Check for updates

Received: 11 December 2022 | Revised: 8 March 2023 | Accepted: 12 March 2023

DOI: 10.1002/sta4.560

ORIGINAL ARTICLE

WILEY

Similarity-based clustering for patterns of extreme values

Miguel de Carvalho[1] | Raphael Huser[2] | Rodrigo Rubio[3]

[1]School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, UK

[2]CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

[3]Data Analytics Section, BCI Bank, Santiago 8320000, Chile

Correspondence
Miguel de Carvalho, School of Mathematics, The University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK.
Email: Miguel.deCarvalho@ed.ac.uk

Funding information
The research was partially funded by *Fundação para a Ciência e a Tecnologia* (Portuguese NSF) through the project UID/MAT/00006/2020 and by King Abdullah University of Science and Technology (KAUST).

Statistical modelling of the magnitude and the frequency of extreme observations is fundamental for a variety of sciences. In this paper, we develop statistical methods of similarity-based clustering for heteroscedastic extremes, which allow us to group time series of independent observations according to their extreme-value index and scedasis function (i.e., the magnitude and frequency of extreme values, respectively). Clustering scedasis functions and extreme-value indices involves the challenge of grouping objects composed of both a function (scedasis) and a scalar (extreme-value index), and thus the need to partition a product-space. Our analysis reveals an interesting mismatch between the magnitude and frequency of extreme losses on the London Stock Exchange and the corresponding economic sectors of the affected stocks. The analysis further suggests that the dynamics governing the comovement of extreme losses in the exchange contains information on the business cycle.

KEYWORDS
cluster analysis, cluster functions and scalars, extreme values, risk diversification, statistics of extremes

1 | INTRODUCTION

Small-probability events—such as a stock market crash—often lead to devastating economic and financial aftershocks. The need to assess the likelihood of such rare events is now widely understood, and statistics of extremes (Balkema & Embrechts, 2007; Beirlant et al., 2004; Coles, 2001; Davison & Huser, 2015; de Haan & Ferreira, 2006; Embrechts et al., 1997; Resnick, 2007) provides an appropriate probabilistic framework to address this issue in a mathematically rigorous manner. An overarching principle of statistics of extremes is that any sensible assessment of risk requires the application of resilient methods that are able to extrapolate into the tails of a distribution—often beyond observed extremes in a dataset. For random samples, a key result in statistics of extremes is that if there is a nontrivial limiting distribution for the normalized sample maxima, then it must be a generalized extreme-value (GEV) distribution (McNeil et al., 2015, Theorem 7.3). The shape parameter of the GEV distribution governs the rate of tail decay, and is also known as extreme-value index, or tail index.

The main goal of this paper is to develop cluster analysis methods for highly volatile time series of extremes. Applying methods from statistics of extremes to finance has a long history, including, Danielsson and de Vries (1997), Longin and Solnik (2001), Poon et al. (2003), Herrera and Schipp (2013), Hilal et al. (2014), and Chavez-Demoulin et al. (2014). The monograph recently edited by Longin (2016) contains an up-to-date survey of methods and applications for statistical modelling of extreme values in finance. To model the dynamics of extremes over time, parametric methods to handle nonstationary data have been proposed by Davison and Smith (1990), Chavez-Demoulin and Davison (2005), and Eastoe and Tawn (2009), among others. More recently, Einmahl et al. (2016) have developed a semiparametric modelling and inference framework for heteroscedastic extremes; their two main objects of interest being the scedasis function, which describes the dynamics of extremes over time, and the extreme-value index, which describes the magnitude of extremes.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Stat* published by John Wiley & Sons Ltd.

A recent paper by Ando and Bai (2017), which relates to ours in terms of applied motivation, considers the issue of clustering financial time series based on observable and unobservable factors. In contrast, our approach clusters financial time series based on their resemblance in terms of risk. More precisely, we build on Einmahl et al. (2016) and develop a clustering algorithm that allows us to group stocks that share similarities in their univariate distribution in terms of the overall *magnitude* and the *frequency* of extreme losses. Since the extreme-value index carries information on the rate of tail decay while the scedasis tracks the evolution of extreme losses over time, these parameters form a natural basis for clustering stocks to address our motivating question.

Clustering is an unsupervised learning problem, in the sense that the 'true' cluster labels are unknown and need to be estimated from the data. Introductions to the subject of cluster analysis include Hastie et al. (2009), Everitt et al. (2011), and King (2014). The huge literature on data clustering is difficult to survey in a few paragraphs, but a concise description of mainstream approaches is offered by Hennig and Liao (2013). The most popular clustering approaches may be classified among similarity-based, model-based, and/or hierarchical clustering techniques. Similarity-based methods include K-means (MacQueen, 1967) and K-medoids (Kaufman & Rousseeuw, 1987). Model-based clustering is typically based on mixture models (Fraley & Raftery, 2002), whereas hierarchical clustering builds hierarchies of clusters, often represented in so-called dendograms (Hastie et al., 2009). Hennig and Liao (2013) recently highlighted the lack of clear guidance on choosing an appropriate clustering algorithm for a given problem, despite the many approaches that have been proposed in the literature. The literature discussing extremes is much sparser, although much has been written about the clustering of extreme events within time series induced by short-term temporal dependence (see, e.g., Leadbetter et al., 1983). Clustering algorithms for parallel time series, tailored to specific extreme-value applications, are often intrinsically related to the concept of dimension reduction. For example, Bernard et al. (2013) developed a clustering method for spatial climate extremes, and Chautru (2015) and Vettori et al. (2020) proposed clustering and dimension reduction techniques for multivariate extreme events based on tree mixtures.

Clustering stocks with a similar scedasis function and extreme-value index entails clustering objects defined by the combination of a function (scedasis) and a scalar (extreme-value index); recently, there has been much interest in clustering complex objects such as functions, and the development of methods for clustering functional data is still an area of ongoing research, both from applied and theoretical viewpoints (Delaigle et al., 2012; Peng & Müller, 2008; Wang et al., 2015).

To our knowledge, our paper is pioneer on tackling the challenging problem of performing a cluster analysis of both functions and scalars; to tackle this problem, we develop a clustering approach that can be seen as a K-means method, but where clustering is executed in a product-space defined in terms of the space of all scedasis functions and the positive real line. A weight parameter is then applied as defined by the analyst to control whether the clustering algorithm should prioritize the frequency or the magnitude of extreme losses, which are respectively controlled by the scedasis function and the extreme-value index. The proposed approach also allows for clustering stocks with similar risk loss patterns, by identifying affinities in time-varying value-at-risk functions. To examine the periods of highest market uncertainty and how these affect different stocks, we also explore how the maxima of the stocks' scedasis functions are spread over time. Intuitively, if the maxima of all scedasis functions coincided at a single time point, this would imply a perfect synchronization among all the examined stocks, and thus the largest market fluctuations would reflect similarly on all stocks. To summarize this information, we define the mode mass function, which can be used to assess the degree of synchronization of the maximum frequency of extreme losses between stocks. The latter approach has some connections with multivariate extreme value modelling, which has been widely used by practitioners after the seminal papers of Longin and Solnik (2001) and Poon et al. (2004). Note, however, that in the current paper we focus on univariate, but time-varying, characteristics of extremes, and as such we do not estimate extremal dependence.

In Section 2, we discuss a K-cluster configuration for statistics of heteroscedastic extremes, and we introduce our similarity-based clustering algorithm, as well as the mode mass function. Numerical experiments are reported in Section 3, and the analysis of the London Stock Exchange data is documented in Section 4. Section 5 concludes with a discussion.

2 | CLUSTER ANALYSIS FOR PATTERNS OF EXTREME VALUES

2.1 | Cluster-based proportional tails model

Our starting point for modelling relies on an extended version of Einmahl et al. (2016) formulated for K time series clusters. We assume that the true data generating process consists of N time series partitioned into blocks with homogeneous scedasis functions and extreme-value indices. Specifically, the N time series are partitioned into $K \leqslant N$ blocks $\boldsymbol{Y}^{[1]},\ldots,\boldsymbol{Y}^{[K]}$, where the kth block consists of $N^{[k]}$ time series of length T, that is,

$$\boldsymbol{Y}^{[k]} = \begin{bmatrix} Y_1^{[k,1]} & \ldots & Y_T^{[k,1]} \\ \vdots & \vdots & \vdots \\ Y_1^{[k,N^{[k]}]} & \ldots & Y_T^{[k,N^{[k]}]} \end{bmatrix}. \tag{1}$$

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In Section 2.2, we assume that we observe all time series, but we ignore their underlying cluster labels (i.e., the first element in the superscripts in (1)), which need to be estimated. Here, we assume that the jth time series in the kth cluster, that is, the jth row in group $\mathbf{Y}^{[k]}$, is formed by independent random variables with infinite upper endpoints

$$Y_1^{[k,j]},\ldots,Y_T^{[k,j]}$$

and with respective continuous distributions $F_1^{[k,j]},\ldots,F_T^{[k,j]}$.

Following Einmahl et al. (2016), we assume that there exists a non-negative limit function $c^{[k]}$, called the scedasis function, such that for each time series j in the kth cluster,

$$c^{[k]}\left(\frac{t}{T}\right)=\lim_{y\to\infty}\frac{1-F_t^{[k,j]}(y)}{1-F^{[k,j]}(y)}, \tag{2}$$

for some unknown continuous baseline distribution function $F^{[k,j]}$ with an infinite upper endpoint. Notice that in each cluster (k) the different time series may possess various baseline distributions, but they share the same scedasis function ($c^{[k]}$), as the left-hand side of (2) does not depend on j. For identifiability reasons, we assume that $\int_0^1 c^{[k]}(w)\,\mathrm{d}w=1$, such that $c^{[k]}$ is a valid density on the unit interval $[0,1]$. Generally speaking, the scedasis density carries information on the time dynamics of extremes; for instance, a uniform scedasis function (for which $c^{[k]}(w)=1$, $w\in[0,1]$) corresponds to a constant frequency of extremes over time. We further assume that for each cluster k, time series j, and time point t, one has

$$\lim_{m\to\infty}\frac{U_t^{[k,j]}(my)}{U_0^{[k,j]}(m)\{c^{[k]}(t/T)\}^{\gamma^{[k]}}}=y^{\gamma^{[k]}} \tag{3}$$

for some constant $\gamma^{[k]}>0$, where $U_t^{[k,j]}=\{1/(1-F_t^{[k,j]})\}^{\leftarrow}$ and '$\leftarrow$' is the left continuous inverse function. In other words, the data in each cluster k are heavy-tailed with rate of tail decay controlled by the extreme-value index $\gamma^{[k]}$. Thus, despite the fact that all variables in cluster k have their own distributions, we assume that their tail behaviour is characterized by the pair $(c^{[k]},\gamma^{[k]})$. Therefore, the extremal behaviour of these $N^{[k]}$ time series can be summarized by a single function $c^{[k]}$, describing the time variation in the frequency of extremes and a scalar $\gamma^{[k]}>0$, controlling the overall amplitude of the largest extremes.

For example, (2) and (3) are satisfied when the data have a Pareto-like tail, that is,

$$1-F_t^{[k,j]}(y)\sim\left\{\frac{y}{a^{[k,j]}c^{[k]}(t/T)}\right\}^{-1/\gamma^{[k]}},\ y\to\infty, \tag{4}$$

where $a^{[k,j]}>0$ is an overall scale parameter, in which case the baseline distribution may be chosen as the Pareto distribution $F_0^{[k,j]}(y)=1-(y/a^{[k,j]})^{-1/\gamma^{[k]}}$, with $y>a^{[k,j]}$. From this example, it is clear that the extreme-value index $\gamma^{[k]}$ controls the tail weight, while $c^{[k]}(t/T)$ is a time-varying *relative* scale component, both specific to the kth cluster. The extra time-invariant *overall* scale parameter $a^{[k,j]}>0$ accounts for possible differences in scales between the different time series grouped together in the same cluster.

2.2 | Estimation and inference

As mentioned above, in practice the cluster labels are unknown, so instead of having data partitioned as in (1), we observe N time series written as $Y_1^i,\ldots,Y_T^i$, which consist of independent observations respectively distributed as $F_1^i,\ldots,F_T^i$, for all i, but we ignore the data structure and their tail properties. Instead, the first step of our approach targets estimations for a family of scedasis densities and extreme-value indices denoted by

$$\{(c^i,\gamma^i)\}_{i=1}^N.$$

Under the K-cluster proportional tails model described in Section 2.1, some of these scedasis functions and extreme-value indices are the same, but we estimate them separately.

Let $Y_{(1)}^i\leqslant\ldots\leqslant Y_{(T)}^i$ be the order statistics from the ith time series, $Y_1^i,\ldots,Y_T^i$, and $k\equiv k_T$ be an intermediate sequence, that is, $k_T\to\infty$ and $k_T/T\to 0$, as $T\to\infty$. To estimate each element in the family of scedasis functions, we use the nonparametric kernel-based estimator of Einmahl et al. (2016) for all i, that is

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\hat{c}^i(w) = \frac{1}{k}\sum_{t=1}^{T} I(Y_t^i > Y_{(T-k)}^i) K_b(w - t/T),\ w \in [0,1], \tag{5}$$

where $K_b(\cdot) = K(\cdot/b)/b$ is a kernel with bandwidth $b \equiv b_T > 0$ such that $b_T \to 0$ and $b_T k_T \to \infty$ as $T \to \infty$. It is well known that the choice of kernel does not affect the inference much, whereas bandwidth selection is crucial. Following Marron (2001, p. 533), we advocate conducting inference over a wide range of bandwidths as a way to assess the sensitivity and reliability of the inference to the bandwidth.

To estimate the family of extreme-value indices, we follow Einmahl et al. (2016) and use the Hill estimator defined as

$$\hat{\gamma}^i = \frac{1}{k}\sum_{t=1}^{k} \log\left(\frac{Y_{(T-t+1)}^i}{Y_{(T-k)}^i}\right), \tag{6}$$

for all i. As shown in Einmahl et al. (2016), the estimator (6) is strongly consistent and asymptotically normal under the heteroscedastic proportional tail model described in Section 2.1 and mild additional regularity conditions. Furthermore, it does not depend on the associated scedasis function c^i as $T \to \infty$.

Asymptotic normality, and thus weak consistency, for the kernel-based estimator (5) is also established in Einmahl et al. (2016, p. 49), but under rather stringent conditions. These appealing large sample properties suggest that our clustering algorithm developed in Section 2.3 should perform well in practice, which is confirmed by our numerical experiments in Section 3, especially for a large sample size T. The outcome of this estimation stage yields N pairs of scedasis and extreme-value estimates, $\{(\hat{c}^i, \hat{\gamma}^i)\}_{i=1}^N$, and we next describe how such pairs can be partitioned into K clusters.

2.3 | Similarity-based clustering for heteroscedastic extremes

Clustering scedasis functions and extreme-value indices requires overcoming the challenge of clustering objects composed of both a function (scedasis) and a scalar (extreme-value index). To approach this problem, we consider a similarity-based clustering algorithm. Similarity-based clustering methods usually require the true number of clusters $K \leqslant N$ to be either known or specified a priori. For these methods, a data point $\mathbf{x}^i = (x_1^i, \ldots, x_d^i)^\mathsf{T}$ should be assigned to a certain cluster if its distance to the cluster centre

$$\mathbf{m}^{[k]} = (m_1^{[k]}, \ldots, m_d^{[k]})^\mathsf{T}$$

is minimal. Different variants lead to different concepts of a cluster centre, the most well-known being K-means (MacQueen, 1967) and K-medoids (Kaufman & Rousseeuw, 1987). Whereas in the case of the K-means algorithm a cluster centre $\mathbf{m}^{[k]}$ is defined by a statistic (e.g., a cluster mean), in the K-medoids algorithm a cluster centre is defined by a data point itself. To measure the dissimilarity of a data point $\mathbf{x}^i$ from a cluster centre $\mathbf{m}^{[k]}$, it is common to resort to a dissimilarity function $D : \mathbb{R}^{2d} \to [0, \infty)$, such that for all possible data points and cluster centres the following properties are satisfied:

1. Non-negativity: $D(\mathbf{x}^i, \mathbf{m}^{[k]}) \geqslant 0$;
2. Identifiability: $D(\mathbf{x}^i, \mathbf{m}^{[k]}) = 0$ if and only if $\mathbf{x}^i = \mathbf{m}^{[k]}$;
3. Symmetry: $D(\mathbf{x}^i, \mathbf{m}^{[k]}) = D(\mathbf{m}^{[k]}, \mathbf{x}^i)$.

The most widely used dissimilarity function is the squared Euclidean distance

$$D(\mathbf{x}^i, \mathbf{m}^{[k]}) = \sum_{j=1}^{d} (x_j^i - m_j^{[k]})^2.$$

As mentioned above, here we are faced with the need of extending the standard similarity-based clustering framework so to be able to cluster both a function (scedasis) and a scalar (extreme value index). We mostly focus on the case where each cluster $\mathbf{Y}^{[k]}$ in (1) is characterized by both the same scedasis function *and* extreme-value index, which we refer to as *partitioning in the product-space* $\Pi = \mathcal{C} \times (0, \infty)$, where $\mathcal{C}$ denotes the space of all scedasis functions (i.e., all densities defined on the unit interval). However, it might also be interesting in practice to cluster time series according to their scedasis function—resulting in K^c clusters—or extreme-value index—resulting in K^γ clusters. We refer to these cases as *partitioning in the profile-spaces*$\mathcal{C}$ and $(0, \infty)$. Clearly, K does not necessarily coincide with K^c and K^γ. This is the case when there are no two time series that have the same scedasis function but a different extreme-value index, or vice versa. In general, one has $\max(K^c, K^\gamma) \leqslant K \leqslant K^c K^\gamma$.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Clustering is performed here with the aid of an *encoder*, that is, a map $\mathcal{L}:\{1,\ldots,N\}\to\{1,\ldots,K\}$ assigning each time series $i=1,\ldots,N$ to a cluster label $k\in\{1,\ldots,K\}$. Standard encoders assign single observations to a cluster (Hastie et al., 2009), while here whole sets of observations are assigned simultaneously to a cluster.

In the time series context of heteroscedastic extremes, suppose that we are given a family of estimated scedasis functions and extreme-value indices, $\{(\hat{c}^i,\hat{\gamma}^i)\}_{i=1}^N$, obtained using the procedure discussed in Section 2.2. We measure dissimilarity by relying on a family of functions

$$D_\alpha:\Pi\times\Pi\to[0,\infty),\ \alpha\in(0,1),\ \Pi=\mathcal{C}\times(0,\infty),$$

such that

$$\lim_{\alpha\to 1}D_\alpha(\pi^i,\pi^j)=D^c(c^i,c^j),\ \lim_{\alpha\to 0}D_\alpha(\pi^i,\pi^j)=D^\gamma(\gamma^i,\gamma^j), \tag{7}$$

where $\pi^i=(c^i,\gamma^i)\in\Pi$ and $\pi^j=(c^j,\gamma^j)\in\Pi$; with

$$D^c:\mathcal{C}\times\mathcal{C}\to[0,\infty),\ D^\gamma:(0,\infty)\times(0,\infty)\to[0,\infty),$$

being dissimilarity measures specific to the scedasis function and to the extreme-value index, respectively. Equation (7) requires dissimilarity in the product-space to be compatible with the dissimilarities in the corresponding profile-spaces, when $\alpha\to 0$ and $\alpha\to 1$. Roughly speaking, $\alpha\in(0,1)$ can also be interpreted as a tuning parameter, controlling whether the dissimilarity function should apply more weight to the dissimilarity in terms of the scedasis density or the extreme-value index. Examples 1 and 2 below are two possible valid families of dissimilarity functions obeying (7), although many others may be designed.

Example 1. (Convex linear dissimilarity in product-space). Let $D_\alpha^{\text{conv}}(\pi^i,\pi^j)=\alpha D^c(c^i,c^j)+(1-\alpha)D^\gamma(\gamma^i,\gamma^j)$,

where $\pi^i=(c^i,\gamma^i)\in\Pi$ and $\pi^j=(c^j,\gamma^j)\in\Pi$, with $D^c:\mathcal{C}\to\mathcal{C}$ and $D^\gamma:(0,\infty)\to[0,\infty)$ being dissimilarity measures.

Example 2. (Max-linear dissimilarity in product-space). Let $D_\alpha^{\max}(\pi^i,\pi^j)=\max\{\alpha D^c(c^i,c^j),(1-\alpha)D^\gamma(\gamma^i,\gamma^j)\}$,

where $\pi^i=(c^i,\gamma^i)\in\Pi$ and $\pi^j=(c^j,\gamma^j)\in\Pi$, with $D^c:\mathcal{C}\to\mathcal{C}$ and $D^\gamma:(0,\infty)\to[0,\infty)$ being dissimilarity measures.

Algorithm 1 *K*-means algorithm for heteroscedastic extremes

- **Step 0 (Initialization)**
  Randomly select K cluster centres
  $$\boldsymbol{\mu}^{(\text{old})}=((\bar{c}^{[1]},\bar{\gamma}^{[1]})^{(\text{old})},\ldots,(\bar{c}^{[K]},\bar{\gamma}^{[K]})^{(\text{old})}).$$
- **Step 1 (Cluster (re)assignment)**
  Allocate time series to clusters using the rule
  $$\mathcal{L}^{(\text{new})}(i)=\arg\min_k D_\alpha((\hat{c}^i,\hat{\gamma}^i),(\bar{c}^{[k]},\bar{\gamma}^{[k]})^{(\text{old})}).$$
- **Step 2 (Cluster-centre updating)**
  Search for cluster centres which minimize total dissimilarity, given the latest cluster assignment made; we thus look for
  $$\boldsymbol{\mu}^{(\text{new})}=\arg\min_{\mu\in\Pi}\sum_{i=1}^N\sum_{k=1}^K I(\mathcal{L}^{(\text{new})}(i)=k)D_\alpha((\hat{c}^i,\hat{\gamma}^i),(\bar{c}^{[k]},\bar{\gamma}^{[k]})).$$

Update cluster assignment and centres iteratively until a stopping criteria is fulfilled.

It can be shown that if D^c and D^γ are distance functions, then so are D_α^{conv} and $D_\alpha^{\max}$ in Examples 1 and 2, respectively, for every value of $\alpha\in(0,1)$.

Algorithm 1 describes the *K*-means type of algorithm underlying our clustering procedure, detailing the three main steps: Initialization, cluster (re)assignment and cluster-centre updating, of which the last two need to be recursively iterated until convergence. Below, we focus on $D_\alpha^{\text{conv}}(\pi^i,\pi^j)$, with $D^c(c^i,c^j)=\int_0^1\{c^i(w)-c^j(w)\}^2\mathrm{d}w$ and $D^\gamma(\gamma^i,\gamma^j)=(\gamma^i-\gamma^j)^2$, although there are many other variants, for example, replacing the ℓ_2 dissimilarity with the ℓ_1 dissimilarity, or focusing on specific features of the scedasis function (e.g., its maximum) rather than its overall behaviour. Unreported results show that the choice of dissimilarity measure (e.g., ℓ_1 versus ℓ_2) does not affect the results much. With our choice of

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

dissimilarity measure, the algorithm outlined in Algorithm 1 becomes a genuine K-means algorithm in the sense that cluster centres are indeed found using the actual means; more precisely, Step 2 can be replaced with $\boldsymbol{\mu}^{(\text{new})} = \{(\overline{c}^{[k]}, \overline{\gamma}^{[k]})\}_{k=1}^{K}$, where

$$\overline{c}^{[k]}(w) = \frac{1}{\hat{N}^{[k]}} \sum_{\mathcal{L}(i)=k} \hat{c}^i(w), \quad \overline{\gamma}^{[k]} = \frac{1}{\hat{N}^{[k]}} \sum_{\mathcal{L}(i)=k} \hat{\gamma}^i, \tag{8}$$

with $\hat{N}^{[k]} = \sum_{i=1}^{N} I(\mathcal{L}(i) = k)$, for all k; see Appendix A for a proof. Note that $\overline{c}^{[k]}(w)$ and $\overline{\gamma}^{[k]}$ are compatible with the assumptions made in Section 2.1, namely, $\int_0^1 \overline{c}^{[k]}(w)\mathrm{d}w = 1$ and $\overline{\gamma}^{[k]} > 0$. This implies that the cluster centres are formed by valid scedasis functions and extreme-value indices, corresponding to a distribution in the maximum domain of attraction of the Fréchet distribution.

A K-medoids type of algorithm can also be applied, in which case the cluster centres correspond to the scedasis and extreme-value index of the observed time series. In practice, this implies that Steps 0 and 1 are the same as for the K-means algorithm but that the cluster-centre updating rule in Step 2 needs to be modified as follows: For each cluster k, search for the time series in this cluster that minimizes the total dissimilarity to other time series in the same cluster

$$i_k^* = \arg \min_{\{i:\mathcal{L}(i)=k\}} \sum_{\mathcal{L}(i')=k} D_\alpha((\hat{c}^i, \hat{\gamma}^i), (\hat{c}^{i'}, \hat{\gamma}^{i'})),$$

and set $\overline{c}^{[k]} \equiv \hat{c}^{i_k^*}$, $\overline{\gamma}^{[k]} \equiv \hat{\gamma}^{i_k^*}$, for all k, and $\boldsymbol{\mu}^{(\text{new})} = ((\hat{c}^{i_1^*}, \hat{\gamma}^{i_1^*}), \ldots, (\hat{c}^{i_K^*}, \hat{\gamma}^{i_K^*}))$. The K-medoids algorithm requires however a larger computational investment than the K-means approach.

The cluster centres of the K-medoids algorithm, $\overline{c}^{[k]} = \hat{c}^{i_k^*}$ and $\overline{\gamma}^{[k]} = \hat{\gamma}^{i_k^*}$, are obviously compatible with the assumptions made in Section 2.1; that is, $\overline{c}^{[k]}$ is a valid scedasis function and $\overline{\gamma}^{[k]} > 0$, for all k. Other variants of the K-means and K-medoids algorithms for clustering heteroscedastic extremes can also be constructed similarly, for example, using Partition Around Medoid (PAM) and Clustering Large Applications (CLARA) (Kaufman & Rousseeuw, 2009, Ch. 2–3).

2.4 | Clustering risk loss patterns

This section capitalizes on the setup introduced so far to offer another natural way to cluster financial time series. Time-varying value-at-risk (VaR), at confidence level $p \in (0,1)$, can be defined as $\mathrm{VaR}_t(p) = \inf\{y : F_t(y) \geqslant p\}$, where p is typically taken as a high probability such as $p = 0.95$ or $p = 0.99$.

More specifically, using the scedasis function and extreme-value index estimators (5) and (6), respectively, we can obtain an estimator for the (functional) value-at-risk for each time series $i = 1, \ldots, N$, by

$$\widehat{\mathrm{VaR}}_w^i(p) = \frac{\hat{a}^i}{(1-p)^{\hat{\gamma}^i}} \hat{c}^i(w), \; w \in [0,1], \; p \to 1, \tag{9}$$

where the estimated scale parameter $\hat{a}^i = Y_{(T-k)} k^{\hat{\gamma}^i} / \left\{ \sum_{t=1}^{T} \hat{c}^i(t/T)^{1/\hat{\gamma}^i} \right\}^{\hat{\gamma}^i}$ is obtained in closed form by setting $k/T = T^{-1} \sum_{t=1}^{T} \{1 - F_t^i(Y_{(T-k)})\}$, combined with the Pareto-like tail specification (4).

Algorithm 2 K-geometric means algorithm for value-at-risk

- **Step 0 (Initialization)**
  Randomly select K cluster centres
  $$\boldsymbol{\mu}^{(\text{old})} = (\overline{\mathrm{VaR}}^{[1]}, \ldots, \overline{\mathrm{VaR}}^{[K]}).$$
- **Step 1 (Cluster (re)assignment)**
  Allocate time series to clusters using the rule
  $$\mathcal{L}^{(\text{new})}(i) = \arg \min_k D(\widehat{\mathrm{VaR}}_w^i(p), \overline{\mathrm{VaR}}^{[k]}).$$
- **Step 2 (Cluster-centre updating)**
  Search for cluster centres which minimize total dissimilarity, given the latest cluster assignment made; we thus look for
  $$\boldsymbol{\mu}^{(\text{new})} = \arg \min_{\mu \in \Pi} \sum_{i=1}^{N} \sum_{k=1}^{K} I(\mathcal{L}^{(\text{new})}(i) = k) D(\widehat{\mathrm{VaR}}_w^i(p), \overline{\mathrm{VaR}}^{[k]}).$$

Update cluster assignment and centres iteratively until a stopping criteria is fulfilled.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Below, we focus on $D(\widehat{\mathrm{VaR}}_w^i(p),\widehat{\mathrm{VaR}}_w^j(p)) = \int_0^1 \{\log\widehat{\mathrm{VaR}}_w^i(p) - \log\widehat{\mathrm{VaR}}_w^j(p)\}^2 \, dw$. With this choice of dissimilarity measure, Algorithm 2 becomes a K-geometric means algorithm in the sense that cluster centres can be shown to be geometric means; more precisely, Step 2 can be replaced with $\boldsymbol{\mu}^{(\mathrm{new})} = \{\overline{\mathrm{VaR}}_w^{[k]}(p)\}_{k=1}^K$, where

$$\overline{\mathrm{VaR}}_w^{[k]}(p) = \left(\prod_{\{i:\mathcal{L}(i)=k\}} \widehat{\mathrm{VaR}}_w^i(p) \right)^{1/\hat{N}^{[k]}}, \tag{10}$$

with $\hat{N}^{[k]} = \sum_{i=1}^N I(\mathcal{L}(i)=k)$; the proof is tantamount to that in Appendix A.

2.5 | Mode mass function

To summarize the periods of highest market uncertainty and to see how these periods reflect on each specific stock, we suggest studying how the modes of the various scedasis functions for different stocks are spread over time through the *mode mass function*. Assume that each estimated scedasis function $\hat{c}^i$ has a unique global maximum and define its unique mode $m^i \in [0,1]$ as

$$m^i = \arg\max_{w \in [0,1]} \hat{c}^i(w), \; i = 1,\ldots,N. \tag{11}$$

Note that (11) can also be written as $m^i = \arg\max_{w \in [0,1]} \widehat{\mathrm{VaR}}_w^i(p)$, under the Pareto-like tail specification in (4); recall Equation (9). A smoothed version of the mode mass function may be computed using a beta kernel density as

$$\hat{M}(w) = \frac{1}{N}\sum_{i=1}^N \beta\{w;\nu m^i,\nu(1-m^i)\}, \; w \in [0,1], \tag{12}$$

where $\beta(w;a,b)$ denotes the Beta density function with parameters $a,b>0$. In the parametrization (12), the beta kernels have mean m^i and $\nu>0$ is a concentration parameter. From an analytical perspective, the smoothed mode mass function in (12) is related to the smooth Euclidean likelihood spectral density estimator proposed by de Carvalho et al. (2013, p. 16). Conceptually, it operates similarly to standard kernel smoothing, in the sense that it 'centres' a beta kernel at each scedasis mode, but here the kernels are asymmetric for $m^i \neq 0.5$ and they are tailored for observations defined in the unit interval.

If the modes of all scedasis functions coincide at a single point in time, that is, $\hat{M}(w)$ concentrate around a certain $w_0 \in [0,1]$, then the periods at which the frequencies of extreme losses are the largest for the different stocks are perfectly synchronized; on the other hand, if the modes of the scedasis functions are uniformly spread over time, that is, $\hat{M}(w) \approx 1$, $w \in [0,1]$, then this suggests that the different stocks' maximum frequencies of extremes are not related at all.

3 | NUMERICAL EXPERIMENTS

3.1 | Simulation setting and preliminary experiments

We start by describing the data-generating processes used to assess the performance of our clustering algorithm. We simulate independent observations according to the structure (1) using a generalized extreme-value distribution defined as

$$F_t^{[k,j]}(y) = \exp\left[-\left\{1+\gamma^{[k]}\left(\frac{y-c^{[k]}(t/T)}{c^{[k]}(t/T)}\right)\right\}_+^{-1/\gamma^{[k]}}\right],$$

for time $t=1,\ldots,T$, clusters $k=1,\ldots,K$, and time series $j=1,\ldots,N^{[k]}$, where $a_+ = \max(0,a)$. We consider three simulation scenarios.

In Scenario A, the true scedasis densities are defined as $c^{[k]}(w) = \tau(w;p_k)$, $p_k \in [0,1]$, where,

$$\tau(w;p_k) = \begin{cases} 4(1-p_k)w + p_k, & 0 \leqslant w < 0.5; \\ -4(1-p_k)(w-1) + p_k, & 0.5 \leqslant w < 1, \end{cases}$$

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

yielding triangular-shaped scedasis functions with peaks at the centre point $w = 0.5$, and the true extreme-value indices are set to $\gamma^{[k]} = 0.7$ or 1. In Scenario B, the true scedasis functions $c^{[k]}(w) = \beta(w;\alpha_k,\beta_k)$ are Beta(α_k,β_k) densities and the extreme-value indices are set to $\gamma^{[k]} = 0.7$ or 1. In Scenario C, the scedasis densities are either $\tau(w;p_k)$ or $\beta(w;\alpha_k,\beta_k)$, and the extreme-value index is set to $\gamma^{[k]} = 0.7$.

The simulation settings and our choice of parameter values are reported in Table 1 and illustrated in Figure 1; for Scenarios A and B, we consider $K = 4$ true clusters in the product-space, but $K^c = K^\gamma = 2$ clusters in the respective profile-spaces, comprising a total of $N = 30$ time series. We simulate five independent time series with parameters (c_1,γ_1), five with (c_2,γ_1), 10 with (c_1,γ_2), and 10 with (c_2,γ_2); the scedasis densities c_1,c_2 with their parameters and the extreme-value indices γ_1,γ_2 are specified in Table 1. Scenario C differs from this setting, as it is characterized by three different scedasis functions and a single extreme-value index, which yields $K = K^c = 3$ and $K^\gamma = 1$. For each cluster, we simulate $N^{[k]} = 10$ time series independently for a total of $N = 30$ time series.

For Scenarios A, B, and C, a single-run experiment was performed using our methods to obtain the estimated cluster centres and scedasis functions. These are compared with the true scedasis functions in Figure 1. To estimate each scedasis function, we consider the sample sizes $T = 500,1000,2000,5000$ with the number of upper order statistics $k = 34,62,112,250$, chosen according to the intermediate sequence $k_T \propto \lfloor T/\log T \rfloor$, which corresponds to the increasing threshold probabilities 93.2%, 93.8%, 94.4%, and 95.0%, respectively. We use the kernel-based estimator of Einmahl et al. (2016) with $K_b(w) = 15\{1-(w/b)^2\}^2/(16b)$, for $w \in [-1,1]$. We select the bandwidth values based on $b = Ak^{-1/5}$, with the constant $A > 0$ set by visual inspection in order to improve estimation in a sensitivity analysis (not shown) considering a grid of possible values for A; using this approach, the values of the bandwidth b that we consider here are 0.2, 0.15, 0.1, and 0.1 for each sample size, respectively. The extreme-value index is estimated using the Hill estimator.

Here and below, we use the convex linear combination dissimilarity from Example 1, that is, $D_\alpha^{\text{conv}}(\pi^i,\pi^j) = \alpha D^c(c^i,c^j) + (1-\alpha)D^\gamma(\gamma^i,\gamma^j)$, with $D^c(c^i,c^j) = \int_0^1 \{c^i(w) - c^j(w)\}^2 \mathrm{d}w$, and $D^\gamma(\gamma^i,\gamma^j) = (\gamma^i - \gamma^j)^2$; recall that in such a case the cluster centres can be obtained as in (8). In Figure 1, we set $\alpha = 0.5$. While Figure 1 suggests that our clustering method performs rather well, firm conclusions can only be drawn from a rigorous Monte Carlo study. This is investigated in the next section.

3.2 | Monte Carlo simulations

Here, we conduct an extensive simulation study to assess the performance of our clustering algorithm in grouping different time series into separate clusters defined in terms of magnitude and dynamics of the extremes. We calculate the Rand (Rand, 1971) and silhouette indices (Rousseeuw, 1987), whose mathematical definition are given in the Supporting Information, based on 1000 runs of our algorithm under the settings described in Section 3.1. Notice that the silhouette index can be computed even if the true data partition is unknown. By contrast, the Rand index is used to measure the similarity between two different partitions, one of which is considered the true partition in our simulation study.

Table 2 reports the Monte Carlo means of the Rand and silhouette indices for the simulation settings described in Section 3.1, as a function of the sample size T and the weight parameter α. We can observe that both validity measures are almost always maximized when $\alpha = 0.5$ for all scenarios (i.e., 'weighting' equally the scedasis density and the extreme-value index). However, as $T \to \infty$, the clustering performance in Scenario C should improve as $\alpha \to 1$ (i.e., more weight on the scedasis density than on the extreme-value index), because $K^\gamma = 1$ and therefore all information for identifying distinct clusters is contained in the scedasis profile space. At subasymptotic levels (in this case, $T = 5000$ and $k = 250$), values of α close to 0.5 seem to work better. Furthermore, as expected, performance improves as T increases; so with larger sample sizes, the estimation of the scedasis functions and extreme-value indices in the first step of our procedure is more accurate, which then in the second step reflects on the estimated clusters.

When the number of clusters K is unknown or uncertain, as is typically the case in practice, it is useful to perform clustering over a range of values for K, and then compare the resulting diagnostics. We now explore the use of the elbow method, proposed by Tibshirani et al. (2001) and adapt it to clustering heteroscedastic extremes. The method is presented here specifically for the K-means algorithm and convex combination D_α of ℓ_2 dissimilarity measures (recall Example 1), but it can also be designed more generically. The elbow method is based on the statistic $D^{[k]}$, defined as

TABLE 1 Data-generating scenarios. For Scenarios A, B, and C, each row reports the number of clusters in the product-space (K), the number of clusters in the respective profile-spaces (K^c and K^γ), the different scedasis functions and extreme-value indices involved, and the cluster sizes.

Scenario	Number of clusters (K,K^c,K^γ)	Scedasis density c_i	Extreme-value index γ_j	Cluster size N_{ij}
A	(4,2,2)	$\tau(w;0.001);\tau(w;0.8)$	0.7;1	$N_{1,1} = N_{2,1} = 5, N_{1,2} = N_{2,2} = 10$
B	(4,2,2)	$\beta(w;2,5);\beta(w;5,2)$	0.7;1	$N_{1,1} = N_{2,1} = 5, N_{1,2} = N_{2,2} = 10$
C	(3,3,1)	$\beta(w;2,5);\beta(w;5,2);\tau(w;0.5)$	0.7	$N_{1,1} = 10, N_{2,1} = 10, N_{3,1} = 10$

Note: N_{ij} is the number of time series simulated independently in a specific cluster characterized by the scedasis c_i and extreme-value index γ_j.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

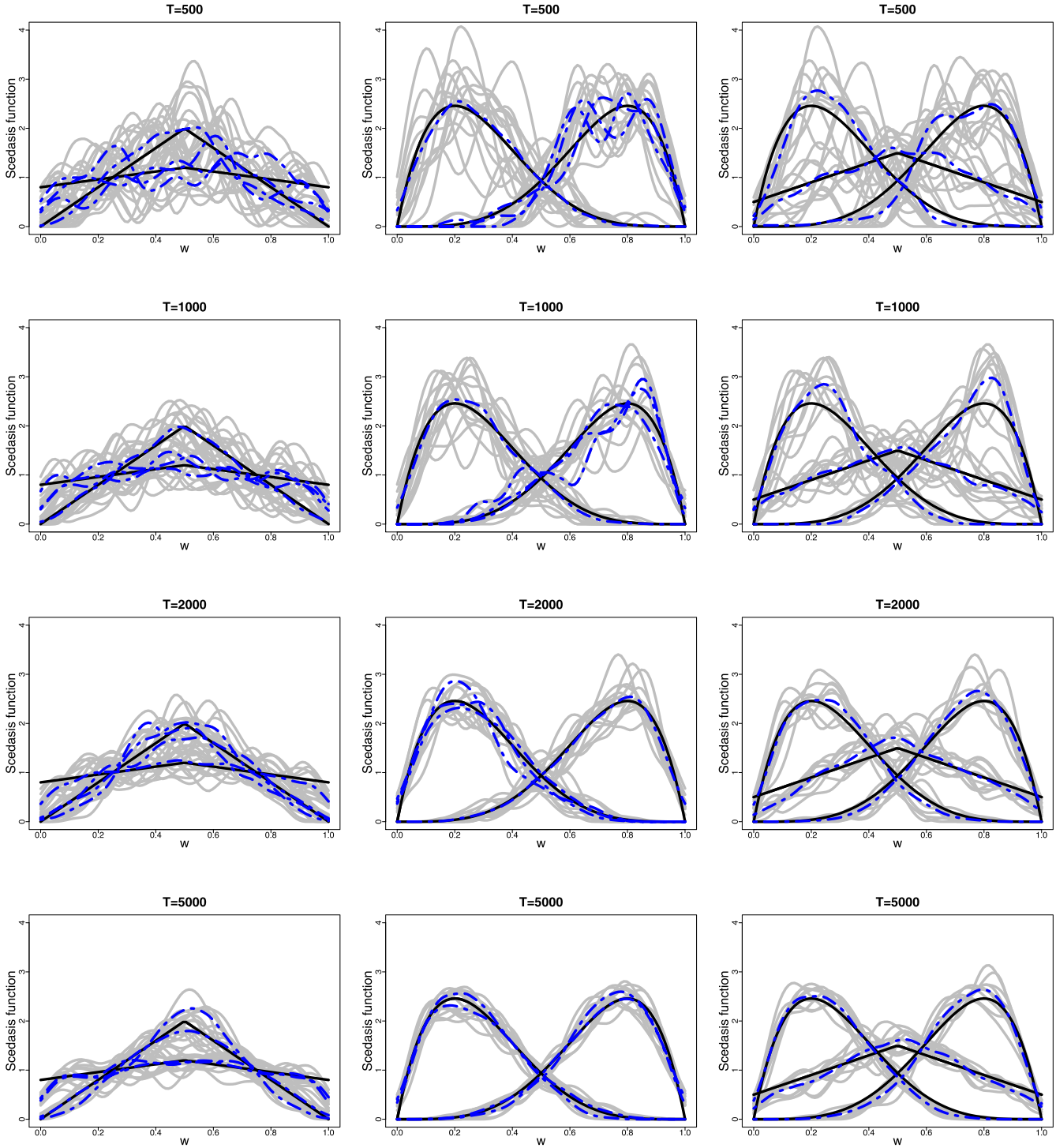

FIGURE 1 The dashed blue lines depict the cluster centre estimates produced by our method, plotted against the true scedasis functions (solid black) defined in Table 1; the scedasis function estimates are depicted in grey. Scenarios A, B, and C (left to right) are presented for sample sizes $T = 500, 1000, 2000, 5000$ (top to bottom).

$$D^{[k]} = \sum_{i,j \in I^{[k]}} D_\alpha(\pi^i, \pi^j) = 2N^{[k]} \sum_{i \in I^{[k]}} D_\alpha(\pi^i, \mu^{[k]}),\ \mu^{[k]} = (\bar{c}^{[k]}, \bar{\gamma}^{[k]}),$$

where $I^{[k]}$ is the index set corresponding to the kth cluster $C^{[k]}$, and $\mu^{[k]} = (\bar{c}^{[k]}, \bar{\gamma}^{[k]})$ is the cluster centre obtained using the K-means algorithm; $D^{[k]}$ represents the sum of intracluster dissimilarities between 'points' in a given cluster $C^{[k]}$, and the corresponding cluster centres. The sum of normalized intracluster sums of squares measures the overall 'compactness' of our clustering procedure:

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Validity measures for Scenarios A, B, and C as defined in Section 3.1 obtained from 1000 Monte Carlo simulations. The Rand and silhouette indices evaluate the clustering of the K-means algorithm for heteroscedastic extremes as a function of the sample size (T) and the weight parameter (α).

	Rand index				Silhouette index			
$\alpha \setminus T$	500	1000	2000	5000	500	1000	2000	5000
Scenario A								
0.1	0.74	0.85	0.86	0.87	0.47	0.52	0.64	0.76
0.3	0.77	0.79	0.86	0.88	0.49	0.64	0.81	0.81
0.5	0.79	0.83	0.89	0.93	0.50	0.72	0.87	0.95
0.7	0.78	0.81	0.85	0.87	0.42	0.44	0.53	0.92
0.9	0.68	0.73	0.78	0.81	0.31	0.42	0.51	0.74
Scenario B								
0.1	0.76	0.86	0.88	0.90	0.33	0.50	0.76	0.79
0.3	0.80	0.83	0.87	0.94	0.48	0.61	0.79	0.86
0.5	0.79	0.89	0.90	0.92	0.41	0.65	0.77	0.95
0.7	0.74	0.81	0.85	0.88	0.40	0.53	0.75	0.85
0.9	0.70	0.74	0.77	0.82	0.36	0.43	0.62	0.80
Scenario C								
0.1	0.77	0.81	0.86	0.90	0.54	0.66	0.74	0.82
0.3	0.80	0.83	0.89	0.92	0.65	0.71	0.82	0.88
0.5	0.82	0.84	0.90	0.94	0.70	0.64	0.90	0.94
0.7	0.79	0.82	0.85	0.91	0.68	0.66	0.85	0.89
0.9	0.74	0.79	0.86	0.88	0.65	0.69	0.74	0.80

Note: Values closer to one correspond to better performance.

$$W_K = \sum_{k=1}^{K} \frac{1}{2N^{[k]}} D^{[k]}. \tag{13}$$

The elbow method runs the K-means algorithm for heteroscedastic extremes on the dataset for a range of values of K and then plots the sum of squared errors W_K as a function of K. The plot should, in principle, be monotonically decreasing with K until the true value of K is reached, and then stabilize for larger values of K. The optimal number of clusters is therefore found at the 'elbow' of the plotted curve. To illustrate this approach, we consider the three aforementioned simulation Scenarios A, B, and C with a sample size $T = 5000$ and weight parameter $\alpha = 0.5$. The results are displayed in Figure 2. As expected, the 'elbow' is found at $K = 4$ for Scenarios A and B, but at $K = 3$ for Scenario C, which all correspond to the true number of clusters.

4 | ANALYSIS OF STOCK MARKET DATA FROM THE LONDON STOCK EXCHANGE

4.1 | London Stock Exchange: Data and main modelling goals

We gathered the data from Yahoo Finance. After some preliminary analysis detailed in Section 4.2, we retained 139 stocks from the London Stock Exchange, and Figure 3 presents the daily negative log-returns for a selection of 26 stocks, including GlaxoSmithKline, Royal Dutch Shell B, Tesco, among others. The stocks displayed in Figure 3 cover nine economic sectors of the exchange, namely, oil and gas, basic materials, industrials, healthcare, consumer goods, consumer services, financials, utilities, and technology; about 40% of these selected stocks are FTSE 100 companies. The period under analysis ranges from December 1989 to May 2016, and thus it includes stock market crashes clearly visible from the time series plots, such as black Wednesday (1992), the dot-com bubble (2000), and the most recent financial crisis (2008–2009); it also takes into account the turbulence in the eminence of the Brexit referendum, which took place on 23 June 2016. The preparations for post-Brexit have actually been creating a wealth of challenges and opportunities for institutional investors, professional money managers, and traders; for example, there are plans for the exchange to take on board further companies from the Middle East in a near future (Reuters, 2017). One of our applied goals is to develop and apply statistical methods that can be used to partition stocks such as those presented in Figure 3 into meaningful clusters from the

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

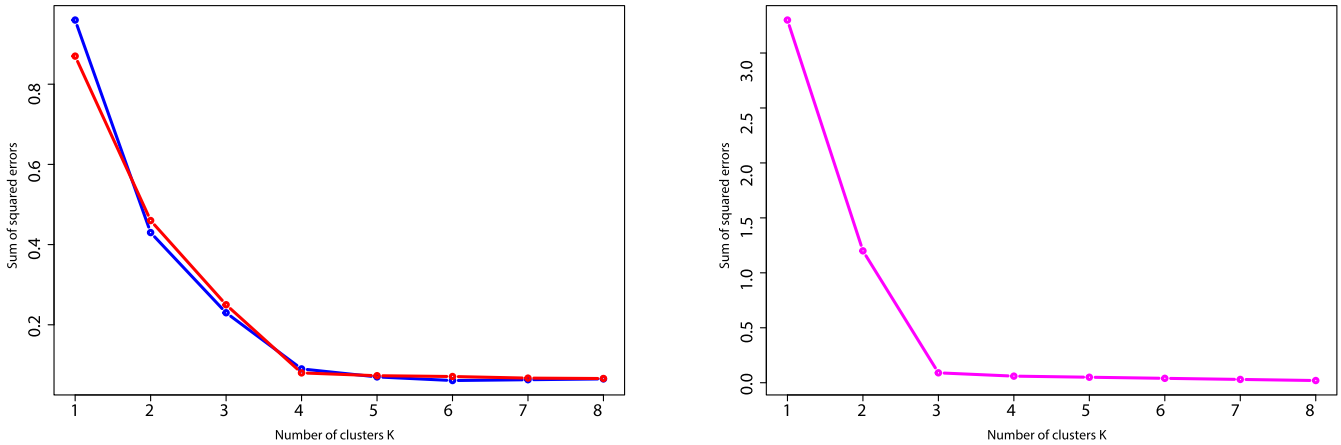

FIGURE 2 Sum of squared errors W_K in (13) for Scenarios A, B (left), and C (right) described in Section 3.1 with $T = 5000$ and $\alpha = 0.5$, as a function of the number of clusters and averaged over 1000 Monte Carlo simulations. The blue and red curves correspond to Scenarios A and B respectively, while the magenta curve corresponds to Scenario C.

viewpoint of risk; the latter might, or might not, reflect the nine economic sectors. More precisely, an overall objective of our analysis is to group stocks according to the *magnitude* and *frequency* of univariate extreme losses. As negative log-returns can be regarded as a proxy for losses, we focus on the right tail of the data shown in Figure 3. Our quantitative analysis is here intended to complement the overall picture provided by exploratory, qualitative or expert-based diagnostics. For example, a quick visual inspection of Figure 3 could perhaps suggest that Stocks 1 (Royal Dutch Shell B) and 5 (Johnson Matthey) look reasonably similar in terms of magnitude of extreme losses, but whether the dynamics of their extreme losses over time is similar is far from clear. The methods presented in this paper will offer answers to these and related questions.

4.2 | Preliminary analysis and selection of stocks

We now apply our model to the London Stock Exchange data briefly presented in Section 4.1. To use our model, we need to test whether the extreme-value index γ is constant over time for each stock. For this, we consider Test 4 of Einmahl et al. (2016, p. 37), based on the comparison of γ estimates from four subperiods. For each stock, we assume that the negative log-returns on each day follow possibly different heavy-tailed distributions as in (1), and we test whether the extreme-value index of the negative log-returns is constant between 1989 and 2016. More concretely, first, we identify the stocks for which we had data available in the period between 1989 and 2016, obtaining 240 companies; second, we consider all stocks for which we do not reject the null hypothesis of constant extreme-value index using the same criteria as in Einmahl et al. (2016, Test 4, p. 37). The main findings are similar for the analysis for the 139 stocks retained according to these two step procedure and for the selection of 26 stocks displayed in Figure 3; therefore, we focus below on the latter for presentation clarity; in the Supporting Information, we also present the extended analysis based on the full dataset with 139 stocks. The 26 selected time series, plotted in Figure 3, are of length $T = 6893$ days. In the Supporting Information, we show that extremes of these stocks are weakly dependent, and that thus that the methodology proposed in Section 2 can be applied. Table 3 presents a list of the 26 selected firms, their economic sectors, their estimated scedasis functions (recall (5)) and the Hill estimates (recall (6)) corresponding to the negative log-returns of each of these stocks, setting $k = 332$ (corresponding approximately to taking the threshold as the 95%-percentile) and $b = 0.1$. At first glance, we can see that the estimated scedasis functions and extreme-value indices do not appear to be homogeneous within each economic sector, but rather there is a remarkable match between each stock's estimated scedasis function and its corresponding cluster centre as determined by our new clustering procedure; more details about our results are provided in Section 4.3.

4.3 | Do clusters mirror economic sectors?

In this section, we examine whether the clusters obtained through our approach have any connection with the stocks' economic sectors. Ex-ante, an investor could be tempted to believe that stocks belonging to the same economic sector might be characterized by similar extreme-value properties, but whether this actually holds true is another matter.

To conduct the proposed inquiry, we run the proposed K-means algorithm for heteroscedastic extremes from Section 2.3 to group the selected companies into new clusters. To run our clustering procedure, we first need to choose the number of clusters K; Figure 4 displays the

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

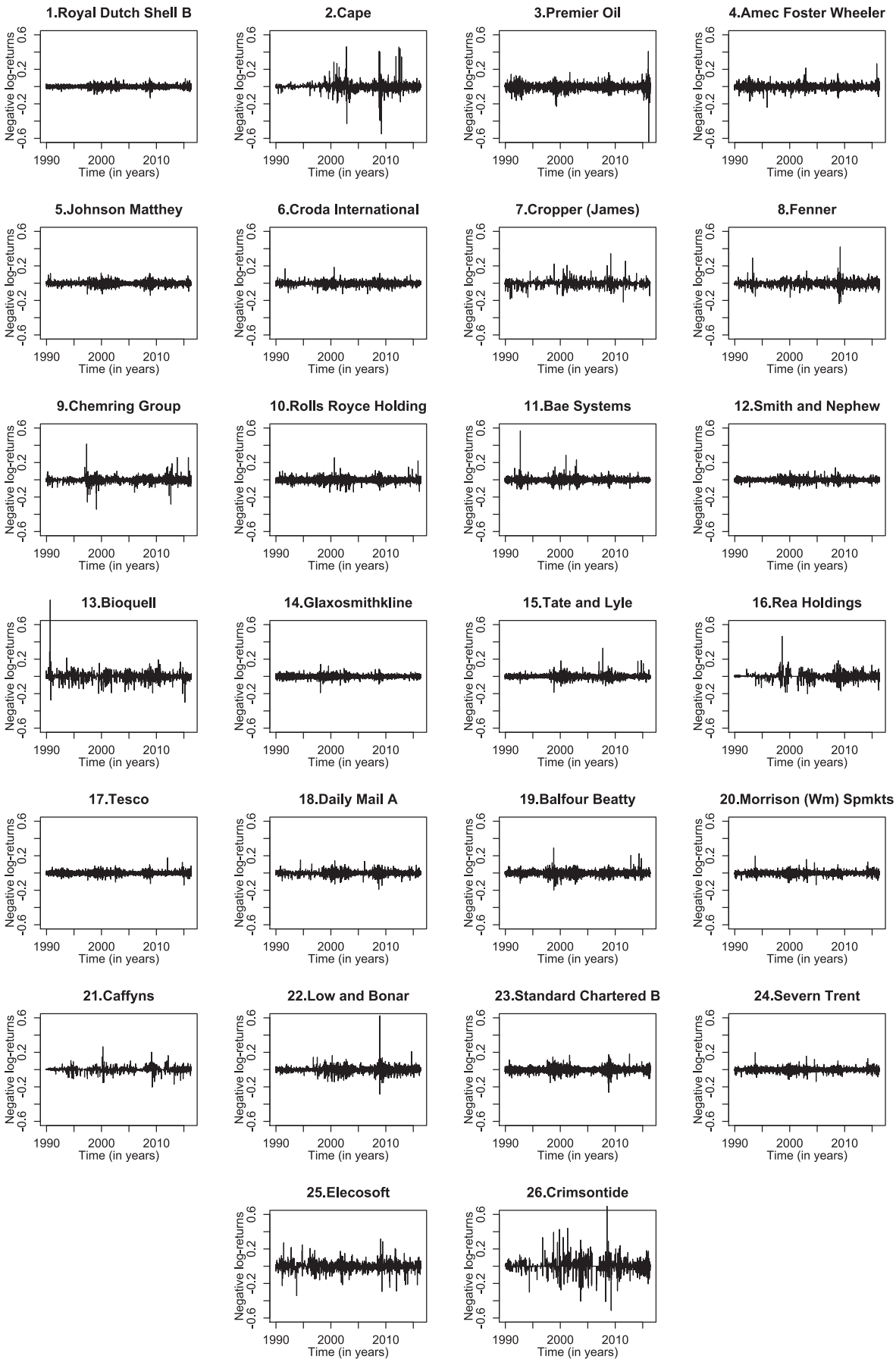

FIGURE 3 Negative log-returns for 26 selected stocks from the London Stock Exchange.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Estimated scale parameters ($\hat{a}$), extreme-value indices ($\hat{\gamma}$), and scedasis functions ($\hat{c}$) (solid black) plotted over time, for the stocks under analysis and corresponding economic sectors. The dashed blue curves correspond to the scedasis' cluster centres estimated by our algorithm. The daggers (†) represent FTSE 100 companies.

Stocks and economic sectors	$(\hat{a}, \hat{\gamma})$	$\hat{c}$	Stocks and economic sectors	$(\hat{a}, \hat{\gamma})$	$\hat{c}$
Oil and gas			*Basic materials*		
1. Royal Dutch Shell B†	(0.006,0.363)		5. Johnson Matthey†	(0.007,0.377)	
2. Cape	(0,005,0.551)		6. Croda International†	(0.007,0.363)	
3. Premier Oil	(0.011,0.385)		7. Cropper (James)	(0.003,0.603)	
4. Amec Foster Wheeler	(0.009,0.382)				
Industrial			*Health care*		
8. Fenner	(0.006,0.467)		12. Smith and Nephew†	(0.006,0.369)	
9. Chemring Group	(0.006,0.432)		13. Bioquell	(0.006,0.499)	
10. Rolls-Royce Holdings†	(0.009,0.343)		14. GlaxoSmithKline†	(0.007,0.310)	
11. Bae Systems†	(0.007,0.377)				
Consumer goods			*Consumer services*		
15. Tate and Lyle	(0.005,0.410)		17. Tesco†	(0.008,0.327)	
16. Rea Holdings	(0.005,0.499)		18. Daily Mail A	(0.006,0.404)	
Technology			19. Balfour Beatty	(0.010,0.332)	
25. Elecosoft	(0.007,0.503)		20. Morrison (Wm) Spmkts†	(0.007,0.353)	
26. Crimson tide	(0.002,0.843)		21. Caffyns	(0.002,0.695)	
			22. Low and Bonar	(0.005,0.449)	
Financials			*Utilities*		
23. Standard Chartered B†	(0.009,0.392)		24. Severn Trent†	(0.006,0.364)	

sum of squared errors W_K defined in (13) as a function of $K = 1,\ldots,20$ for various weight parameters $\alpha = 0,0.1,\ldots,0.9,1$. Following the elbow method described in Section 3.2, the number of clusters K should be only about two or three when α is small (placing heavier weight on the extreme-value index), perhaps due to the high variability in the estimation of the extreme-value index γ. For larger values of α (placing heavier weight on the scedasis function), the elbow method suggests that $K > 3$. However, the optimal value of K is not obvious as the curve does not completely stabilize. Here, we fix $K = 9$ for the rest of the analysis to contrast our results with the nine economic sectors. To reduce the possibility that the K-means algorithm for heteroscedastic extremes generates a suboptimal clustering, we set the initial cluster centres using the approach by Arthur and Vassilvitskii (2007). In addition, for each value of α, we run the algorithm 500 times and select the solution $\mathcal{L}^*$ with the smallest value of the objective function

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

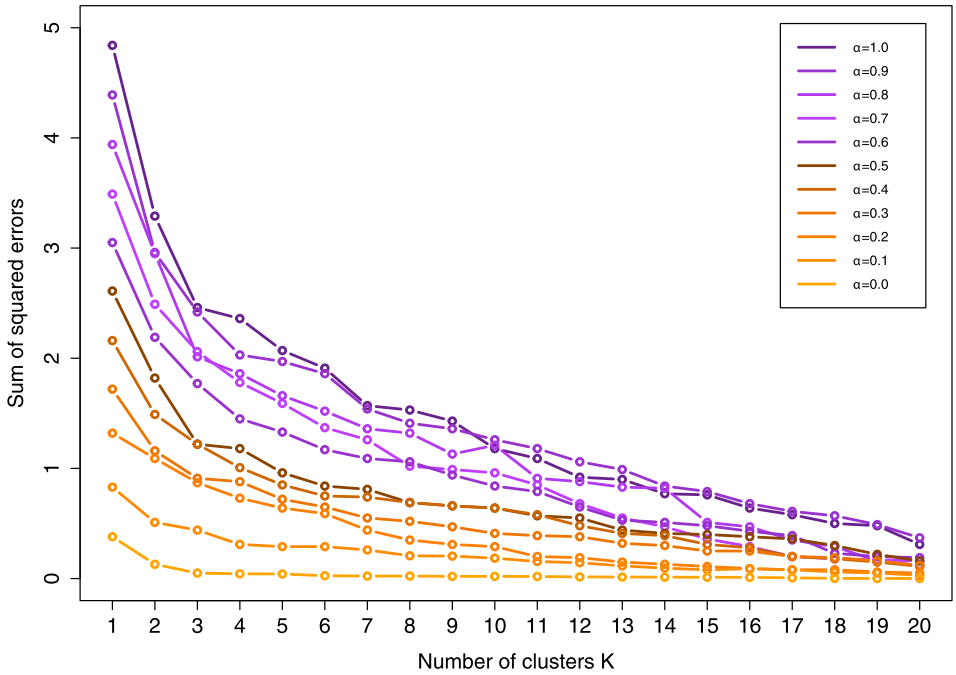

FIGURE 4 Sum of squared errors W_K in (13) for the data described in Section 4.2, plotted as a function of the number of clusters for various values of the weight parameter $\alpha = 0,0.1,\ldots,0.9,1$ (bottom to top).

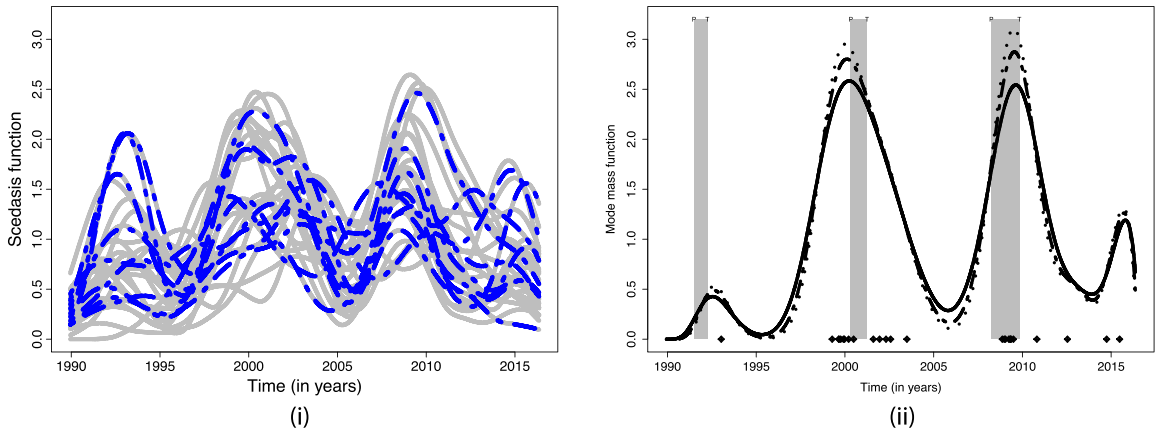

FIGURE 5 (i) The dashed blue lines correspond to cluster centres estimates obtained by our method, plotted against the scedasis function estimates (grey lines) for the stocks under analysis. (ii) The mode mass function using $\nu = 75$ (solid), $\nu = 100$ (dashed), and $\nu = 120$ (dotted); the grey rectangles correspond to contraction periods in the US economy as dated by the NBER, and the rug corresponds to the modes of the estimated scedasis functions.

$$W(\mathcal{L}^*) = \sum_{i=1}^{N} \sum_{k=1}^{K} I(\mathcal{L}^*(i) = k) D_\alpha((\hat{c}^i, \hat{\gamma}^i), (\bar{c}^{[k]}, \bar{\gamma}^{[k]})).$$

Figure 5(i) and Table 4 show the scedasis function estimates of the selected companies and the cluster centre estimates produced by our method, when $\alpha = 0.5$. The choice of the latter value is motivated by the numerical experiments reported in Section 3, which suggest that the validity measures are almost always maximized when $\alpha = 0.5$; a sensitivity analysis with respect to the choice of α is reported in the Supporting Information. We can see a significant difference between the grouping of companies in Table 3 and the new classifications given in Table 5, suggesting that there is no direct connection between the stocks' economic sectors and their classification in terms of magnitude and frequency of extreme losses.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Cluster centre estimates corresponding to the nine extreme-value indices, for the partition of stocks obtained with $\alpha = 0.5$. Ticker symbols are the same as in Table 3.

	Cluster centre extreme-value index ($\bar{\gamma}_k$)								
	0.385	0.466	0.575	0.346	0.843	0.385	0.432	0.382	0.377
Stocks	1, 5, 12, 15, 18	2, 6, 8, 16, 22	7, 13, 21, 25	10, 14, 17, 19, 20, 23, 24	26	3	9	4	11

TABLE 5 Nine clusters of heteroscedastic extremes of companies over a grid of values of α. Ticker symbols are the same as in Table 3.

	Tuning parameter (α)		
Groups	0	0.5	1
1	3, 4, 5, 11, 15, 18, 23	1, 5, 12, 15, 18	1, 5, 10, 11, 15, 18, 20, 23, 24
2	8, 9, 22	2, 6, 8, 16, 22	2, 6, 9, 22
3	13, 16, 25	7, 13, 21, 25	13
4	1, 6, 12, 20, 24	10, 14, 17, 19, 20, 23, 24	14, 17, 19, 21, 25
5	26	26	26
6	2	3	3
7	21	9	8, 16
8	10, 14, 17, 19	4	4
9	7	11	7, 12

Our algorithm defines new 'sectors', which are more homogeneous in terms of the magnitude and frequency of extreme losses, and highlights an interesting mismatch between stocks' economics sectors and clusters of risk in terms of frequency and magnitude of extreme losses. In more detail, Table 5 shows the partition of stocks on the product and profile spaces obtained using our clustering method. Here, $\alpha = 0.5$ corresponds to the scenario where we put equal interest on the frequency and magnitude of the extremes losses. If an investor is more interested in the magnitude of extreme losses than their frequency, then (s)he should consider a classification from Table 5 with $\alpha = 0$; on the other hand, if (s)he is more interested in the frequency of extreme losses than in their magnitude, then (s)he should consider a classification from Table 5 with $\alpha = 1$. All in all, either in the product-space or in the profile-space, our analysis suggests that clusters of magnitude and dynamics of extreme losses present no straightforward connection with the economic sectors of the corresponding stocks. The same finding holds if clustering is based on risk loss patterns as introduced in Section 2.4. Table 6 contains the partition of stocks obtained using time-varying value-at-risk. Although the two clustering algorithms in Tables 5 and 6 are based on different criteria, they still produce remarkably similar results. In addition, as can be seen from Table 6 the partition of stocks obtained using time-varying value-at-risk is quite robust to changes in the confidence levels. All in all, both results in Tables 5 and 6 agree that the estimated clustering of stocks presents no straightforward connection with the economic sectors of the corresponding stocks, which implies that stocks from the same economic sector may have very different risk loss patterns.

4.4 | Economic contraction periods and the mode mass function

In this section, we examine the mode mass function as defined in (12) and compare its shape with the contraction periods for the UK and US economies. Stylized facts on the UK business cycle can be found in Chadha and Nolan (2002). Business cycle dating for the US economy is conducted by the National Bureau of Economic Research (NBER). While it might seem odd to compare financial information from one country (UK) with economic figures from another country (US), it is well known that the US economy is a major player in the world economy, and some studies suggest that the UK business cycle is actually more synchronized with the US than with countries in the Euro area zone, such as Germany (Albulescu, 2017). Figure 5(ii) shows how the modes of the scedasis functions are spread over time. The mode mass function in this figure suggests four periods of great synchronization in terms of the maximum frequency of extremes losses. The grey rectangles represent contraction periods of the US economy, from peak (P) to trough (T). We can see that the local maxima of the mode mass function are relatively close to these economic contraction periods. The scedasis function maxima tend to be reasonably well aligned with the contraction periods dated by the NBER, which appears compatible with the hypothesis that around periods of contraction, the maximum frequencies of extremes losses are more synchronized. Our findings are also in line with the business cycle peak to through dates for the UK from the Economic Cycle Research Institute (ECRI). Interestingly, however, for the period under analysis, ECRI only identifies the peak and through dates May 1990–March 1992 and May 2008–January 2010, and thus their chronology does not consider that the 2000 dot-com bubble has left its footprint on the cycle.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Nine clusters of value-at-risk functions of companies over a grid of values of p. Ticker symbols are the same as in Table 3.

	Confidence level (p)				
Groups	0.95	0.98	0.99	0.995	0.999
1	1, 5, 12, 15, 18, 22	1, 5, 12, 15, 18, 22	1, 5, 12, 15, 18	1, 12, 15	1, 5, 12, 14, 15
2	6, 8, 9	6, 8, 9	6, 8, 9, 22	8, 9, 16	8, 9, 16
3	13, 19	13, 19	13	13	13
4	4, 10, 11, 14, 17	4, 10, 11, 14, 17	4, 10, 11, 14, 17, 19	4, 10, 11, 14, 17, 19	4, 10, 11, 17, 19
	20, 23, 24	20, 23, 24	20, 23, 24	20, 23, 24	20, 23, 24
5	26	26	26	26	26
6	3	3	3	3	3
7	2, 16	2, 16	2, 16	2, 5, 6, 18, 22	2, 6, 18, 22
8	25	25	25	25	25
9	7, 21	7, 21	7, 21	7, 21	7, 21

5 | DISCUSSION

This paper develops and applies statistical similarity-based clustering techniques for heteroscedastic extremes, which can be used to group time series of independent observations that are alike in terms of magnitude and dynamics of extreme observations. Our method yields a data-driven partition of stocks into clusters of risk, and can be used to uncover which stocks resemble each other the most in terms of probabilistic features that are critical from a risk management perspective.

Clustering scedasis functions and extreme-value indices involves the challenge of grouping objects composed of both a function (scedasis) and a scalar (extreme-value index), and thus the need to partition a product-space. To address this issue, we have proposed valid dissimilarity measures able to transition from one profile-space to the other, characterized by a tuning parameter to be set by the analyst, deciding where the clustering algorithm should place more weight. To our knowledge, this is the first paper tackling the challenging problem of conducting a cluster analysis of both functions and scalars. Similarly to Einmahl et al. (2016), our approach cannot treat stocks for which the extreme-value index varies in time.

Our clustering algorithm is designed to group stocks whose *marginal* distributions of extremes over time have similar characteristics. In particular, considering the scedasis function and extreme-value index as objects of interest, our clustering method puts strong focus on the dynamics of univariate extremes over time (through the scedasis function) and the overall magnitude of the most extreme losses (through the extremal index); in addition to the scedasis function ('relative' scale parameter) and extreme-value index (tail shape parameter), it is possible to use a Pareto-like tail specification to estimate the 'overall' scale parameter and deduce the time-varying value-at-risk, which provides a more complete description of extreme losses. As described in our paper, this can be used to cluster univariate risk loss patterns. In future work, it would also be interesting to devise a clustering algorithm that takes extremal *dependence* into account as a way of tracing and outlining portfolios exhibiting joint (i.e., simultaneous) extreme losses. In particular, the characterization of extremal dependence among stocks is crucial for assessing the risk of an index, defined for example as a weighted sum of the prices of the individual stocks. We leave such an important question for future research.

Some final remarks on our financial illustration are in order. In practice, return series may exhibit serial dependence and volatility clustering. While an option to come closer to the independence assumption is to work with monthly returns, that also leads to a substantial reduction in the amount of data, hence weakening the case for learning about limiting objects—such as the scedasis and the extreme value index. In addition, it should be noted that the allocations devised via our method are not static as the scedasis evolves over time. Thus, recent events such as the COVID-19 pandemic, the Black Monday 2020 crashes, rising inflation and the war in Ukraine, can naturally impact the allocations reported in our illustration—and so can any other future event that influences stock markets. Finally, we acknowledge survivorship bias in our empirical findings, as the analysis focuses on companies that have succeeded in the market—while ignoring those that have failed.

DATA AVAILABILITY STATEMENT

The data are publicly available from the R package `extremis`.

The data that support the findings of this study are available in Yahoo Finance at https://finance.yahoo.com. These data were derived from the following resources available in the public domain: - R package extremis, https://cran.r-project.org/web/packages/extremis/index.html.

ORCID

Raphael Huser https://orcid.org/0000-0002-1228-2071

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

Albulescu, C. T. (2017). UK Business cycle synchronization with Germany and the US: New evidence from time–frequency domain and structure of economic growth. *Applied Economics Letters*, *24*, 67–71.

Ando, T., & Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, *112*, 1182–1198.

Arthur, D., & Vassilvitskii, S. (2007). *K*-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035.

Balkema, A. A., & Embrechts, P. (2007). *High risk scenarios and extremes: A geometric approach*. Amsterdam: European Mathematical Society.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes*. Chichester: Wiley.

Bernard, E., Naveau, P., Vrac, M., & Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, *26*, 7929–7937.

Chadha, J. S., & Nolan, C. (2002). A long view of the UK business cycle. *National Institute Economic Review*, *182*, 72–89.

Chautru, E. (2015). Dimension reduction in multivariate extreme value analysis. *Electronic Journal of Statistics*, *10*, 109–133.

Chavez-Demoulin, V., & Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*, 207–222.

Chavez-Demoulin, V., Embrechts, P., & Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, *181*, 44–52.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer.

Danielsson, J., & de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of empirical Finance*, *4*, 241–257.

Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and its Application*, *2*, 203–235.

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*, 393–442.

de Carvalho, M., Oumow, B., Segers, J., & Warchol, M. (2013). A Euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics-Theory and Methods*, *42*, 1176–1192.

de Haan, L., & Ferreira, A. (2006). *Extreme value theory: An introduction*. New York: Springer.

Delaigle, A., Hall, P., & Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika*, *99*, 299–313.

Eastoe, E. F., & Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *58*, 25–45.

Einmahl, J. H. J., de Haan, L., & Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *78*, 31–51.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. New York: Springer.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. London: Wiley.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*, 611–631.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining inference and prediction* (2nd ed.). New York: Springer.

Hennig, C., & Liao, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*, 309–369.

Herrera, R., & Schipp, B. (2013). Value at risk forecasts by extreme value models in a conditional duration framework. *J. Empir. Fin.*, *23*, 33–47.

Hilal, S., Poon, S.-H., & Tawn, J. (2014). Portfolio risk assessment using multivariate extreme value methods. *Extremes*, *17*, 531–556.

Kaufman, L., & Rousseeuw, P. J. (1987). *Statistical data analysis based on the L1-norm and related methods*. Amsterdam: North Holland.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

King, R. S. (2014). *Cluster analysis and data mining: An introduction*. Virginia: Mercury Learning and Information.

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extreme and related properties of random sequences and processes*. New York: Springer.

Longin, F. (Ed.) (2016). *Extreme events in finance: A handbook of extreme value theory and its applications Edited by* Longin, F. Wiley: Hoboken.

Longin, F., & Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, *56*, 649–676.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5Th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). Oakland, CA, USA.

Marron, J. S. (2001). Discussion of inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, *96*, 532–533.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management*. Princeton: Princeton University Press.

Peng, J., & Müller, H.-G. (2008). Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, *2*, 1056–1077.

Poon, S.-H., Rockinger, M., & Tawn, J. A. (2003). Modelling extreme-value dependence in international stock markets. *Statistica Sinica*, *13*, 929–953.

Poon, S.-H., Rockinger, M., & Tawn, J. A. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies*, *17*, 581–610.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850.

Resnick, S. I. (2007). *Heavy-tail phenomena: Probabilistic and statistical modeling*. New York: Springer.

Reuters (2017). London Stock Exchange looks to Middle East for revival after Brexit. Retrieved from www.reuters.com

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*, 411–423.

Vettori, S., Huser, R., Segers, J., & Genton, M. G. (2020). Bayesian model averaging over tree-based dependence structures for multivariate extremes. *Journal of Computational and Graphical Statistics*, *29*, 174–190.

Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2015). Review of functional data analysis. arXiv preprint 1507.05135.

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: de Carvalho, M., Huser, R., & Rubio, R. (2023). Similarity-based clustering for patterns of extreme values. *Stat*, *12*(1), e560. https://doi.org/10.1002/sta4.560

APPENDIX A: DERIVATION OF CLUSTER CENTRES IN PRODUCT-SPACE

This appendix provides details on the cluster-centre updating step in the K-means algorithm for heteroscedastic extremes. Recall that in Step 0 of the K-means algorithm in Section 2.3, we select K cluster centres randomly, $\boldsymbol{\mu}^{(\text{old})} = ((\bar{c}^{[1]}, \bar{\gamma}^{[1]})^{(\text{old})}, \ldots, (\bar{c}^{[K]}, \bar{\gamma}^{[K]})^{(\text{old})})$ in the product-space $\Pi = \mathcal{C} \times (0, \infty)$, then we classify all times series using the estimates $\{(\hat{c}^i, \hat{\gamma}^i)\}_{i=1}^N$, and in the last step we update the cluster centres $\boldsymbol{\mu}^{(\text{new})} = ((\bar{c}^{[1]}, \bar{\gamma}^{[1]})^{(\text{new})}, \ldots, (\bar{c}^{[K]}, \bar{\gamma}^{[K]})^{(\text{new})})$, as

$$
\begin{aligned}
\boldsymbol{\mu}^{(\text{new})} &= \arg\min_{\boldsymbol{\mu} \in \Pi} \sum_{i=1}^{N} \sum_{k=1}^{K} I(\mathcal{L}^{(\text{new})}(i) = k) D_\alpha((\hat{c}^i, \hat{\gamma}^i), (\bar{c}^{[k]}, \bar{\gamma}^{[k]})) \\
&= \arg\min_{\boldsymbol{\mu} \in \Pi} \sum_{k=1}^{K} \left\{ \alpha \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} \int_0^1 \{\hat{c}^i(w) - \bar{c}^{[k]}(w)\}^2 \mathrm{d}w + (1-\alpha) \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} (\hat{\gamma}^i - \bar{\gamma}^{[k]})^2 \right\} \\
&= \arg\min_{\boldsymbol{\mu} \in \Pi} \sum_{k=1}^{K} \left[\alpha \left\{ \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} \int_0^1 \{\hat{c}^i(w) - \bar{s}^{[k]}(w)\}^2 \mathrm{d}w + \hat{N}^{[k]} \int_0^1 \{\bar{s}^{[k]}(w) - \bar{c}^{[k]}(w)\}^2 \mathrm{d}w \right\} \right. \\
&\quad \left. + (1-\alpha) \left\{ \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} (\hat{\gamma}^i - \bar{\psi}^{[k]})^2 + \hat{N}^{[k]} (\bar{\psi}^{[k]} - \bar{\gamma}^{[k]})^2 \right\} \right],
\end{aligned}
$$

where $\bar{s}^{[k]}(w) = (\hat{N}^{[k]})^{-1} \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} \hat{c}^i(w)$, $\bar{\psi}^{[k]} = (\hat{N}^{[k]})^{-1} \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} \hat{\gamma}^i$, and $\hat{N}^{[k]} = \sum_{i=1}^N I(\mathcal{L}^{(\text{new})}(i) = k)$. Since,

$$
\sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} \int_0^1 \{\hat{c}^i(w) - \bar{s}^{[k]}(w)\}^2 \mathrm{d}w \geqslant 0, \qquad \sum_{\{i:\mathcal{L}^{(\text{new})}(i)=k\}} (\hat{\gamma}^i - \bar{\psi}^{[k]})^2 \geqslant 0,
$$

for each $k \in \{1, \ldots, K\}$, the expression in Step 2 of the K-means algorithm outlined in Section 2.3 is indeed minimized with $\bar{c}^{[k]}$ and $\bar{\gamma}^{[k]}$ as in (8).

20491573, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.560 by University Of Edinburgh, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License