 presents the percentage of times each model passed the given backtest for each significance level α . Based on these statistics, it can be concluded that for the unconditional ES test, the t -GARCH(1,1) ES forecasts perform best for both levels of α . Looking at the rejection rates for the conditional coverage ES backtest, it can be seen that at the 2.5% level the t -GARCH(1,1) model performs best but at the 1% level the rejection rate of the tail entropy ES is comparable with the same rate within the t -GARCH(1,1) model. Also, according to the Z_2 test, the rejection rates are the lowest for tail entropy ES for $\alpha = 2.5\%$, and stand at around 10%. Based on these results the t -GARCH(1,1) and tail entropy ES show similar performances and overperform the other models considered. Table 5 presents the test statistics of each backtest over the entire testing period and for α levels of 1% and 2.5%. The tail entropy ES, like the other ES models considered, fails both the unconditional and conditional backtests at the 5% level, but it passes the Z_2 test at the 5% level, which the t -GARCH(1,1) ES forecasts fail to pass.

Table 4. Backtesting rejection rates of the null hypothesis from January 1980 to December 2018. Legend: UC, unconditional exception frequency; CC, conditional independence.

Model	$\alpha = 1\%$			$\alpha = 2.5\%$		
	Du's UC Test	Du's CC Test	Z_2 Test	Du's UC Test	Du's CC Test	Z_2 Test
Gaussian ES	62.56%	56.43%	55.55%	74.74%	59.17%	43.93%

Students’s <i>t</i> ES	65.93%	42.80%	38.04%	73.41%	66.03%	34.35%
Historical ES	59.02%	38.14%	39.82%	73.39%	69.79%	33.00%
n-GARCH(1,1) ES	74.83%	33.05%	77.03%	66.91%	46.57%	45.77%
<i>t</i> -GARCH(1,1) ES	29.30%	23.17%	26.73%	27.44%	41.97%	23.32%
Tail Entropy ES	58.35%	24.11%	28.29%	69.50%	69.22%	10.25%

Note: rolling windows of length 1000 were used to carry out the backtests.

To get an insight into the dynamic behavior of the test statistics of the Z_2 backtest, Figure 7 depicts time-varying test statistics (using a rolling window of length 1000) for ES estimated at significance levels $\alpha = 1\%$ and 2.5% . As can be seen, the expected shortfall estimated with the Student’s *t*-GARCH(1,1) model underperformed, for both α levels, during the financial crisis of 2008.

Table 5. Test statistics of ES backtests from January 1980 to December 2018.

Model	$\alpha = 1\%$			$\alpha = 2.5\%$		
	Du’s UC Test	Du’s CC Test	Z_2 Test	Du’s UC Test	Du’s CC Test	Z_2 Test
Gaussian ES	18.26	97.55	−1.88	11.18	84.61	−0.73
Student’s <i>t</i> ES	5.24	45.58	−0.33 *	5.48	74.74	−0.19 *
Historical ES	5.50	44.70	−0.62 *	4.56	81.62	−0.32 *
n-GARCH(1,1) ES	11.43	22.88	−1.39	7.78	18.39	−0.73
<i>t</i> -GARCH(1,1) ES	2.82	16.21	−2.78	3.39	22.37	−3.37
Tail Entropy ES	6.11	20.58	−0.23 *	20.71	66.28	0.20 *

Note: critical values for each test are the following. Du’s UC test (5%, two-sided, normal): −1.96, 1.96. Du’s CC test (5%, one-sided, Chi-squared): 3.84. Z_2 test (5%, one-sided, normal): −0.70. * stands for test statistics that passed the backtests at the 5% level.

The illustration of tail-Entropy-based ES in Figure 6 indicates that at similar backtesting rejection rates tail-entropy-based ES seems to lead to overly high risk quantifications; in addition, the time to recover from a bad shock (like Black Monday on 19 October 1987) is much longer than for historical expected shortfall (which itself usually leads to long recovery times). These two observations seem to indicate that the proposed entropy-based version of expected shortfall tends to react fiercely in the presence of outlying observations. This may be a sign that tail-entropy-based ES is a more conservative risk measure than traditional ones due to the way it is defined, as it takes into account the tail distribution, which has a higher inertia and a longer recovery time from a negative shock.

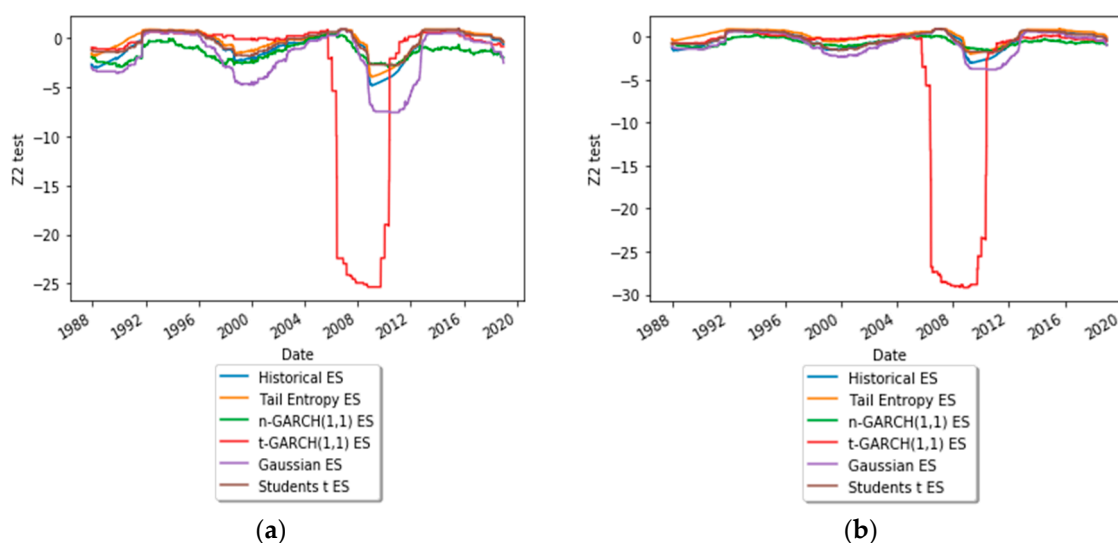


Figure 7. Dynamics of the Z_2 test statistic: (a) $\alpha = 1\%$ and (b) $\alpha = 2.5\%$. See the Github page [SFM TE ES](#).

4. Conclusions

In this paper we propose a nonparametric estimator of ES based on TE, as well as an extension of this model based on kernel smoothing, and compare it with the classical estimates of ES, i.e., historical ES, Gaussian ES, Student's t ES, Gaussian GARCH(1,1) ES and Student's t -GARCH(1,1) ES. The main advantage of the measure we propose is that it has a low dependency on the actual values of observations in the tail of the distribution, making it a more stable measure of tail risk.

We illustrate an application of tail entropy in financial risk management based on S&P 500 daily log-returns between 1980 and 2018 (9835 observations). Comparing backtest rejection rates based on rolling windows, it can be concluded that the performance of tail entropy ES is comparable with the performance of the t -GARCH(1,1) model. However, when we backtest over the entire sample period, tail-entropy-based ES is the only model which passed the Z_2 backtest.

Further research is needed to fully acknowledge the efficiency of the tail entropy ES estimator. This new measure may be beneficial for risk measurement as it is fully non-parametric and has a reduced sensitivity to actual observations in the tail. One possible extension of this approach will be to assess the predictive performance of tail entropy ES in a non-standard environment, for example by assuming a fat-tailed distribution of the returns as the generalized error distribution (see Cerqueti et al., 2019 [62]). In addition, our approach can be applied to other fields such as operational risk, where distributions are usually fat-tailed and highly skewed.

Our research can be easily reproduced using the Python codes uploaded to www.github.com.

Author Contributions: All authors are contributed equally to this paper.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Unconditional and Conditional Backtests for ES (Du and Escanciano)

Let x_t , $t = 1$ to T and denote the returns series, which are assumed to follow a model with estimated cdf \hat{F}_t and estimated parameters $\hat{\theta}_1$. Denoting by Ω_t the information available at time t , we define the cumulative probability under the estimated distribution and the cumulative violations, i.e.,

$$\hat{u}_t = \hat{F}_t(x_t, \Omega_{t-1}, \theta_1), \tag{A1}$$

$$\hat{H}_t(\alpha) = \frac{1}{\alpha}(\alpha - \hat{u}_t)1_{\hat{u}_t \leq \alpha}. \tag{A2}$$

The null hypothesis for the unconditional backtest is

$$H_{0u} : E(H_t(\alpha, \theta_0) - \alpha/2) = 0. \tag{A3}$$

If the estimation period is much larger than the evaluation period, the unconditional test statistic and its distribution will be

$$U_{ES} = \frac{\sqrt{n} \left(\frac{1}{n} \sum_{t=1}^n \hat{H}_t(\alpha) - \frac{\alpha}{2} \right)}{\sqrt{\alpha(1/3 - \alpha/4)}} \xrightarrow{d} N(0, 1). \tag{29}$$

For the conditional backtest the null hypothesis is:

$$H_{0c} : E(H_t(\alpha, \theta_0) - \alpha/2 | \Omega_{t-1}) = 0. \tag{A5}$$

The conditional test statistic (for first order dependence) and its distribution is

$$C_{ES}(1) = \frac{n^3}{(n-1)^2} \cdot \left(\frac{\sum_{t=2}^n (\hat{H}_t(\alpha) - \alpha/2)(H_{t-1}(\alpha) - \alpha/2)}{\sum_{t=1}^n (\hat{H}_t(\alpha) - \alpha/2)^2} \right)^2 \xrightarrow{d} \chi^2(1). \quad (A6)$$

Appendix B. Exception Frequency and Magnitude Test for ES (Acerbi and Szekely)

Let x_t , $t = 1$ to T and denote the returns series with unknown tail distribution cdf F_t . When forecasting the risk of the returns, namely $VaR_{\alpha,t}$ and $ES_{\alpha,t}$, we obtain a predictive tail distribution P_t .

The hypotheses to be tested are

$$H_0 : P_t^{[\alpha]} = F_t^{[\alpha]}$$

$$H_1 : ES_{\alpha,t}^F \geq ES_{\alpha,t} \text{ for all } t \text{ and } > \text{ for some } t$$

$$VaR_{\alpha,t}^F \geq VaR_{\alpha,t}, \text{ for all } t.$$

The test statistic is computed as

$$Z_2 = \sum_{t=1}^T \frac{x_t \mathbf{1}_{x_t < VaR_{\alpha,t}}}{T \alpha ES_{\alpha,t}} + 1. \quad (A7)$$

Under the null hypothesis we have $E_{H_0}[Z_2] = 0$ and $E_{H_1}[Z_2] < 0$. If the test statistic is below -0.7 then the null hypothesis is rejected and the model fails the ES backtest.

References

1. Artzner, P.F.; Delbaen, F.; Eber, J.-M.; Heath, D. Thinking coherently. *Risk* **1997**, *10*, 68–71.
2. Artzner, P.; Delbaen, F.; Eber, J.-M.; Heath, D. Coherent measures of risk. *Math. Financ.* **1999**, *9*, 203–228.
3. Yamai, Y.; Yoshida, T. Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization. *Monet. Econ. Stud.* **2002**, *20*, 87–122.
4. Föllmer, H.; Schied, A. Convex measures of risk and trading constraints. *Financ. Stoch.* **2002**, *6*, 429–447.
5. Basel Committee on Banking Supervision. *Fundamental Review of the Trading Book*; ICMA: Basel, Switzerland, 2012.
6. Basel Committee on Banking Supervision. *Fundamental Review of the Trading Book: A Revised Market Risk Framework*; Bank for International Settlements: Basel, Switzerland, 2013.
7. McNeil, A.; Frey, R.; Embrechts, P. *Quantitative Risk Management*; Princeton University Press: Princeton, NJ, USA, 2005.
8. Guégan, D.; Hassani, B.K. *Risk Measurement. From Quantitative Measures to Management Decisions*; Springer: Cham, Germany, 2019.
9. Nadaraj, S.; Zhang, B.; Chang, S. Estimation methods for Expected Shortfall. *Quant. Financ.* **2014**, *14*, 271–291.
10. McNeil, A.; Frey, R. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *J. Empir. Financ.* **2000**, *7*, 271–300.
11. Embrechts, P.; Klueppelberg, C.; Mikosch, T. *Modeling Extremal Events for Insurance and Finance*; Springer: Berlin/Heidelberg, Germany, 1997.
12. Embrechts, P.; Kaufmann, R.; Patie, P. Strategic long-term financial risks: Single risk factors. *Comput. Optim. Appl.* **2005**, *32*, 61–90.

13. Taylor, J. Estimating value at risk and expected shortfall using expectiles. *J. Financ. Econom.* **2008**, *6*, 231–252.
14. Aigner, D.T. A.; Poirier, D. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. *J. Econom. Rev.* **1976**, *17*, 377–396.
15. Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **1982**, *50*, 987–1007.
16. Schnabel, S.K.; Eilers, P.H. Optimal expectile smoothing. *Comput. Stat. Data Anal.* **2009**, *53*, 4168–4177.
17. Schall, R. Estimation in generalized linear models with random effects. *Biometrika* **1991**, *78*, 719–727.
18. Daouia, A, Girard, S and Stupfler, G. Estimation of Tail Risk based on Extreme Expectiles. *J. R. Stat. Soc. Ser. B* **2018**, *80*, 263–292.
19. Nadaraj, S.; Chan, S.; Afuecheta, E. Tabulations for Value at Risk and Expected Shortfall. *Commun. Stat.—Theory Methods* **2017**, *46*, 5956–5984.
20. Broda, S.A.; Paoletta, M.S. *Expected Shortfall for Distributions in Finance*; Statistical Tools for Finance and Insurance; Cížek, P., Härdle, W., Rafał, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 57–99.
21. Krause, J.; Paoletta, M. A fast, accurate method for value-at-risk and expected shortfall. *Econometrics* **2014**, *2*, 98–122.
22. Chen, S.X. Nonparametric estimation of Expected Shortfall. *J. Financ. Econom.* **2008**, *6*, 87–107.
23. Inui, K.; Kijima, M. On the significance of expected shortfall as a coherent risk measure. *J. Bank. Financ.* **2005**, *29*, 853–864.
24. Jadhav, D.; Ramanathan, T.V.; Naik-Nimbalkar, U.V. Modified estimators of the expected shortfall. *J. Emerg. Mark. Financ.* **2009**, *8*, 87–107.
25. Scaillet, O. Nonparametric estimation and sensitivity analysis of expected shortfall. *Math. Financ.* **2004**, *14*, 115–129.
26. Alemany, R.; Bolancé, C.; Guillén, M. A nonparametric approach to calculating value-at-risk. *Insur. Math. Econ.* **2013**, *52*, 255–262.
27. Cai, Z.; Wang, X. Nonparametric estimation of conditional VaR and expected shortfall. *J. Econom.* **2008**, *147*, 120–130.
28. Hill, J.B. Expected shortfall estimation and Gaussian inference for infinite variance time series *J. Financ. Econom.* **2015**, *13*, 1–44.
29. Richardson, L.F. The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philos. Trans. R. Soc. A* **1911**, *210*, 307–357.
30. Fan, Y.; Li, G.; Zhang, M. Expected shortfall and its application in credit risk measurement. In Proceedings of the 2010 International Conference on Business Intelligence and Financial, Engineering, Hong Kong, China, 13–15 August 2010; pp. 359–363.
31. Martins-Filho, C.; Yao, F.; Torero, M. Nonparametric estimation of conditional Value-at-Risk and Expected Shortfall based on Extreme Value Theory. *Econom. Theory* **2018**, *34*, 23–67.
32. Philippatos, G.C.; Wilson, C. Entropy, Market Risk and the Selection of Efficient Portfolios. *Appl. Econ.* **1972**, *4*, 209–220.
33. Ebrahimi, N.; Maasoumi, E.; Soofi, E.S. Ordering Univariate Distributions by Entropy and Variance. *J. Econom.* **1999**, *90*, 317–336.
34. Dionisio, A.; Menezes, R.; Mendes, D.A. An econophysics approach to analyse uncertainty in financial markets: An application to the Portuguese stock market. *Eur. Phys. J. B* **2006**, *50*, 161–164.
35. Ebrahimi, N.; Maasoumi, E.; Soofi, E.S. Measuring Informativeness of Data by Entropy and Variance. In *Advances in Econometrics: Income Distribution and Methodology of Science, Essays in Honor of Camilo Dagumi*; Slottje, D., Ed.; Springer: Berlin/Heidelberg, Germany, 1999; Chapter 5.
36. Allen, D.E.; McAleer, M.; Powell, R.; Singh, A.K. A non-parametric and entropy based analysis of the relationship between the VIX and S&P 500. *J. Risk Financ. Manag.* **2013**, *6*, 6–30.
37. Pele, D.T.; Lazar, E.; Dufour, A. Information entropy and measures of market risk. *Entropy* **2017**, *19*, 226.
38. Liu, L.Z.; Qian, X.Y.; Lu, H.Y. Cross-sample entropy of foreign exchange time series. *Phys. A Stat. Mech. Appl.* **2010**, *389*, 4785–4792.
39. Zunino, L.; Zanin, M.; Tabak, B.M.; Pérez, D.G.; Rosso, O.A. Forbidden patterns, permutation entropy and stock market inefficiency. *Phys. A Stat. Mech. Appl.* **2009**, *388*, 2854–2864.
40. Risso, A. The Informational Efficiency and the Financial Crashes. *Res. Int. Bus. Financ.* **2008**, *22*, 396–408.

41. Oh, G.; Kim, S.; Eom, C. Market efficiency in foreign exchange markets. *Phys. A Stat. Mech. Appl.* **2007**, *382*, 209–212.
42. Wang, Y.; Feng, Q.; Chai, L. Structural Evolutions of Stock Markets Controlled by Generalized Entropy Principles of Complex Systems. *Int. J. Mod. Phys. B* **2010**, *24*, 5949–5971.
43. Maasoumi, E.; J. Racine. Entropy and predictability of stock market returns. *J. Econom.* **2002**, *107*, 291–312.
44. Billio, M.; Casarin, R.; Costola, M.; Pasqualini, A. An entropy-based early warning indicator for systemic risk. *J. Int. Financ. Mark. Inst. Money* **2016**, *45*, 42–59.
45. Stutzer, M.J. Simple entropic derivation of a generalized Black-Scholes option pricing model. *Entropy* **2000**, *2*, 70–77.
46. Stutzer, M.J.; Kitamura, Y. Connections between entropic and linear projections in asset pricing estimation. *J. Econom.* **2002**, *107*, 159–174.
47. Yang, J.; Qiu, W. A measure of risk and a decision-making model based on expected utility and entropy. *Eur. J. Oper. Res.* **2005**, *164*, 792–799.
48. Gradojevic, N.; Gencay, R. Overnight interest rates and aggregate market expectations. *Econ. Lett.* **2008**, *100*, 27–30.
49. Gencay, R.; Gradojevic, N. Crash of '87—was it expected? Aggregate market fears and long-range dependence. *J. Empir. Financ.* **2010**, *17*, 270–282.
50. Gradojevic, N.; Caric, M. Predicting systemic risk with entropic indicators. *J. Forecast.* **2017**, *36*, 16–25.
51. Bowden, R.J. Directional entropy and tail uncertainty, with applications to financial hazard. *Quant. Financ.* **2011**, *11*, 437–446.
52. Geman, D.; Geman, H.; Taleb, N.N. Tail risk constraints and maximum entropy. *Entropy* **2005**, *17*, 3724–3737.
53. Zhou, R.; Cai, R.; Tong, G. Applications of entropy in finance: A review. *Entropy* **2013**, *15*, 4909–4931.
54. Lorentz, R. On the entropy of a function. *J. Approx. Theory* **2009**, *158*, 145–150.
55. Pele, D.T. Information Entropy and Occurrence of Extreme Negative Returns. *J. Appl. Quant. Methods* **2011**, *6*, 23–32.
56. Miśkiewicz, J. Improving quality of sample entropy estimation for continuous distribution probability functions. *Phys. A Stat. Mech. Appl.* **2016**, *450*, 473–485.
57. Liu, D.; Wang, D.; Wang, Y.; Wu, J.; Singh, V.P.; Zeng, X. Entropy of hydrological systems under small samples: Uncertainty and variability. *J. Hydrol.* **2016**, *532*, 163–176.
58. DeDeo, S.; Hawkins, R.X.D.; Klingenstein, S.; Hitchcock, T. Bootstrap Methods for the Empirical Study of Decision-Making and Information Flows in Social Systems. *Entropy* **2013**, *15*, 2246–2276.
59. Nolan, J.P. *Stable Distributions—Models for Heavy Tailed Data*; Birkhauser: Boston, MA, USA, 2011.
60. Du, Z.; Escanciano, J.C. Backtesting expected shortfall: Accounting for tail risk. *Manag. Sci.* **2017**, *63*, 940–958.
61. Acerbi, C.; Szekely, B. Backtesting expecting shortfall. *Risk* **2014**, *27*, 76–81.
62. Cerqueti, R.; Giacalone, M.; Panarello, D. A Generalized Error Distribution Copula-based method for portfolios risk assessment. *Phys. A Stat. Mech. Appl.* **2019**, *524*, 687–695.

