Master Thesis Econometrics & Management Science

Modelling heteroscedastic extreme events in multivariate asset loss observations

Darby Macaraeg

373712

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Supervisor:

Dr. Chen Zhou

Second Assessor:

Dr. Alberto Quaini

Date final version:

April 21 2023

Abstract

Classical extreme value theory states that observations are identically distributed in each point in time. This view is rather reductive when considering asset return variables, as they often exhibit heteroscedasticity. This paper uses the scedasis function to model time variation in the frequency at which extreme asset losses occur, and to predict the one-step-ahead Value-at-Risk (VaR) and Expected Shortfall (ES). The results show that modelling for heteroscedastic extremes can improve prediction of the risk measures if one carefully approaches the bias-variance trade-off. Furthermore, I examine three methods to select the optimal amount of extreme events to use in estimating the scedasis, VaR and ES.

Contents

1 Introduction

When extreme events for financial observations are modelled by means of *classical extreme value theory* (CEVT), a widely used assumption is that the tail distributions are identical and independent over time (Jansen et al., 2000, Fernandez, 2005, Hyung and De Vries, 2007). However, the frequency in which extreme events occur may very well differ between different time periods. For instance, in times of crisis, extreme losses in stock returns might occur more often than in economically more stable times. Therefore, CEVT may underestimate or overestimate tail risk over certain periods of time. Einmahl et al. (2016) address modelling time variation in the tail distributions and refer to this phenomenon as *heteroscedastic extremes*. They discuss how to model this heteroscedasticity by means of a so called *scedasis* function, which allows us to model mild variation in the tail distributions, under the condition that they stem from a common family of distribution functions.

In this paper, I will examine how we can use the scedasis function to model heteroscedasticity in the tails of a multivariate financial data set, and whether we can use this model to analyse possible differences between the marginal tail distributions. Furthermore, I will investigate whether modelling for heteroscedasticity can improve the predictive performance of the marginal one-step ahead Value-at-Risk (VaR) and Expected Shortfall (ES) models when compared to the models with homoscedastic extremes. In order to conduct this research with the best possible estimates and predictions, I will additionally compare the performance of three methods of optimally selecting the amount of extremes used for estimating the scedasis and predicting the two risk measures. The findings of this research will help to improve the estimation of tail risk and provide more accurate predictions of extreme events in asset returns.

Modelling heterosedasticity in the context of extreme value theory (from now on referred to as *heteroscedastic extreme value theory* (HEVT)) is treated in, among others, De Haan et al. (2015). They introduce a semi-parametric estimator of a trend in the tail distributions, under the condition that they are within the maximum domain of attraction of a common distribution function. As a consequence, variation in the tail distribution must come solely from variation in the (scale) parameter values. For the marginal estimations and predictions in this research, I rely on the findings of Einmahl et al. (2016), who build on the idea of mild variation in the tail distributions and provide a non-parametric estimator and test procedures for the scedasis function in the univariate case. Their scedasis estimator makes use of kernel density estimation.

For the multivariate analysis, I will use the modified estimator and test procedures from Einmahl

et al. (2022), who expand to modelling heteroscedastisity in multivariate data on rainfall and provide a multivariate estimator for the scedases. Furthermore, the papers provide a combined set of assumptions which must hold for the marginal tail distributions in order to permissibly make use of the estimators and testing procedures. The first assumption is that the data set does not suffer from serial dependency. As this concept solely addresses the violation of the CEVT assumption that distributions are *identical*, it is still tied to the assumption that observations are *independent*. The second assumption is that the marginal tail indices remain constant over time. The scedasis effectively measures the differences in the tail distributions only up to a scale, which implies that it is not able to model time variation in the extreme value index. The last assumption is that the marginal tail indices are equal in the cross-section. In order to compare difference in frequency of extreme losses between the variables, their magnitudes (i.e. the tail index value) must be equal. This last assumption logically only applies to a multivariate data set and may be ignored when considering the (univariate) marginal distributions.

When modelling heteroscedastic extremes for financial observations, the question arises as to whether the data set satisfies the necessary set of assumptions provided in Einmahl et al. (2022). Therefore, a substantial part of this paper dedicated to testing the validity of the assumptions stated in Einmahl et al. (2022) on a financial data set, and to investigating how to adapt to possible violations. For this research, I will use the time series of loss return observations from 10 Industry Portfolios constructed by the Fama and French website.

Possibly the most disputed of the assumptions, when applied to a financial data set, is the one on serial independence. For instance, Tversky and Kahneman (1985) and Poterba and Summers (1988) find evidence of mean reversion in stock returns, Jegadeesh and Titman (1993) claim the presence of *momentum* in stock returns and there is an extensive amount of research on volatility clustering (e.g. Granger and Ding, 1995, Jacobsen and Dannenburg, 2003, Cont, 2007). However, when a data set exhibits serial dependence, (most of) it may stem from the center of the distribution. As this research focuses on the tails of the distribution, I will only consider serial dependency in the extreme events. The widely used *extremal index* (Smith and Weissman, 1994, Ancona-Navarrete and Tawn, 2000, Moloney et al., 2019) allows us to measure the existence and magnitude of clustering in the extremes. The value of the extremal index tells us how many consecutive observations the dependence lasts. Therefore, the data set can be cleaned by periodically discarding observations, such that the cleaned data set has independent observations.

To test for constant marginal tail indices over time I will use the test procedure from Einmahl

et al. (2016). They suggest dividing the data sample into subsets over which the scedasis estimates are equal, such that possible variation in the tail distribution over these subsets must come from variation in the tail index. To test for equal tail indices in the cross-section I turn to Stupfler (2019), who provide a testing procedure for equal tail indices when the marginals are spatially dependent and heavy-tailed.

After achieving a processed data set for which the assumptions reasonably hold, I will examine whether the heteroscedastic risk measure models are an improvement with respect to their homoscedastic counterparts. If one can improve the estimation of tail risk, it should also improve the prediction of tail risk measures. Einmahl et al. (2016) provide a quantile predictor which I will use to predict one-step-ahead forecasts of the VaR and ES. To examine whether predictive performance can in fact be improved, I will use backtesting methods to compute test statistics on the validity of the VaR and ES (see Shirazi (2014) and Acerbi and Szekely (2014), respectively) for both HEVT and CEVT models and compare their performances.

As both the marginal and multivariate scedasis estimators rely on a fraction of the tail of the empirical distribution, one may be able to improve the estimations by optimally choosing the amount extreme events to take into account. Einmahl et al. (2016) prove that the tail index estimator from Hill (1975) is still a suitable estimator when extremes are heteroscedastic, and they show that the threshold used for the tail index estimator can be used equivalently as a threshold for the scedasis estimator. Whereas both Einmahl et al. (2016) and Einmahl et al. (2022) use a fixed percentage of the sample size, there exist method of selecting the optimal fraction size that have more theoretic support. For this research I will consider three methods. Two of the methods are proposed by Danielsson et al. (2019). The first optimises the sample fraction size by minimising the distance between the empirical quantile and a theoretical quantile. The second is a heuristic quantification of looking at a Hill plot and somewhat guessing where the variance is sufficiently low and the bias not too high. The third method is from Hall (1990) and uses bootstrap methods to optimally balance bias and variance in the Hill estimator. Performance will be tested by running simulations and evaluate how accurate the scedasis estimates and risk measure predictions are when using each of the three sample size selection methods.

This research concludes that HEVT models can predict the VaR and ES with higher accuracy than the CEVT models. However, one should pay attention to the variance of the VaR predictor and the bandwidth in kernel density estimator in the scedasis predictor, as they can cause high amounts of undesired VaR violations. Furthermore, this research reveals how the bias-variance

trade off dilemma applies to the heteroscedastic risk measure models and that the best performing sample fraction selecting method is the one that produces the lowest variance in the predictor.

The remainder of this paper is outlined as follows: Sections 2 and 3 are devoted to an elaborate description of HEVT and the sample fraction size selection methods, respectively. Section 4 contains a comparison of the performance of the three threshold selection methods and an analysis of the HEVT and CEVT prediction models based on various simulations. In Section 5, I examine whether modelling for homoscedastic extremes is justified in an empirical multivariate data set consisting of asset loss observations, and I use the best performing models from Section 4 to analyse the differences in the marginal tail distributions. Section 6 covers an overall conclusion and discussion.

2 Heteroscedastic EVT

In the first part of this section I will discuss how the scedasis function can be used to model heteroscedasticity in the tails of a distribution. In the second part I will explain how to extend this to predicting risk measures. In the third part I address the set of assumptions that is required to justify the use of the scedasis function.

2.1 Scedasis

2.1.1 Univariate

For a basic understanding of the scedasis function, I will first discuss it in a univariate context. Moreover, some univariate results and procedures are used for testing validity of the required assumptions, and for estimating and forecasting the marginal tail distributions. Now let $X_1, X_2, \ldots, X_n$ be a time series of serially independent asset loss-return observations. In order to model the fat tails of an asset loss distribution, classical extreme value theory (CEVT) states that the observations are i.i.d. with distribution function F, and that F satisfies the condition

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha}, \tag{2.1}$$

where $\alpha > 0$ is called the *tail index*. In the case that $X_1, X_2, \ldots, X_n$ follow the respective continuous and possibly varying distributions $F_1, F_2, \ldots, F_n$, such that $F_{i_1} = F_{i_2}$ does not necessarily hold for

$i_1 \neq i_2$, Einmahl et al. (2016) additionally suggest the condition

$$\lim_{x\to\infty} \frac{1 - F_i(x)}{1 - F_0(x)} = c\Big(\frac{i}{n}\Big), \tag{2.2}$$

for all $1 \leq t \leq n$, distribution function F_0 and *scedasis function* c. As F_0 satisfies the EVT Condition (2.1), I will refer to this expansion as *heteroscedastic extreme value theory* (HEVT). Specifically, F_0 belongs to the maximum domain of attraction of a Fréchet distribution. The scedasis function c, is a positive function defined on $[0, 1]$ and has the imposed condition

$$\int_0^1 c(s)ds = 1, \tag{2.3}$$

which makes the scedasis uniquely defined. Combining conditions (2.1) and (2.2) gives the following limit relation

$$\lim_{t\to\infty} \frac{1 - F_i(tx)}{1 - F_0(t)} = \lim_{t\to\infty} \frac{1 - F_i(tx)}{1 - F_0(tx)} \times \frac{1 - F_0(tx)}{1 - F_0(t)} = c\Big(\frac{i}{n}\Big)x^{-\alpha}. \tag{2.4}$$

Equation (2.4) essentially states that each distribution F_i, for $i = 1, 2, \ldots, n$, belongs to the maximum domain of attraction of a Fréchet distribution. Furthermore, it implies that the tails of the distribution functions F_i must be comparable up to a scale to the tail of a common distribution function F_0, and not necessarily identical. This condition requires that the tail indices are equal for all i, i.e. the tail index remains constant over time. This is one of the assumptions that must be tested on the data set in order to justify the use of the scedasis function. Complete equality of the tail distributions occurs when $c(s) = 1$, which is a useful condition for hypothesis testing as this indicates that extremes are homoscedastic. Another assumption that must be tested for is the assumption of serial independence. As HEVT deviates from CEVT in the sense that it allows for relaxation of the identical distribution assumption, we are still tied to assuming serially independent distributions.

Due to Condition (2.3) and the fact that Equation (2.2) can never be negative, $c(s)$ can be considered a density function. Consequently, the integrated scedasis function $C(s) = \int_0^s c(u)du$ represents a cumulative distribution function. Given this perspective of the scedasis function, c can be interpreted as the frequency of extremes and C should be proportional to the number of exceedances of a high threshold on the first $\lfloor ns \rfloor$ observations. The latter is the inspiration for

Einmahl et al. (2016) to come up with the following estimator for the integrated scedasis function

$$\hat{C}(s) := \frac{1}{k} \sum_{i=1}^{\lfloor ns \rfloor} \mathbb{1}\{X_i \geq X_{(n-k)}\}, \tag{2.5}$$

where $\mathbb{1}\{\cdot\}$ is an indicator function, $\lfloor \cdot \rfloor$ is the floor function, $X_{(j)}$ is the jth ranked observation such that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ and $k = k(n)$ is an *intermediate* sequence for which holds

$$\lim_{n\to\infty} k = \infty \tag{2.6}$$

and

$$\lim_{n\to\infty} \frac{k}{n} = 0. \tag{2.7}$$

When testing the integrated scedasis function, one can test the null hypothesis $c = c_0$ without having to estimate c, by equivalently testing $C = C_0$, where $C_0(s) := \int_0^s c_0(u)du$. For instance, testing for homoscedastic extremes (i.e. $c(s) = 1$) against heteroscedastic extremes ($c(s) \neq 1$) would come down to testing whether C is the identity function. Einmahl et al. (2016) provide two types of test statistic to execute such hypothesis testing, one of Kolmogorov -Smirnov-type and one of Cramér-von-Mises-type. Both test statistics can be obtained by computing the estimator in Equation (2.5).

While estimating C is sufficient for hypothesis testing, we still need an estimator for c to forecast, or conduct a more thorough analysis of, tail risk when extremes are heteroscedastic. Einmahl et al. (2016) propose a kernel density estimation method with continuous, symmetric kernel function G on $[-1, 1]$ such that $\int_{-1}^{1} G(u)du = 1$ and $G(u) = 0$ for $|u| > 1$. For bandwidth $h = h(n) > 0$, where $h \to 0$ and $kh \to \infty$ as $n \to \infty$, the function c can be estimated non-parametrically by

$$\hat{c}(s) := \frac{1}{kh} \sum_{i=1}^{n} \mathbb{1}\{X_i \geq X_{(n-k)}\} G\left(\frac{s - \frac{i}{n}}{h}\right), \tag{2.8}$$

for $s \in (0, 1)$. Now let $\gamma = 1/\alpha$, such that the well known tail index estimator from Hill (1975) is denoted as

$$\hat{\gamma}(k) = \frac{1}{k} \sum_{i=1}^{k} \log(X_{(n+1-i)}) - \log(X_{(n-k)}).$$

Mason (1982) show that for intermediate sequence $k = k(n)$, the Hill estimator is consistent and Einmahl et al. (2016) prove that the asymptotic distribution of the Hill estimator does not depend

on c or $\hat{C}$. They use this knowledge to conclude that *gamma* is still a valid estimator for the extreme value index in the case of heteroscedastic extremes. Therefore, the optimal threshold used in the Hill estimator can be used for estimating the scedasis function and integrated scedasis function as well.

2.1.2 Multivariate

To make the switch to multivariate data, consider $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m$ to be m time series of asset loss observations, where $\boldsymbol{X}_j = (X_{1,j}, X_{2,j}, \ldots, X_{n,j})'$ for asset $j = 1, \ldots, m$. Furthermore, denote $F_{i,j}$ as the distribution function of asset j at time i. Similar to Einmahl et al. (2022), I assume the multivariate variant of the limit relation (2.2)

$$\lim_{x\to\infty} \frac{1 - F_{i,j}(x)}{1 - F_0(x)} = c\Big(\frac{i}{n}, j\Big), \tag{2.9}$$

which must hold for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Here, $c(\cdot, j)$ is the scedasis function for asset j and F_0 still satisfies the heavy tail condition in Equation (2.1). Equation (2.9) implies that the tail index must be equal for all $X_{i,j}$, i.e. it must remain constant within each time series and in the cross-section. Equality across the marginal tail indices must be tested on a multivariate data set, in addition to the two assumptions mentioned in the univariate context. This condition is necessary to ensure that possible heteroscedasticity in the tail distributions lies in the frequency in which extreme events occur, and not in their magnitude. Particularly, since we have that

$$\lim_{t\to\infty} \frac{1 - F_{i,j}(tx)}{1 - F_0(x)} = c\Big(\frac{i}{n}, j\Big)x^{-\alpha}, \tag{2.10}$$

each $F_{i,j}$ belongs to the maximum domain of attraction of a common Fréchet distribution. To ensure identification of the scedasis function, they provide the condition

$$\sum_{j=1}^{m} C_j(1) = 1,$$

where for $0 \leq s \leq 1$

$$C_j(s) = \frac{1}{m}\int_0^s c(u, j)du$$

is the integrated scedasis function for asset j. As we are interested in differences of the scedasis functions between assets, C_j should be estimated with the same threshold for all m assets. Einmahl

et al. (2022) propose to take a high quantile from all $N := n \times m$ observations as a common threshold. Now let $X_{(N-k):N}$ be the $(N-k)$-th ranked observation of all N observations, such that $X_{(1):N} \leq X_{(2):N} \leq \cdots \leq X_{(N):N}$, they then present the following estimator for C_j

$$\widehat{C}_j(s) := \frac{1}{k} \sum_{i=1}^{\lfloor ns \rfloor} \mathbb{1}\{X_{i,j} \geq X_{(N-k):N}\}, \tag{2.11}$$

where conditions (2.6) and (2.7) still hold for intermediate sequence for $k = k(n)$. A value for k can be chosen optimally by choosing the optimal sample fraction size to estimate the tail index of all N observations. They establish the joint asymptotic behaviour of estimators $\hat{C}_j$ in order to test equality of the scedasis over time and over the cross section.

To test whether the scedasis is constant over all assets, I test $H_0 : C_j(1) = \frac{1}{m}$ for $j = 1, \ldots, m$, and carry out testing by checking whether the limit vector $D = \Big(\sqrt{k}\big(\widehat{C}_1 - \frac{1}{m}\big), \sqrt{k}\big(\widehat{C}_2 - \frac{1}{m}\big), \ldots, \sqrt{k}\big(\widehat{C}_m - \frac{1}{m}\big)\Big)$ has mean zero via an adapted χ^2 test. Specifically, let $B = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}'_m$, where I_m and $\mathbf{1}_m$ are the identity matrix and unit vector with dimension m, respectively. From Einmahl et al. (2022) we have test statistic

$$T_1 := D'_{m-1}\big((B\hat{\Sigma}B')_{m-1}\big)^{-1}D_{m-1}. \tag{2.12}$$

Here, subscript $m-1$ indicates the first $m-1$ elements of the corresponding vector. The invertible matrix $\hat{\Sigma}$ is an empirical estimator for the covariance matrix of the bivariate Wiener processes $W_j\big(1, C_j(1)\big)$, for $j = 1, 2, \ldots, m$, which is used in a Skorokhod construction to derive the asymptotic distribution of estimator $\hat{C}$. For an elaborate explanation of the specific Wiener process I refer to Einmahl et al. (2022). Now, under H_0, we have that $T_1 \xrightarrow{d} \chi^2(m-1)$ as the amount of time series observations grows to infinity.

In order to test for a constant scedasis over time, I follow Einmahl et al. (2022) and test $H_0 : C_j(s) = sC_j(1)$, for $0 \leq s \leq 1$ and asset $j = 1, 2, \ldots, m$. They show that the limit distribution of the process $\big\{\sqrt{k}\big(\widehat{C}_j(s) - s\widehat{C}_j(1)\big)\big\}_{0\leq s\leq 1}$ is a Brownian bridge, such that the following Kolmogorov-Smirnov type test statistic can be used. Let B be a Brownian bridge, then for fixed asset j we have that under H_0

$$T_{2,j} = \left\{\sqrt{k\hat{C}_j(1)}\Big(\frac{\hat{C}_j(s)}{\hat{C}_j(1)} - s\Big)\right\}_{s\in[0,1]} \xrightarrow{d} \big\{B(s)\big\}_{s\in[0,1]}. \tag{2.13}$$

2.2 Tail risk measures

2.2.1 Value-at-risk

If we are able to model heteroscedasticity in extreme asset losses, it should also be incorporated in risk measures. Here I will discuss the one-step ahead prediction of the marginal Value-at-Risks. Recall that the VaR is a quantile value that is exceeded by a random variable with a certain (small) probability. Specifically, for asset j at time i and small probability p, we have

$$\mathrm{VaR}_{1-p;i,j} = \{x : P[X_{i,j} > x] = p\} = F^{\leftarrow}_{i,j}, (1-p) \tag{2.14}$$

where $\leftarrow$ denotes the left inverse. From Einmahl et al. (2016), we have the following procedure that will lead to the one-step ahead VaR predictor $\widehat{\mathrm{VaR}}_{1-p:n+1,j}$. Let U_0 denote the quantile function

$$U_0 := \Big(\frac{1}{1-F_0}\Big)^{\leftarrow}.$$

Then, for given asset j, we have

$$U_{i,j} := \Big(\frac{1}{1-F_{i,j}}\Big)^{\leftarrow},$$

at time points $i = 1, 2, \ldots, n$. Equation (2.9) and the assumption that F_0 satisfies Condition (2.1) imply that

$$\begin{aligned}
\lim_{t\to\infty} \frac{U_{i,j}(tx)}{U_0(t)} &= \lim_{t\to\infty} \left(\frac{1-F_0(t)}{1-F_{i,j}(tx)}\right)^{\leftarrow} \\
&= \lim_{t\to\infty} \left[\left(\frac{1-F_{i,j}(tx)}{1-F_0(t)}\right)^{-1}\right]^{\leftarrow} \\
&= \left[\left(\lim_{t\to\infty} \frac{1-F_{i,j}(tx)}{1-F_0(tx)} \times \frac{1-F_0(tx)}{1-F_0(t)}\right)^{-1}\right]^{\leftarrow} \\
&= \left[\left(c\Big(\frac{i}{n}, j\Big)x^{-\alpha}\right)^{-1}\right]^{\leftarrow} \\
&= c\Big(\frac{i}{n}, j\Big)^{1/\alpha} x^{1/\alpha}.
\end{aligned}$$

Using the definition of the VaR from Equation (2.14), with very small probabilities p and q for which $p < q$, in the limit relation from Equation (2.10), we get

$$\frac{1-F_{i,j}(\mathrm{VaR}_{1-p;i,j})}{1-F_0(\mathrm{VaR}_{1-q;i,j})} = \frac{p}{q} \approx c\Big(\frac{i}{n}, j\Big)\left(\frac{\mathrm{VaR}_{1-p;i,j}}{\mathrm{VaR}_{1-q;i,j}}\right)^{-\alpha},$$

which after rewriting gives

$$\text{VaR}_{1-p;i,j} = U_{i,j}\left(\frac{1}{p}\right) \approx \text{VaR}_{1-q;i,j}\left(\frac{c\left(q\frac{i}{n}, j\right)}{p}\right)^{1/\alpha}.$$

They suggest to substitute empirical probability k/n for q and empirical quantile $X_{(n-k),j}$ for $\text{VaR}_{1-q;i,j}$. The value for k is chosen to be the amount of extreme events used for estimating both the scedasis and the extreme value index. We now obtain the VaR estimator

$$\widehat{\text{VaR}}_{1-p:i,j} = X_{(n-k),j}\left(\frac{k\hat{c}\left(\frac{i}{n}, j\right)}{np}\right)^{\hat{\gamma}}, \tag{2.15}$$

where $X_{(n-k),j}$ is the $(n-k)$-th ranked observation of asset j such that $X_{(1),j} \leq X_{(2),j}, \cdots \leq X_{(n),j}$, and $\hat{c}$ is the scedasis estimator from Equation (2.8).

To extend this estimator to a one-step ahead predictor, we must estimate the scedasis $\hat{c}(1, j)$ at boundary point 1. The one-step ahead forecast of a high quantile is computed as

$$\widehat{\text{VaR}}_{1-p:n+1,j} = X_{(n-k),j}\left(\frac{k\hat{c}(1, j)}{np}\right)^{\hat{\gamma}}, \tag{2.16}$$

where $\hat{c}(1, j)$ is the out-of-sample scedasis predictor

$$\hat{c}(1, j) := \frac{1}{kh}\sum_{i=1}^{n} \mathbb{1}\{X_{i,j} \geq X_{(n-k),j}\} G_b\left(\frac{1-\frac{i}{n}}{h}\right), \tag{2.17}$$

with boundary kernel function G_b, which is defined as

$$G_b(x) = \frac{\int_0^1 u^2 G(u)du - x\int_0^1 uG(u)du}{\frac{1}{2}\int_0^1 u^2 G(u)du - \left(\int_0^1 uG(u)du\right)^2} G(x), \tag{2.18}$$

for a "regular" kernel function G. This particular boundary corrected kernel has as disadvantage that it can take on negative values. A negative value for the kernel may produce a negative estimated scedasis value, for which we know it to be strictly positive. Therefore, I follow Jones and Foster (1996) to correct for the boundary kernel estimate which take on a negative value. For kernel function G, I follow Einmahl et al. (2016) by using the bi-weight kernel

$$G(x) = \frac{15}{16}(1-x^2)^2,$$

for $x \in [-1, 1]$ and $G(x) = 0$ otherwise, and setting the bandwidth equal to $h = 0.1$.

To test the performance of the VaR predictions when modelling for heteroscedastic extremes, I will use the binomial backtesting method and compare the test results with the results obtained when homoscedastic extremes are modelled. The test is based on the amount of VaR violations over that occur over a test sample. For a test sample of size n_T and fixed asset j, we have the count variable

$$v_{i,j} = \mathbb{1}\{X_{i,j} \geq \widehat{\text{VaR}}_{1-p:i,j}\},$$

for $i = n+1, n+2, \ldots, n+n_T$. If $\widehat{\text{VaR}}_{1-p:i,j}$ is specified correctly, we have that $v_{i,j} \sim B(1, p)$, i.e. $v_{i,j}$ follows a Bernoulli distribution with probability p. Consequently, denote $V_j = \sum_{i=n+1}^{n+n_T} v_{i,j}$ as the total violations count of the predicted VaR for asset j. We now have that $V_j \sim B(n_T, p)$. To test the null hypothesis of a correct coverage of the VaR model, we can approximate the rejection region by exploiting the following convergence in distribution. For asset j and small probability p, we have under the null hypothesis that

$$T_3 = \frac{V_j - n_T p}{\sqrt{n_T p(1-p)}} \xrightarrow{d} N(0, 1), \tag{2.19}$$

as $n_T \to \infty$.

2.2.2 Expected shortfall

Another widely used risk measure is the expected shortfall, which essentially measures the expected size of a VaR violation. For fixed asset j, the expected shortfall can be expressed in mathematical terms as

$$\text{ES}_{1-p;i,j} = \text{E}[X_{i,j} | X_{i,j} > \text{VaR}_{1-p;i,j}] = \frac{1}{p} \int_{1-p}^{1} \text{VaR}_{u;i,j} du, \tag{2.20}$$

for loss $X_{i,j}$ on day i and small probability p. In order to estimate the expected shortfall I use the limit relation

$$\lim_{p \to 0} \frac{\text{ES}_{1-p;i,j}}{\text{VaR}_{1-p;i,j}} = \frac{\alpha}{\alpha - 1}, \tag{2.21}$$

for tail index α. Now, when we have obtained an estimate $\widehat{\text{VaR}}_{1-p}$ for a very small probability p, one can estimate ES_{1-p} quite easily by

$$\widehat{\text{ES}}_{1-p;i,j} = \widehat{\text{VaR}}_{1-p;i,j} \frac{\hat{\alpha}}{\hat{\alpha} - 1}, \tag{2.22}$$

where $\hat{\alpha} = 1/\hat{\gamma}$ is the reciprocal of the Hill estimate $\hat{\gamma}$. This estimator can now be extended to the one-step ahead predictor

$$\widehat{\mathrm{ES}}_{1-p;n+1,j} = \widehat{\mathrm{VaR}}_{1-p;n+1,j} \frac{\hat{\alpha}}{\hat{\alpha} - 1}, \tag{2.23}$$

To test the performance of the expected shortfall estimator, I use the first of the three backtesting statistics proposed in Acerbi and Szekely (2014). Technically, their method jointly tests the performances of the ES model and the VaR model, as their performances can not be distinguished from one another through backtesting. However, given the fact that heteroscedasticity of the extremes is modelled in the ES through the VaR, ranking the ES and VaR performances jointly is not necessarily a bad thing when comparing HEVT models with CEVT models. The statistic is used to test the magnitude of the realized violations against the model predictions, after the validity of the VaR_{1-p} has been tested. We can rewrite Equation (2.20) as

$$\mathrm{E}\left[\frac{X_{i,j}}{\mathrm{ES}_{1-p;i,j}} - 1 | X_{i,j} > \mathrm{VaR}_{1-p;i,j}\right] = 0, \tag{2.24}$$

and their testing procedure aims to estimate this expected value. I will perform the backtest on a test sample of size n_T, such that for $H = \sum_{i=n+1}^{n+n_T} \mathbb{1}\{X_{i,j} > \widehat{\mathrm{VaR}}_{1-p;i,j}\}$ the test statistic is defined as

$$T_{4,j} = \frac{\sum_{i=n+1}^{n+n_T} \frac{X_{i,j} \mathbb{1}\{X_{i,j} > \widehat{\mathrm{VaR}}_{1-p;i,j}\}}{\widehat{\mathrm{ES}}_{1-p;i,j}}}{H} - 1, \tag{2.25}$$

under the condition that $H > 0$, i.e. at least one exceedance of the VaR_{1-p} has occurred over $i = n+1, n+2, \ldots, n+n_T$. Given the fact that magnitude of the violations is averaged over all violations, this test is insensitive to the amount of violations. Therefore, performance of the ES model is isolated from the VaR model as much as possible. I will use this statistic to perform a one sided test of the null hypothesis that the ES model is adequate, against the alternative hypothesis that risk is underestimated. Under the null it must hold that $\mathrm{E}[T_4 | H > 0] = 0$ and under the alternative that $\mathrm{E}[T_4 | H > 0] > 0$. A critical value for the test is obtained by means of Monte Carlo simulations.

2.3 Testing assumptions

2.3.1 Serial dependence

As this research is focused on extreme events, I will test for potential serial dependency solely in the tails of the distributions. One can suitably test for such dependency by examining the so called

extremal index, which can be defined as follows. Let $X_i^* = \max\{X_{i,1}, X_{i,2}, \ldots, X_{i,m}\}$ on day i, such that $(X_1^*, X_2^*, , \ldots, X_n^*)$ is a sequence of maximum daily loss observations with distribution F^* and u a high threshold. Furthermore, let there exist some d for which it holds that exceedances in different blocks of the sequence are independent, and that within a block exceedances are not too far from one another (conditions from Leadbetter et al. (1983) and Leadbetter and Nandagopalan (1989)), then the extremal index θ

$$\theta = \lim_{n\to\infty} \mathrm{P}\Big[\max_{2\leq i\leq d} X_i^* \leq u_n \Big| X_1^* > u_n\Big].$$

can be interpreted as the reciprocal of the mean cluster size of the extremes in the limiting point process of exceedance times over high threshold (Robert et al., 2008), i.e. $\theta \approx \frac{1}{\text{mean clustering size}}$. This means that if $\theta = 1$, the extremes are serially independent. If, alternatively, $\theta < 1$, extremes occur in clusters of size $1/\theta$ on average.

For estimating the extremal index I will use the sliding blocks estimator from Berghaus and Bücher (2018). Robert et al. (2008) provide evidence that, even though the sliding blocks are strongly dependent, the sliding block estimator is asymptotically more efficient than an estimator based on disjoint blocks. Berghaus and Bücher (2018) show that the sliding blocks estimator outperforms other blocks estimators, and that it is competitive to runs- and inter-exceedance estimators in various models. The sliding blocks estimator is constructed as follows. For block size r, $0 < r \leq n$, and block maximum $M_t = \max\{X_t^*, X_{t+1}^*, \ldots, X_{t+r-1}^*\}$, let

$$Y_t := r(1 - F^*(M_t)) \tag{2.26}$$

and

$$\tilde{\theta} = \left(\frac{1}{n-r+1}\sum_{t=1}^{n-r+1} Y_t\right)^{-1}. \tag{2.27}$$

By plugging in the empirical distribution

$$\bar{F}^*(M_t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[X_i^* \leq M_t]$$

for F^* in Equation (2.26) we obtain an estimator $\hat{Y}$ and, subsequently, plugging in $\hat{Y}$ in Equation (2.27) provides an estimator $\hat{\theta}$ for the extremal index. The extremal index estimates are computed for a large range of block sizes, such that an optimal block size can be chosen empirically by taking

a value r^*, such that taking any value r, for which holds $r > r^*$, would not affect the value of the estimate by much.

2.3.2 Constant tail indices

As the scedasis function measures differences in tail distributions only up to a scale, the tail index should be constant over time and in the cross-section in order to make comparisons between the tail distributions at different time points and between the assets. To test whether the extreme value index is constant over time for a given asset j, I will use a test statistic from Einmahl et al. (2016), where they suggest dividing the observations of asset j into $L > 1$ subsamples, such that $0 = s_{0,j} < s_{1,j} < s_{2,j} \cdots < s_{L-1,j} < s_{L,j} = 1$ and $s_{l,j} = \sup\{s : \hat{C}_j(s) \leq l/L\}$. Now each block has equal intensity $\hat{C}_j(s_{l,j}) - \hat{C}_j(s_{l-1,j}) = 1/l$, such that differences between the tail distributions can not be explained by the scedasis functions, but solely by varying extreme value indices. Then, I will calculate the Hill estimates for each of the $l = 1, 2, \ldots, L$ subsamples, which exclude their infimum but include their supremum, denoted by $\hat{\gamma}_{(s_{l-1},s_l]}$, by using the highest $k_j/L+1$ observations. Here, k_j is the fraction size used to compute the Hill estimate for the complete data set of asset j, $\hat{\gamma}_j$. Under the null hypothesis that the tail index $\alpha = 1/\gamma$ is constant over time, the test statistic

$$T_{5,j} = \frac{1}{L} \sum_{l=1}^{l} \left(\frac{\hat{\gamma}_{(s_{l-1,j},s_{l,j}]}}{\hat{\gamma}_j} - 1 \right)^2$$

has limit relation

$$k_j T_{5,j} \xrightarrow{d} \chi^2(L-1). \tag{2.28}$$

2.3.3 Equal tail indices

For testing equality of the extreme value indices across different assets I turn to Stupfler (2019), who provide a test for a multivariate distribution with heavy-tailed and cross-sectionally dependent marginals. Specifically, I use Corollary 4 of the paper, which illustrates the bivariate case. Suppose we have j_1 and j_2, two different assets with loss distributions F_{j_1} and F_{j_2}. Assume there is a function R on $[0,\infty]^2 \backslash (\infty, \infty)$, for which

$$\lim_{t\to\infty} t\mathrm{P}\left(1 - F_{j_1}(X_{j_1}) \leq \frac{x_{j_1}}{t}, 1 - F_{j_2}(X_{j_2}) \leq \frac{x_{j_2}}{t}\right) = R(x_{j_1}, x_{j_2}), \tag{2.29}$$

for any $(x_{j_1}, x_{j_2}) \in [0,\infty]^2 \backslash (\infty,\infty)$. Let $k_1 = k_1(n)$ and $k_2 = k_2(n)$ be two intermediate sequences and auxiliary functions A_1 and A_2 with constant sign and converging to 0 at infinity, such that they additionally hold

$$\sqrt{k_1} A_1\Big(\frac{n}{k_1}\Big) \to 0, \qquad \sqrt{k_2} A_2\Big(\frac{n}{k_2}\Big) \to 0$$

and

$$\frac{k_1}{k_2} \to a \in (0,\infty).$$

Furthermore, let

$$R_{j_1,h} = \sum_{i=1}^{n} \mathbb{1}\{X_{i,j_1} \le X_{h,j_1}\} \quad \text{and} \quad R_{j_2,h} = \sum_{i=1}^{n} \mathbb{1}\{X_{i,j_2} \le X_{h,j_2}\}$$

be the ranks of observations X_{h,j_1} and X_{h,j_2}, respectively. To estimate the upper tail dependence structure in Equation (2.29) I use their estimator $\hat{R}_{k_1}(x_{j_1}, x_{j_2})$, which is defined as

$$\hat{R}_{k_1}(x_{j_1}, x_{j_2}) = \frac{1}{k_1} \sum_{h=1}^{n} \mathbb{1}\{R_{j_1,h} \ge n - k_1 x_{j_1},\ R_{j_2,h} \ge n - k_1 x_{j_2}\}.$$

Then, to test for an identical pair of extreme value indices, $H_0 : \gamma_1 = \gamma_2$ for two assets $j_1 \neq j_2$, we have for test statistic T_6 that the following limit relation holds under the null hypothesis

$$T_{6,j_1,j_2} = k_1 \frac{(\hat{\gamma}_1 - \hat{\gamma}_2)^2}{\hat{\gamma}_1^2 + c\hat{\gamma}_2^2 - 2\hat{\gamma}_1\hat{\gamma}_2\hat{R}_{k_1}(a,1)} \xrightarrow{d} \chi^2(1), \tag{2.30}$$

when $a \neq 1$.

3 Sample fraction selection

The estimations of the scedasis function and the integrated scedasis function, as well as the prediction of the scedasis function, depend on k, the number of (positive) values considered to be extreme. This number is also used in the Hill-estimator and the sequence $k = k(n)$ is assumed to have the same asymptotic arguments for all three estimators, which suggest we may use the one and the same value for k. This section compares three existing methods for selecting an optimal value for k, which will be used to optimise the estimator of the scedasis and the predictors of the VaR and ES ($\hat{c}$, $\widehat{\mathrm{VaR}}_{1-p:i,j}$ and $\widehat{\mathrm{ES}}_{1-p:i,j}$ from Equations (2.8), (2.15) and (2.20), respectively).

Most existing methods for selecting an optimal fraction size k are computed by balancing the

bias-variance trade-off of the tail index estimator. For relatively smaller values of k the variance of the Hill-estimator is large, and for larger values of k its variance reduces but the bias increases. This trade-off is not as prominently featured in the scedasis estimator, as the variance of the scedasis estimator is rather low for any value of k. This can be explained by considering where the variance in the Hill-estimator comes from, namely the large differences in magnitude between extreme events. When threshold $X_{(n-k)}$ is high, (ergo, k is low) the difference between observations $X_{(n-k)}$ and $X_{(n-(k+1))}$ is generally large, causing large fluctuations in the Hill-estimator. As the scedasis estimator solely considers the frequency of exceedances of a high threshold and not the magnitude of the exceedance, an increase from k to $k+1$ affects the scedasis estimate only by roughly the weight given by the kernel to an increase of the frequency. Equations (2.15) and (2.20) show that the variance and bias of the risk measure estimates are more sensitive to the trade-off in the tail index estimator when there is no trade-off in the scedasis estimator. This justifies optimising the risk measure estimators by choosing k for which the bias-variance trade-off of the Hill-estimator is balanced properly. Two of the three following sample fraction size selection methods rely on such an objective. The third method selects k by minimizing the distance between the empirical distribution and a theoretical semi-parametric distribution.

3.1 Hall

To choose the optimal number of extreme events, Hall (1990) suggest to base the selection on minimising the mean squared error (MSE) of the Hill-estimator $\hat{\gamma}$ for a data set of size n

$$\mathrm{MSE}_n(k) = \mathrm{E}[(\hat{\gamma}_n(k) - \gamma)^2], \tag{3.1}$$

where the MSE is to be estimated by means of bootstrapping. However, due to the unsmooth nature of the Hill-estimator, drawing bootstrap resamples of the same size as the original sample will cause this estimate to significantly underestimate the bias, even though the bias is a large contributor to the MSE. Therefore, the size of the bootstrap resamples must be taken smaller than n, such that knowledge about the difference between the two sample sizes can be used to select the optimal value for k. Hall (1990) furthermore shows that resample size n_1 for which holds $n_1 \leq n^{1-\delta}$, for some $\delta > 0$, is small enough and grows slowly enough to obtain asymptotic optimality. This results in

the following procedure: take fraction size k_1 which minimizes the bootstrap estimate

$$\widehat{\mathrm{MSE}}_{n_1}(k_1) = \mathrm{E}[(\hat{\gamma}_{n_1}(k_1) - \hat{\gamma}_n(k_0))^2],$$

where $\hat{\gamma}_{n_1}$ and $\hat{\gamma}_n$ are the Hill-estimates of the bootstrap sample and original sample, respectively, and k_0 is a tuning parameter. Now let $k_1^* = \underset{k_1}{\mathrm{argmax}}\{\widehat{\mathrm{MSE}}_{n_1}(k_1)\}$, Hall (1990) prove that for a constant β, for which holds $0 < \beta < 1$, the quantity

$$k_{\mathrm{hall}} = k_1^* \Big(\frac{n}{n_1}\Big)^{\beta}$$

is asymptotic to the k which minimizes Equation (3.1). Algorithm 1 shows how the optimal value is obtained by mean of the Hall-method for values $n_1 = n^{1-\delta}$ with $\delta = 0.045$, $k_0 = 2\sqrt{n}$, $\beta = 2/3$ and $b = 1000$ bootstraps. As this method is based on asymptotic theory, one might question its performance in finite samples.

Algorithm 1: Hall-method

input : Amount of bootstrap resamples b, bootstrap resample size n_1, full sample size n, tuning parameter k_0

output: Optimal fraction of sample size

1 Draw b bootstrap resamples of size n_1

2 **for** $k_1 = 2, \ldots, n_1^+$ **do**

3 | Compute bootstrap estimates $\widehat{\mathrm{MSE}}_{n_1}(k_1) = \mathrm{E}[(\hat{\gamma}_{n_1}(k_1) - \hat{\gamma}_n(k_0))^2]$

4 **end**

5 $k_1^* = \underset{k_1}{\mathrm{argmin}}\{\widehat{\mathrm{MSE}}_{n_1}(k_1)\}$

6 **return** $k_{\mathrm{hall}}^* = k_1^* \big(\frac{n}{n_1}\big)^{2/3}$

3.2 Eye-Ball

The second method to select the optimal fraction size to estimate the scedasis and the extreme value index is a quantification of simply looking at a Hill-plot and choosing the first value after which the variance of the Hill-estimates becomes sufficiently small. It is designed by Danielsson et al. (2019) and called the "Eye-ball"-method for obvious reasons. The optimal fraction size is the smallest value for k such that over a rolling window w, no less than a specified percentage h of the

estimates $[\hat{\gamma}(k+1), \ldots, \hat{\gamma}(k+w)]$ falls within the region $[\hat{\gamma}(k) \pm \epsilon]$, i.e.

$$k^*_{\text{eye}} = \min\left\{k \in 2, ..., n^+ - w | h < \frac{1}{w}\sum_{i=1}^{w} \mathbb{1}\{\hat{\gamma}(k+i) \in [\hat{\gamma}(k) \pm \epsilon]\}\right\},$$

where n^+ is the number of positive observations in the data. Following Danielsson et al. (2019), I take for the rolling window w 1% of the full sample, $h = 90\%$ and for the bound $\epsilon = 0.3$. This method is rather heuristic and might be overly simplified. On the other hand, it is easy and fast to compute and it does not require any asymptotic assumptions, such that it might have an advantage in smaller samples over methods that do rely on asymptotic arguments. Algorithm 2 shows the step-by-step overview of the Eye-ball-method from Macaraeg (2021).

Algorithm 2: Eye-ball-method

input : Moving window size w, size ϵ of region around $\hat{\gamma}(k)$ and percentage h of estimates that should be within this region

output: Optimal fraction of sample size

for $k = 2, \ldots, n^+ - w$ **do**
  $z = \frac{1}{w}\sum_{i=1}^{w} \mathbb{1}[\hat{\gamma}(k+i) \in \{\hat{\gamma}(k) \pm \epsilon\}]$
  if $h < z$ **then**
    return $k^*_{\text{eye}} = k$
  end
end

3.3 KS-Distance metric

The last method for selecting the optimal amount of extreme values is introduced by Danielsson et al. (2019) and selects k by minimizing the distance between the empirical distribution and a theoretical semi-parametric distribution. Triggered by the findings in Bickel and Sakov (2008) that sub-sample bootstrap methods may fail in some important cases, their incentive was to develop a method that does not suffer from this problem but has more theoretical support than certain heuristic methods, such as the Eye-ball-method or taking a fixed percentage of the sample size as k. The result is a Kolmogorov-Smirnov (KS) like test statistic, although the distance is measured in the quantile dimension rather than the probability dimension. As a result this deviation from the classical KS type test statistic, estimation errors have smaller impact on quantile analysis.

In the context of this paper, given the limit condition in Equation (2.10), a scaled Pareto distri-

bution would be the best theoretical distribution to use as a reference to fit the empirical distribution to. The Pickands-Balkema-De Haan-theorem states that the far tails of any distribution within the maximum domain of attraction of a generalized extreme value distribution can be accurately approximated by a generalized Pareto distribution. The quantile function of a Pareto distribution can be empirically estimated by

$$q_{1-\bar{p}_e}(k) = X_{(n-k),j}\Big(\frac{k}{\bar{p}_e n}\Big)^{\hat{\gamma}(k)},$$

where $\bar{p}_e$ is the empirical probability e/n. Now the horizontal distance between the empirical quantile $X_{(e),j}$ and the quantile estimator is optimised over k for which the maximum distance over a range of extreme values is minimal. In mathematical terms, we get the distance measures

$$Q(k) = \sup_e |X_{(n-e),j} - q_{1-\bar{p}_e}(k)|,$$

for $e = 1, 2, \ldots, T$, $k = 1, 2, \ldots, K$ and $K < T$. Then, the optimal value for k obtained by the KS-Distance metric-method is

$$k^*_{\mathrm{KS}} = \operatorname*{argmin}_k \{Q(k)\}.$$

The following algorithm is used with values $T = 0.15 \times n$, i.e. 15% of the total sample, and $K = T - 1$.

Algorithm 3: KS-distance metric-method

input : Region T over which KS-distances are measured and range K of possible values for k

output: Optimal fraction of sample size

```
for k = 1, ..., K do
    for e = 1, ..., T do
        D(e, k) = |X_{(n-e),j} - q_{1-p̄_e}(k)|
    end
    Q(k) = max_e {D(e, k)}
end
return k*_KS = argmin_k {Q(k)}
```

4 Simulations

In this section I will investigate whether modelling for heteroscedastic extreme events can theoretically improve the prediction of the univariate VaR and expected shortfall, with respect to modelling these risk measures conform classical extreme value theory. I simulated four data-generating processes (DGPs) that satisfy the univariate conditions from Einmahl et al. (2016) (hence, serially independent and constant extreme value indices over time) and have various scedasis functions. Along with testing for improvement, I also examine which of the sample fraction size optimisation methods produces the most accurate and efficient estimator of the scedasis and predictors of the VaRs and expected shortfalls. The DGPs, for which I follow Einmahl et al. (2016) to a large extent, follow a Fréchet distribution and are defined as follows

1. DGP 1: Observations are i.i.d. and follow a Fréchet distribution with location parameter 0, scale parameter 1 and shape parameter 2, i.e. $F_{n,i}^{(1)}(x) = \exp(-x^{-2})$ for $x > 0$. Here,

$$c(s) = 1.$$

2. DGP 2: Observations follow a Fréchet distribution with location parameter 0, time varying scale parameter $0.5 + i/n$ and shape parameter 2, i.e. $F_{n,i}^{(2)} = \exp\big(-\big(\frac{x}{0.5+i/n}\big)^{-2}\big)$ for $x > 0$. Here,

$$c(s) = \frac{12}{13}(0.5+s)^2.$$

3. DGP 3: Observations follow a Fréchet distribution with location parameter 0, a time varying scale parameter $h(s)$ which faces a structural break, and shape parameter 2, i.e. $F_{n,i}^{(3)} = \exp\big(-\big(\frac{x}{h(i/n)}\big)^{-2}\big)$ for $x > 0$, where $h(s) = 2s + 0.5$ for $s \in [0, 0.5]$ and $h(s) = -2s + 2.5$ for $s \in (0.5, 1]$. Here,

$$c(s) = \begin{cases} \frac{12}{13}(0.5+2s)^2, & \text{if } 0 \le s \le 0.5 \\ \frac{12}{13}(2.5-2s)^2, & \text{if } 0.5 < s \le 1. \end{cases}$$

4. DGP 4: Observations follow a Fréchet distribution with location parameter 0, a time varying scale parameter $h(s)$ which faces multiple breaks, and shape parameter 2, i.e. $F_{n,i}^{(4)} = \exp\big(-\big(\frac{x}{h(i/n)}\big)^{-2}\big)$ for $x > 0$, where $h(s) = 0.8$ for $s \in [0, 0.4] \cup [0.6, 1]$, $h(s) = 20s - 7.2$ for $s \in (0.4, 0.5]$

and $h(s) = -20s + 12.8$ for $s \in (0.5, 0.6)$. Here,

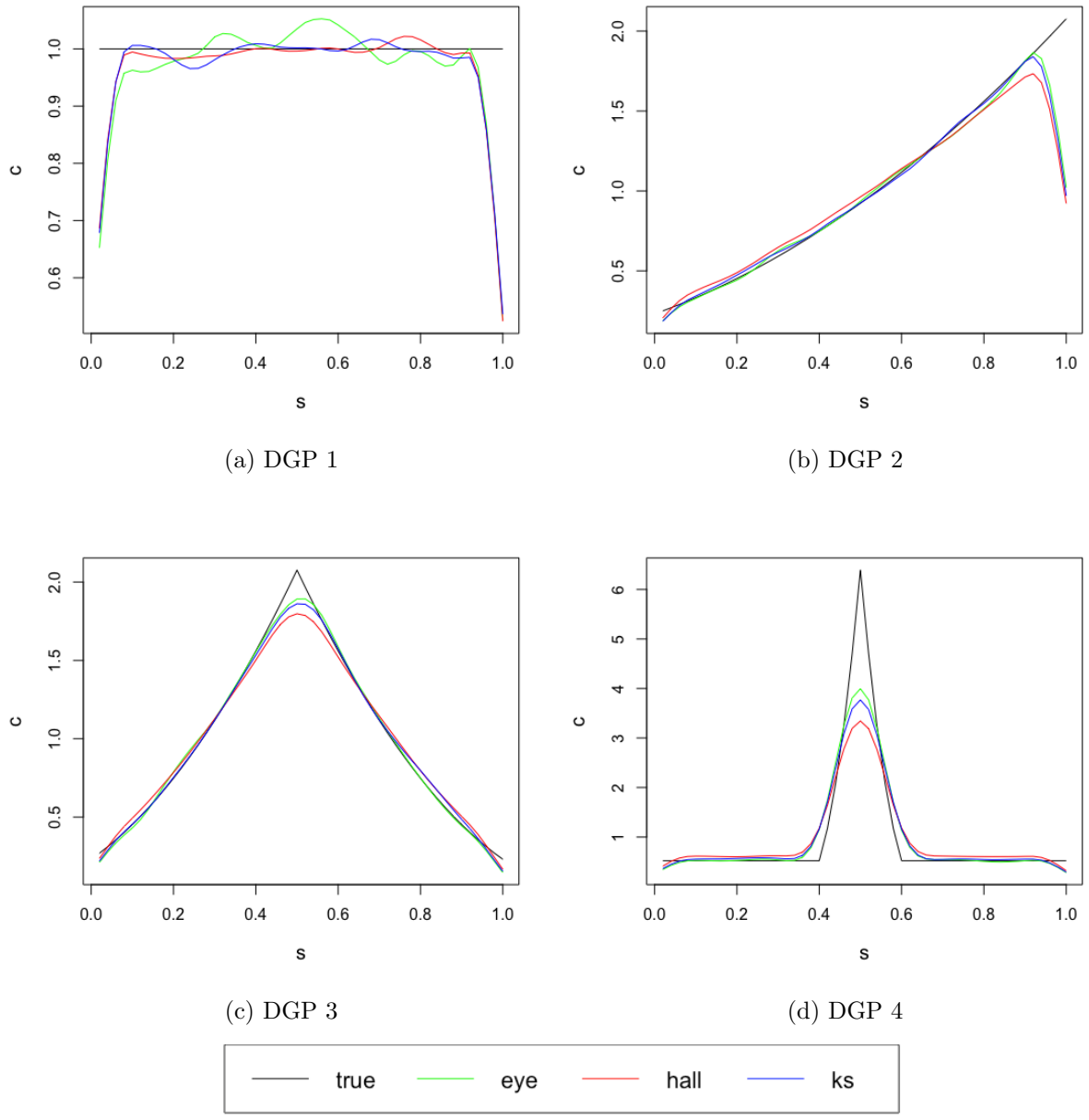
$$c(s) = \begin{cases} \frac{75}{92}0.8^2, & \text{if } 0 \leq s \leq 0.4 \vee 0.6 \leq s \leq 1 \\ \frac{75}{92}(-7.2 + 20s)^2, & \text{if } 0.4 < s \leq 0.5 \\ \frac{75}{92}(12.8 - 20s)^2, & \text{if } 0.5 < s < 0.6s. \end{cases}$$

The derivations of the scedasis functions are shown in Appendix A. For each DGP I simulate 500 data sets of size $n = 1500$. I deviate from Einmahl et al. (2016) by setting each shape parameter equal to 2 (ergo, tail index $\alpha = 2$), rather than 1. As statistical moments of extreme value distributions only exist up to the value of α, setting $\alpha = 2$ ensures that a second moment exists in each of the DGPs, which seems more realistic when dealing with financial data. More importantly, the limit relation in (2.21) for the expected shortfall is undefined when $\alpha = 1$.

4.1 Scedasis estimation performance

For all simulated data sets I computed the estimates $\hat{c}(s)$ over $s \in [0, 1]$ and selected optimal k by means of three different selection methods. Figure 1 shows per DGP the plots of their corresponding true scedasis functions, along with the average scedasis estimates resulting from the previously described "Eye-ball", "Hall" and "KS-distance metric" methods. It appears that the three methods perform quite similar in terms of average estimation accuracy. In DGP 1, the Eye-ball method estimates fluctuate more around the true value, compared to the other two methods. Furthermore, the Hall method seems to perform slightly worse at times where the scedasis suffers from a structural break in DGPs 3 and 4. Another thing that stands out, is the sudden drop in accuracy for all methods when s is close to its boundaries 0 and 1. Due to its symmetric characteristic, the bi-weight kernel assigns equal weights to observations (or, in the case of scedasis estimation, to threshold exceedances) in the left and right neighbourhood of $s \times n$. As s comes closer to 1, the right neighbourhood of $s \times n$ falls beyond the range of the data sample. The same happens to the neighbourhood to the left of $s \times n$ when s is close to zero, causing the estimator to have fewer observations near the sample boundaries. This phenomenon is often called "boundary bias" and is in treated in, among others, Jones (1993) and Marron and Ruppert (1994). For the remainder of this paper it is not of great importance to correct for this bias when estimating the in-sample scedasis estimates, as they are solely used for mutual comparison among a set of loss-return variables. Moreover, the multivariate tests on the scedases are based on estimates of the integrated scedasis

function (Equation (2.11)), which does not require kernel density estimation. When predicting the risk measures, the out-of-sample prediction of the scedasis will be corrected to predict beyond the boundary point 1 without a bias, by implementing the boundary kernel from Equation (2.18).



(a) DGP 1

(b) DGP 2

(c) DGP 3

(d) DGP 4

Figure 1: Average in-sample scedasis estimates computed by each selection method compared to the true scedases for each DGP

For a more thorough comparison of the sample fraction size optimisation methods I computed

the estimation error variables $\hat{c}(s)/c(s) - 1$ over $s \in [0, 1]$, for all DGPs, that result from using each of the three methods. The means and variances mean squared errors (MSE) of these error variables over all 500×1500 estimations per DGP are shown in Table 1, along with the mean squared errors (MSEs). In terms of mean estimation accuracy, differences between the sample selection methods are minimal for DGPs 1, 2 and 3, with the Hall method being either the best or jointly the best performer in each of them. Contrarily, the Hall method is the worst performing method in DGP 4, possibly due to the relatively high amount of structural breaks. In terms of efficiency, the scedasis estimator has the lowest variance in all four DGPs, resulting in the lowest MSE for the Hall method in each case. This leads to the conclusion that, among the three examined methods, it performs best in estimating the in-sample scedasis.

Table 1: Mean, variance and mean squared error of the scedasis estimation error variables for each of the fraction size optimisation methods

DGP	1			2		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	-0.03	0.34	0.34	-0.03	0.48	0.48
Hall	-0.03	0.04	0.04	0.00	0.09	0.09
KS	-0.03	0.12	0.12	-0.02	0.21	0.21
DGP	3			4		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	-0.02	0.49	0.49	0.06	0.68	0.68
Hall	0.01	0.07	0.07	0.17	0.21	0.24
KS	-0.01	0.25	0.25	0.10	0.37	0.38

4.2 VaR prediction

To evaluate the performance of the three sample fraction selection methods on the VaR and ES, I used an estimation sample of 1000 observations to predict the risk measures in the first out-of-sample location for each of the 500 samples. This means that for each sample, the optimal value for k is selected over the first 1000 observations in order to compute a one-step ahead forecasts, which gives us a test sample of 500 observations total. In the same manner as the scedasis estimation error variables, I constructed the prediction error variables $\widehat{\text{VaR}}_{1-p,1000+i}/\text{VaR}_{1-p,1000+i} - 1$ for $i = 1$ and $p = 0.02$. The true VaR in the prediction error variable is the known 98th percentile of the distributions from which the simulated data sets are generated. The total means, variances

and MSEs over all 500 predictions are shown in the black coloured rows of Table 2 for each DGP and for the three threshold selection method. The difference in mean estimation error between the methods is larger than we previously saw when estimating the scedasis. The KS-Distance metric-method produces on average the most accurate estimates for DGPs 2, 3 and 4, whereas the Eye-ball-method is most accurate in DGP 1. The ability of the KS-method to predict a high quantile more accurately can be explained by the fact that it optimises k to match the most extreme events to a high theoretical quantile, whereas the other two address the tail index estimator directly.

The gray rows in Table 2 represent the means and variances of the CEVT prediction error variables. The CEVT next period quantile predictions are computed as $\widehat{\text{VaR}}^{\text{CEVT}}_{1-p,n+1} = X_{(n-k)}k^{\hat{\gamma}}(np)^{-\hat{\gamma}}$. For DGP 1, where extremes are generated homoscedastically, the CEVT predictors logically outperform their heteroscedasticity modelling counterparts. The true scedasis values at the first out-of-sample location are 1.20, 1.14 and 0.80 for DGPs 2, 3 and 4, respectively, and the HEVT models in these DGPs on average produce more or equally accurate predictions than the CEVT models, with one exception for the Eye-method in DGP 3. However, the CEVT methods are much more efficient due to the fact that the CEVT quantile predictions are not scaled by the scedasis, which suffers from a variance itself.

Table 2: Mean, variance and MSE of the VaR estimation error variables for each of the fraction size optimisation methods

DGP	1			2		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	-0.07	0.78	0.78	0.08	0.47	0.48
CEVT	0.02	0.02	0.02	-0.25	0.01	0.08
Hall	0.23	0.11	0.16	0.17	0.07	0.08
CEVT	0.05	0.02	0.02	-0.23	0.01	0.06
KS	0.10	0.25	0.26	0.07	0.15	0.16
CEVT	-0.01	0.01	0.01	-0.28	0.00	0.08
DGP	3			4		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	-0.14	0.62	0.64	-0.39	1.04	1.19
CEVT	-0.02	0.02	0.02	0.54	0.05	0.34
Hall	0.08	0.11	0.12	0.21	0.32	0.37
CEVT	-0.08	0.01	0.02	0.58	0.04	0.38
KS	0.06	0.25	0.25	-0.13	0.52	0.54
CEVT	-0.06	0.01	0.01	0.50	0.02	0.27

Regarding the HEVT models only, the KS method may be the most accurate on average, but

the Hall method is superior in all DGPs when it comes to efficiency. This results in the lowest MSE for the Hall method in all DGPs. Having a quantile estimator with a low variance can be crucially important when judging performance of the quantile predictor in terms of VaR violations. The importance of efficiency becomes apparent in Table 3, which shows per DGP the total amount of violations of the VaR predictions produced by each of the sample fraction selection methods, along with the test statistic T_3 of the binomial distribution backtest and its corresponding p-value in the brackets. The total amount of violations per DGP should ideally be around $500 \times 0.02 = 10$, and the Hall-method is by far the closest to this amount in all four situations even though Table 2 shows that the KS-distance method has a lower bias. The Hall-method is the only method that produces VaR predictions that are accepted at a 5% significance level. DGPs 1, 2 and 3 have respective p-values of 0.11, 0.75 and 0.11. The KS-method outperforms the Eye-method in terms of violation coverage as the former is more efficient, although it still produces too many violations to be considered an acceptable model.

To understand the importance of a low prediction error variance, we will have a look at how the high amount of violations come into being. As the VaR is a high quantile, observations are not symmetrically distributed around this quantile. Above a high quantile, there are simply fewer observations and these are more spread out, whereas below the quantile the amount and the frequency of occurred events increases. Therefore, the excess of violations caused by large underestimation of the VaR can not be offset by an overestimation of the same magnitude. Since a violations surplus or deficit of the same size are punished equally, this means that underestimation of the VaR disproportionately harms a correct coverage. Estimators with a high variance suffer increasingly more damage from underestimation than estimators with a low variance, even though their bias might be higher. To illustrate this, Figure 2 compares the two $\text{VaR}_{0.98}$ predictors produced by the KS-method and the Hall-method over the testing sample from DGP 3. The predictor from the KS-methods has a high variance (25%) and low (absolute) bias (6%) and the predictor from the Hall-method has low variance (11%) and high (absolute) bias (24%). The figure clearly shows that the downward spikes in the KS-method predictions are larger and cause a substantial part of the excess violations.

The difference in variance in Table 2 would argue in favour of modelling the VaR with homoscedastic extremes rather than heteroscedastic extremes. The CEVT methods have a very low variance since the VaR predictors are scaled by 1 over the whole sample, whereas the heteroscedastic predictors are scaled by an estimated scedasis value. Setting the performance of the Hall-method off

Table 3: Total violations with corresponding T_3-statistic and p-value of the $\text{VaR}_{0.98}$ estimates produced by each of the sample fraction selection methods

DGP	1		2		3		4	
	Violations	T_3 (p)	Violations	T_3 (p)	Violations	T_3 (p)	Violations	T_3 (p)
Eye	151	45.04 (0)	75	20.76 (0)	137	40.57 (0)	296	91.36 (0)
Hall	15	1.60 (0.11)	9	-0.32 (0.75)	15	1.60 (0.11)	38	8.94 (0)
KS	36	8.31 (0)	27	5.43 (0)	38	8.94 (0)	130	38.33 (0)

against the best performing CEVT method, which is tied between the Hall-method and KS-method with p-values of 0.52, 0.00, 0.20 and 0.20 for DGPs 1, 2, 3 and 4, respectively (see Appendix B.1, Table 13), confirms the relatively poor performance of heteroscedastic models. Only for DGP 2 we see an improvement when modelling heteroscedastic extremes, compared with CEVT. However, Figure 2 suggests that a large amount of the violations in Table 3 are caused by outliers in the predictions. Looking closer into these outlying underestimations of the VaR in the heteroscedastic models, we find the what causes most of the excess in violations in the HEVT models. These are zero predictions of the VaR, as a result of inadequate scedasis predictions. When the threshold $X_{(k)}$, for optimally chosen k, is not exceeded for a longer, recent period of time, the scedasis predictor in Equation (2.17) predicts a value of 0, even though Equation (2.9) implies that the scedasis is a strictly positive function. Consequently, the quantile predictor in Equation (2.16) predicts a value of 0, which seems an unrealistic value for a 98% quantile of a Fréchet distribution with location parameter 0. The fact that the Hall method tends to select the highest optimal values for k (the overall mean value of selected k is 19, 288 and 101 for the Eye-ball, Hall and KS- methods, respectively), such that the samples contain more threshold exceedances, explains it why it performs so much better than the other two methods in terms of correct coverage of violations. Inspection shows that in some situations, up to 80% of the total violations occur when the VaR is predicted to be 0.

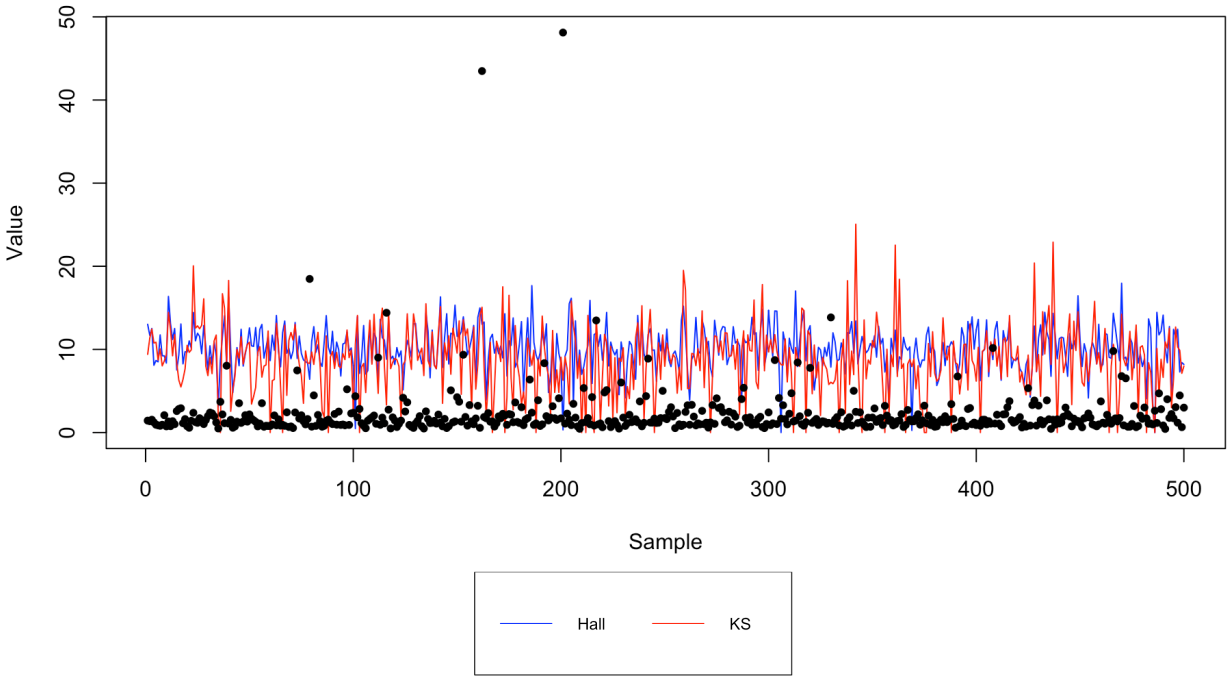


Figure 2: One-step-ahead $\text{VaR}_{0.98}$ predictions and actual simulated observations in DGP 3. The Hall-method has a low variance and high (absolute) bias and the KS-method has a high variance and low (absolute) bias.

4.3 ES prediction

Table 4 shows the means and variances of the $\text{ES}_{0.98}$ prediction error variables for both the heteroscedastic and homoscedastic models. To compute the true ES, I used the limit relation in Equation (2.21) for $\alpha = 2$. Again, the homoscedastic models outperform their heteroscedastic counterparts in DGP 1, but excluding the situations where the optimal k's are computed by the Hall-method in DGPs 2 and 3, the heteroscedastic methods produce on average more accurate predictions than the homoscedastic methods. Given the high variances the Eye-ball-method produces, it not very well suited to predict the ES. The best performing heteroscedastic method in terms of bias and variance is the KS-method, being most accurate and either most efficient or closely second most efficient in all DGPs. As a result, it has the lowest MSE in all DGPs. The efficiency of the predictors that result from the Hall- and KS-methods is a lot closer for the ES predictors than for the VaR predictors, which indicates that the KS-method estimates the tail index in Equation (2.20) more efficiently than the Hall-method does.

Table 4: Mean, variance and MSE of the ES estimation error variables for each of the fraction size optimisation methods

DGP	1			2		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	0.15	1.75	1.77	0.13	7.49	7.51
CEVT	0.15	1.28	1.30	-0.35	6.56	6.66
Hall	0.34	0.28	0.40	0.27	0.25	0.33
CEVT	0.13	0.12	0.14	-0.17	0.08	0.11
KS	0.14	0.34	0.36	0.10	0.21	0.22
CEVT	0.01	0.05	0.05	-0.26	0.02	0.09
DGP	3			4		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	-0.07	0.97	0.98	-0.24	4.73	4.79
CEVT	0.07	0.22	0.22	0.69	4.41	4.89
Hall	0.33	0.28	0.39	0.38	0.73	0.88
CEVT	0.06	0.10	0.10	0.75	0.39	0.95
KS	0.08	0.31	0.32	-0.07	0.65	0.66
CEVT	-0.09	0.03	0.04	0.57	0.12	0.44

4.4 Increased bandwidth

The causes of the poor performance in terms of violations found in Section 4.2 suggest a simple possible improvement of the HEVT VaR models. Increasing the bandwidth in the kernel function is known to increase the bias of the estimator, as it will fit the model less closely to the estimation sample, but it decreases the variance as the model becomes more smooth. Therefore, increasing the bandwidth in the boundary kernel function in Equation (2.17) inherently increases the bias and reduces the variance of the VaR predictor, a trade-off for which our findings show it may be highly beneficial when judging the VaR on correct coverage of violations. Moreover, increasing the bandwidth will cause the kernel to cover a wider range of observations, which will increase the probability of including an extreme event in the scedasis predictions. This may reduce the amount of 0-predictions in the HEVT VaR models which in turn will reduce its bias.

I revised the risk measure predictions done in Sections 4.2 and 4.3 with a bandwidth of $h = 0.2$ in the kernel density estimator in Equation (2.17). Table 5 shows the mean and variance of the resulting $\text{VaR}_{0.98}$ prediction error variables, from which we can see that the Eye-Ball-method benefits the most from the increase in bandwidth. As the Eye-Ball-method tends to choose a lower value for optimal fraction size k, the improvement in bias from reduced 0-predictions is substantially

larger than the setback in bias due to overfitting. As a result, the bias is improved in every DGP. Furthermore, its variance is clearly smaller in all DGPs. Overall, there is no visible improvement in bias for the Hall and KS-Distance methods, nor is there an overall deterioration. However, the variance is smaller for all methods in all DGPs, for which we have seen it can greatly improve the predictions of the VaR and ES.

Table 5: Mean, variance and MSE of the VaR estimation error variables for each of the fraction size optimisation methods with a bandwidth of 0.2

DGP	1			2		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	0.02	0.45	0.45	0.06	0.27	0.27
Hall	0.21	0.06	0.10	0.14	0.05	0.07
KS	0.10	0.13	0.14	0.06	0.09	0.10
DGP	3			4		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	-0.01	0.37	0.37	-0.18	0.67	0.70
Hall	0.23	0.07	0.12	0.27	0.16	0.23
KS	0.09	0.14	0.15	-0.01	0.29	0.29

Table 6 shows per DGP and selection method the amount of amount of total violations of the one-step-ahead $\text{VaR}_{0.98}$ predictions over the test sample of size 500, using a bandwidth of 0.2. Comparing Table 6 with Table 3 (bandwidth 0.1) stresses the importance of a low variance. Except for DGP 2 when using the Hall-method, the amount of violations is clearly improved in all cases, leading to substantially lower absolute T_3-statistics. The Hall-method predictions are now accepted for all four DGPs with p-values of 0.75, 0.20, 0.75 and 0.75. However, the one exception reveals the downside of increasing the bandwidth: over-smoothing. The Hall-method applied to DGP 2 is a very well performing model with a bandwidth of 0.1, having a p-value of 0.75, which means the VaR predictions are all relatively close to the true value. The best predictions suffer most from the increase in bias. Therefore, increasing the bandwidth may worsen the best predictions, but it can greatly improve the worst predictions. This brings us to the time honored bias-variance trade-off puzzle. The Hall-method still outperforms the other two methods when the bandwidth is increased to 0.2, and even beats the best performing CEVT method in all DGPs.

Table 7 shows the mean and variance of the one-step-ahead $\text{ES}_{0.98}$ prediction errors using a bandwidth of 0.2. As for the mean, some cases show a decrease while others are increased. Whereas there is no overall improvement in mean, the variance is substantially reduced in all cases. Overall,

the KS-Distance metric-method remains the best performing method for predicting the ES, as its MSE is lowest in all four DGPs.

Table 6: Total violations and corresponding T_3 test statistics and p-values of the $\text{VaR}_{0.98}$ estimates produced by each of the threshold selection methods and a bandwidth of 0.2

DGP	1		2		3		4	
	Violations	T_3 (p)	Violations	T_3 (p)	Violations	T_3 (p)	Violations	T_3 (p)
Eye	67	18.21 (0)	37	8.62 (0)	61	16.29 (0)	136	40.25 (0)
Hall	9	-0.32 (0.75)	6	-1.28 (0.20)	11	0.32 (0.75)	11	0.32 (0.75)
KS	22	3.83 (0)	19	2.87 (0)	24	4.47 (0)	59	15.65 (0)

Table 7: Mean, variance and MSE of the ES estimation error variables for each of the fraction size optimisation methods with a bandwidth of 0.2

DGP	1			2		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	0.12	1.32	1.34	0.62	6.92	6.96
Hall	0.31	0.22	0.32	0.23	0.21	0.27
KS	0.13	0.20	0.22	0.09	0.14	0.15
DGP	3			4		
	Mean	Variance	MSE	Mean	Variance	MSE
Eye	0.08	0.71	0.72	-0.04	2.79	2.79
Hall	0.33	0.23	0.34	0.43	0.53	0.71
KS	0.11	0.20	0.21	0.05	0.39	0.39

5 Empirical application

In this section, I analyse differences in the scedasis over time and in the cross section on a real life multivariate data set. Furthermore, I exploit the results from the previous section and examine whether they can be applied to improve the prediction of the marginal risk measures.

5.1 Data

For an empirical application of modelling for heteroscedastic extremes in financial data I use the daily loss-return time series of 10 selected Industry Portfolios from the Fama-French website, over the time period from January 4 1988 until October 31 2022. Each stock listed on the NYSE, AMEX or NASDAQ is assigned to an industry portfolio at the end of June of each year based on its four-digit standard industrial classification (SIC) code at that time. In order to save space in the text and tables, I will number the portfolios from 1 to 10. Please find the industry names corresponding to each portfolio number in Appendix B.1, Table 14.

To justify analysing heteroscedasticity in extremes, this data set must satisfy the set of assumptions on multivariate data mentioned in Einmahl et al. (2022). The first assumption I will deal with is the one on serial independence. Like most financial data sets, this raw data set exhibits serial dependence. The extremal index is estimated to have a value of 3, meaning extreme events occur in clusters of, on average, 3 consecutive days. In line with Einmahl et al. (2022), I will clean the serially dependent data set by selecting the largest daily loss among the 10 portfolios, and rank them from high to low. Then, all the observations of the day with the highest daily loss are saved. The observations on the day with the second highest loss are saved when it is not within two days of the previously saved observations. If it is within two days of the previously saved observations, all observations on that corresponding day are discarded. Subsequently, the next daily maximum losses are monitored in descending order. In this manner, all observations are discarded if they fall within two days of any of the previously saved observations, or will otherwise be saved as well. This method of data cleaning allows us to obtain serial independence in the tails of the distribution, while keeping as much information as possible from the most extreme events.

Now that the desired serial independence is obtained, I will test whether the tail index for each portfolio remains constant over the whole time period. Computing the test statistics from Equation (2.28) over the observations from January 4 1988 until October 31 2022 leads to 7 rejections of a constant tail index null hypothesis out of the 10 portfolios. Einmahl et al. (2016) suggest that the

financial crisis in 2008 might have caused a structural change in the tail index value in the daily loss-returns of the S&P 500. Given the fact that the industry portfolios and the S&P 500 are largely intertwined (the portfolios and the index all exist of US stocks) it is very plausible that the financial crisis also is the reason that a constant tail index is rejected for many of the portfolios. Redoing the tests over the time period from January 4 1988 until December 31 2007 for each portfolio gives the p-values given in Table 8. We fail to reject that the tail indices are constant at a 5% significance level for all portfolios, albeit only just for portfolio 2 with a p-value of 0.09. The other 9 portfolios reject the null hypothesis more comfortably. Whereas the original data set contains 8778 daily observations, the workable data now exists of 1380 observations for each portfolio.

Table 8: The p-values that result from testing for a time invariant tail index from January 4 1988 - December 31 2007

Portfolio	1	2	3	4	5	6	7	8	9	10
p-value	0.73	0.09	0.85	0.90	0.24	0.43	0.17	0.69	0.38	0.48

From Section 2.3.3 we know that the test for equality of two tail indices may be rejected because of two reasons: the estimated tail indices differ too much and/or a strong dependence between the marginal tails. Table 9 shows the Hill-estimated tail indices for each of the 10 portfolios by selecting optimal sample fraction k by means of the Hall-method. The largest difference in tail index estimates is between Portfolios 4 and 6, with values of 2.73 and 3.78, respectively. Computing the test statistic from Equation (2.30) gives a value of 0.94, with a corresponding p-value of 0.33. This means we fail to reject the null hypothesis of equal tail indices for any of the commonly used significance levels. The portfolio pair with the lowest probability of equality of the tail indices is between Portfolios 4 and 7, which has a p-value of 0.18. This means that Portfolios 4 and 7 must have a stronger tail dependence than Portfolios 4 and 6. The p-values for the equality tests for each pair of portfolios are displayed in Appendix B.2, Table 15.

Table 9: Tail index estimates and optimal sample fraction sizes for 10 industry portfolios

Portfolio	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}$	2.95	2.94	2.98	2.73	3.20	3.78	3.51	2.77	3.00	3.02
k	46	123	127	96	20	6	20	11	40	31

5.2 Multivariate analysis

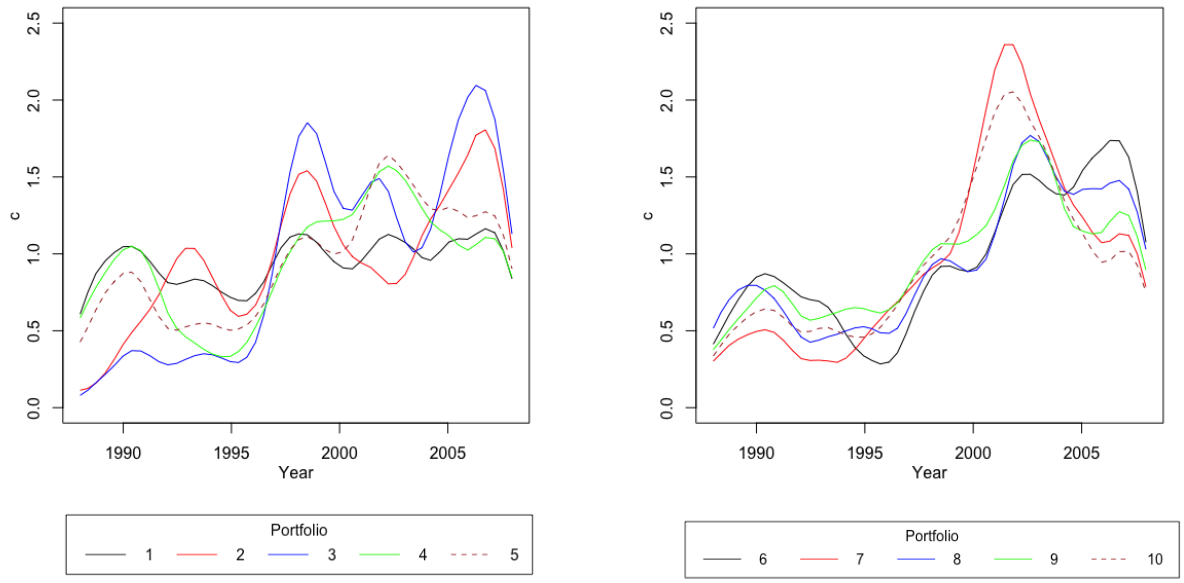
As the data set satisfies the assumptions necessary, we may now conduct the hypothesis tests of testing for homoscedastic extremes against heteroscedastic extremes and testing for variation in the scedasis functions between the different portfolios. For the former test, I computed the test statistic values from Equation (2.13) for each portfolio. Table 10 shows the values for these test statistics and their corresponding p-values. With p-values of basically 0 for each portfolio, we may strongly reject that the extremes are homoscedastic. This justifies the implementation of HEVT risk measures. For the latter test, I computed the test statistic value from Equation (2.12), which resulted in a value of 0.95 and a corresponding p-value of 0. This means we may strongly reject that scedases are equal for each portfolio.

Table 10: Values of the test statistics for a constant scedasis over time and their corresponding p-values

Portfolio	1	2	3	4	5	6	7	8	9	10
T_2	47	144	110	67	62	65	113	65	75	85
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The results of both test become evident in Figure 3, which contains plots of the 10 in-sample scedasis estimates computed by the estimator in Equation (2.8), over the whole time period from 1988-2007. Large variations of the scedasis values are visible both over time and between portfolios. Figure 3 also allows us to examine the different reactions of the portfolios to macro economic situations in respect of frequency of occurring extreme events. We see a sharp rise in scedasis values during and after the burst of the Internet Bubble in early 21st century for Portfolios 4 through 10. Portfolios 1, 2 and 3 seem to have been more resistant to this stock market crash. Portfolio 1 corresponds to the Food Industry portfolio, therefore showing little reaction to the burst of the bubble as people will need food regardless of the economic situation. The scedasis for this industry portfolio shows the least variation over the whole time period out of all 10 portfolios. Portfolios 2 and 3 correspond to the Mining and Oil Industry portfolios, respectively, and show a peak in scedasis value in the late 1990's, whereas the other 8 portfolios barely show any reaction during that time. An economic crisis struck Russia in 1998, a major player in the oil industry, causing oil prices to take a deep dive. The Russian economic crisis was caused by economic crises in South East Asian countries and it consequently put Latin American countries in a deep recession. South East Asia, Russia and Latin America make up for a large part of mining activity world wide (Garside,

2021), which explains the drastically increased frequency of extreme losses in mining stocks. The sudden drop in scedasis values towards the end of the sample shown by all 10 portfolios will be ignored as this is likely caused by the boundary bias in the kernel estimator.



(a) Portfolios 1 through 5

(b) Portfolios 6 through 10

Figure 3: In sample scedasis estimates for the 10 portfolios

5.3 Risk measures

To test the performance of the HEVT VaR models against the CEVT VaR models using empirical data I use rolling estimation samples of 1000 observations to compute 380 one-day-ahead $\text{VaR}_{0.98}$ forecasts per portfolio. The forecasts cover the time period between June 19 2002 and December 31 2007, over which the models will be backtested. In all cases, the Hall method is used to compute the optimal sample fraction size. The amount of violations per portfolio testsample should be around $380 \times 0.02 = 7.6$, given a total ideal amount of 76. Table 11 shows the amount of violations, the value of the T_3-test statistic (Equation (2.19)) and its corresponding p-value for the HEVT model with bandwidths 0.1 and 0.2, and for the CEVT model. Overall, the HEVT model with a bandwidth of 0.2 has the best violations coverage with 53 violations and corresponding T_3-statistic and p-value of -2.67 and 0.01, respectively. The model is accepted at a 5% significance level for 7 out of 10 portfolios, whereas the model with bandwidth 0.1 and the CEVT model are both accepted for 6 portfolios. In total, the CEVT model slightly outperforms the HEVT model with small bandwidth,

with respective T_3-statistics of 6.14 and 7.18. Appendix B.2, Figure 4 shows for each portfolio plots of the observations over the test sample along with the $\text{VaR}_{0.98}$ values predicted by the two HEVT models and the CEVT model.

When looking closer at the differences between the HEVT models with smaller and larger bandwidths, we come to the same findings as we did after running the simulations. The three worst performing portfolios under the smaller bandwidth, Portfolios 3, 7 and 10, make up for 70% of the total amount of violations. These Portfolios are the only portfolios with 0-predictions over the test sample (see Figure 4), which explains the high amount of violations. We see a great improvement when increasing the bandwidth for Portfolios 3, 7 and 10, showing p-values of practically 0 for a bandwidth of 0.1 and respective p-values of 0.05, 0.34 and 0.83 for a bandwidth of 0.2. The fourth worst performing sample, Portfolio 4, is slightly improved with a p-value increasing from 0.02 to 0.04. However, the three best performing samples under the small bandwidth, for Portfolios 1, 5 and 6, show a deterioration in performance when the bandwidth is increased. The three moderately performing samples, Portfolios 2, 8 and 9, show an improvement in only 1 portfolio when the bandwidth is increased. This reaffirms that increasing the bandwidth can greatly improve the worst predictions, at a cost of reducing the accuracy of the best predictions.

Table 11: Violations counted over the 10 test samples of 380 observations each for HEVT models with bandwidths 0.1 and 0.2, along with their corresponding T_3-statistic and p-value

	Violations			T_3			p-value		
	HEVT		CEVT	HEVT		CEVT	HEVT		CEVT
Bandwidth	0.1	0.2		0.1	0.2		0.1	0.2	
1	9	6	11	0.51	-0.59	1.61	0.61	0.56	0.11
2	11	10	25	1.25	0.88	6.38	0.21	0.38	0.00
3	37	13	20	10.77	1.99	4.54	0.00	0.05	0.00
4	1	2	12	-2.42	-2.05	1.61	0.02	0.04	0.11
5	6	3	11	-0.59	-1.69	1.25	0.56	0.09	0.21
6	6	4	18	-0.59	-1.32	3.81	0.56	0.19	0.00
7	28	5	3	7.47	-0.95	-1.69	0.00	0.34	0.09
8	4	2	14	-1.32	-2.05	2.35	0.19	0.04	0.02
9	4	2	11	-1.32	-2.05	1.25	0.19	0.04	0.21
10	32	7	3	8.94	-0.22	-1.69	0.00	0.83	0.09
Total	138	53	129	7.18	-2.67	6.14	0.00	0.01	0.00

For the one-step ahead forecasts of the expected shortfall I will stick with the Hall-method to select the optimal sample fraction size, even though Section 4.3 points out that the KS-Distance-method produces the most accurate and most efficient predictions. The reason is that the KS-Distance-method yet produces more 0-predictions of the VaR and, hence, the ES, than the Hall-method does. This makes the ES backtest statistic in Equation (2.25) useless as dividing by 0 is impossible. The values of the test statistics are displayed in Table 12 for HEVT models with bandwidths 0.1 and 0.2 and the CEVT model. The three portfolios that include 0-predictions with a bandwidth of 0.1 have infinite test statistic values and are the only situations in which the null hypothesis is rejected. As the test is one-sided, the HEVT model with the large bandwidth has the strongest arguments of accepting an adequately modelled ES even though the model with small bandwidth has values closer to 0 in most cases. We may therefore conclude that the model with bandwidth 0.1 produces the most accurate ES estimates when no 0-predictions occur, whereas the model with bandwidth 0.2 and the CEVT model tend to overestimate the risk measure relatively more.

Table 12: Test statistic values of testing for correctly modelling the ES against underestimation of the ES for the HEVT models with small and large bandwidth and the CEVT model and the simulated critical values for each portfolio

	T_4			Critical Value
	HEVT		CEVT	
Bandwidth	0.1	0.2		
1	-0.09	-0.16	-0.10	0.12
2	-0.01	-0.02	-0.04	0.18
3	∞	-0.08	-0.02	0.06
4	0.08	-0.20	-0.12	0.15
5	-0.20	-0.22	-0.17	0.12
6	-0.07	-0.12	-0.11	0.16
7	∞	-0.18	0.01	0.08
8	-0.24	-0.27	-0.37	0.21
9	-0.22	-0.27	-0.17	0.12
10	∞	-0.30	-0.06	0.23
Total	∞	-0.15	-0.07	0.21

6 Conclusion

During this research, I investigated modelling heteroscedastic tail distributions in a multivariate financial data set. Specifically, heteroscedasticity is modelled in the case that its presence is in the form of variation in the frequency of which extreme events occur, rather than variation in the magnitude of the extreme events. I use methods from Einmahl et al. (2016) and Einmahl et al. (2022), who propose to model such heteroscedasticity by means of the scedasis function, a function that scales the tail distributions and represents the variation over time.

The estimator of the scedasis relies on a threshold, for which exceedances are considered extreme events that are counted to estimate the time specific frequency. This threshold can be equivalently used for tail index estimation and Value-at-Risk prediction and there exist several methods of selecting the optimal threshold. In search for the optimal procedure to estimate the scedasis and predict the risk measures VaR and expected shortfall, I tested and compared three methods of selecting the optimal threshold through running Monte Carlo simulations. I considered the Hall-method (Hall, 1990), the Eye-Ball-method (Danielsson et al., 2019) and the KS-Distance-method (Danielsson et al., 2019) and the results of the simulations show that the Hall-method is overall the most useful. For scedasis estimation it is most accurate and efficient. For VaR prediction it is most efficient, which is shown to be an important feature. For ES prediction it was neither most accurate nor most efficient, but it produced predictions that enabled backtestability of the model. Furthermore, I showed that prediction of the one-step ahead VaR and ES can be improved by modelling for heteroscedastic extremes, compared to modelling for homoscedastic extremes.

To justify the use of a scedasis function to model potential heteroscedsticity in extremes of a data set, one must test whether it satisfies a set of assumptions. I showed that these assumptions reasonably hold for a data set containing the daily loss observations of 10 Fama-French Industry portfolios. The scedasis function is estimated and used to examine the differences in heteroscedasticity in the cross section and over time. At last, the scedasis function is used to compute one-step ahead VaR and ES forecasts of the daily losses. Backtests in this practical application confirm that risk measure prediction can be improved by modelling extremes heteroscedastically, rather than homoscedastically.

The amount of threshold selection methods compared in this research is limited to three. This does not cover all existing methods in the literature. In fact, there even exist innovations of the Hall-method by Drees and Kaufmann (1998) and Danielsson et al. (2000). I did not consider them for

this paper due to their exceedingly long computation times, but it could be interesting to examine their performance in scedasis estimation and heteroscedastic risk measure prediction. Furthermore, this research shows the importance of appropriately choosing a bandwidth size when predicting the scedasis by means of a kernel density function. The results of this research can possibly be improved after investigating how to optimally select a value for the bandwidth. Another way to possibly improve scedasis estimation or prediction is by examining the outcomes of different kernel estimation methods. For this research I subjectively chose to use the bi-weight kernel function, however, there exists a wide variety of kernel estimators that are worth considering.

Appendices

A Derivation of the scedasis functions

DGP 1:

We have

$$F_{n,i}(x) = \exp(x^{-2})$$

and

$$1 - F_{n,i}(x) \sim x^2.$$

If we take $1 - F_0(x) = \frac{1}{x^2}$, then

$$\frac{1 - F_{n,i}(x)}{1 - F_0(x)} \sim 1.$$

Since $\int_0^1 1ds = 1$, the scedasis function is defined as

$$c(s) = 1.$$

DGP 2:

We have

$$F_{n,i}(x) = \exp\left(-\Big(\frac{x}{0.5 + i/n}\Big)^{-2}\right)$$

and

$$1 - F_{n,i}(x) \sim \Big(\frac{0.5 + i/n}{x}\Big)^2.$$

Let $g(s) = (0.5 + s)^2$. For $g(s)$ to be considered a scedasis function it must hold that $\int_0^1 g(s)ds = 1$ (identification condition 2.3). However,

$$\int_0^1 (0.5 + s)^2 ds = \frac{13}{12}.$$

If we take $1 - F_0(x) = \frac{13}{12}\frac{1}{x^2}$, we have

$$\frac{1 - F_{n,i}(x)}{1 - F_0(x)} \sim \frac{12}{13}\Big(0.5 + \frac{i}{n}\Big)^2.$$

So the scedasis is defined as

$$c(s) = \frac{12}{13}(0.5 + s)^2,$$

with $\int_0^1 c(s) = 1$.

DGP 3:

We have

$$F_{n,i}(x) = \begin{cases} \exp\left(-\left(\frac{x}{0.5+2*i/n}\right)^{-2}\right), & \text{if } 0 < i \leq n/2 \\ \exp\left(-\left(\frac{x}{2.5-2*i/n}\right)^{-2}\right), & \text{if } n/2 < i \leq n \end{cases}$$

and

$$1 - F_{n,i}(x) \sim \begin{cases} \left(\frac{0.5+2*i/n}{x}\right)^2, & \text{if } 0 < i \leq n/2 \\ \left(\frac{2.5-2*i/n}{x}\right)^2, & \text{if } n/2 < i \leq n. \end{cases}$$

Let

$$g(s) = \begin{cases} (0.5+2s)^2, & \text{if } 0 \leq s \leq 1 \\ (2.5-2s)^2, & \text{if } 0.5 < s \leq 1, \end{cases}$$

such that

$$\int_0^1 g(s)ds = \int_0^{0.5} (0.5+2s)^2 ds + \int_{0.5}^1 (2.5-2s)^2 ds = \frac{13}{24} + \frac{13}{24} = \frac{13}{12}.$$

If we take $1 - F_0(x) = \frac{26}{24}\frac{1}{x^2}$, we have

$$\frac{1 - F_{n,i}(x)}{1 - F_0(x)} \sim \begin{cases} \frac{12}{13}(0.5+2\frac{i}{n})^2, & \text{if } 0 < i \leq n/2 \\ \frac{12}{13}(2.5-2\frac{i}{n})^2, & \text{if } n/2 < i \leq n. \end{cases}$$

So the scedasis function is defined as

$$c(s) = \begin{cases} \frac{12}{13}(0.5+2s)^2, & \text{if } 0 \leq s \leq 0.5 \\ \frac{12}{13}(2.5-2s)^2, & \text{if } 0.5 < s \leq 1, \end{cases}$$

with $\int_0^1 c(s) = 1$.

DGP 4:

We have

$$F_{n,i}(x) = \begin{cases} \exp\left(-\left(\frac{x}{0.8}\right)^{-2}\right), & \text{if } 0 \le i \le 0.4n \vee 0.6n \le i \le n \\ \exp\left(-\left(\frac{x}{-7.2+20*i/n}\right)^{-2}\right), & \text{if } 0.4n < i \le 0.5n \\ \exp\left(-\left(\frac{x}{12.8-20*i/n}\right)^{-2}\right), & \text{if } 0.5n < i < 0.6n \end{cases}$$

and

$$1 - F_{n,i}(x) \sim \begin{cases} \left(\frac{0.8}{x}\right)^2, & \text{if } 0 \le i \le 0.4n \vee 0.6n \le i \le n \\ \left(\frac{-7.2+20*i/n}{x}\right)^2, & \text{if } 0.4n < i \le 0.5n \\ \left(\frac{12.8-20*i/n}{x}\right)^2, & \text{if } 0.5n < i < 0.6n. \end{cases}$$

Let

$$g(s) = \begin{cases} 0.8^2, & \text{if } 0 \le s \le 0.4 \vee 0.6 \le s \le 1 \\ (-7.2+20s)^2, & \text{if } 0.4 < s \le 0.5 \\ (12.8-20s)^2, & \text{if } 0.5 < s < 0.6s, \end{cases}$$

such that

$$\begin{aligned} \int_0^1 g(s)ds = \int_0^{0.4} 0.8^2 ds + \int_{0.4}^{0.5} (-7.2+20s)^2 ds + \int_{0.5}^{0.6} (12.8-20s)^2 ds + \int_{0.6}^1 0.8^2 ds \\ = \frac{32}{125} + \frac{134}{375} + \frac{134}{375} + \frac{32}{125} = \frac{92}{75}. \end{aligned}$$

If we take $1 - F_0(x) = \frac{92}{75}\frac{1}{x^2}$, we have

$$\frac{1 - F_{n,i}(x)}{1 - F_0(x)} \sim \begin{cases} \frac{92}{75}0.8^2, & \text{if } 0 \le i \le 0.4n \vee 0.6n \le i \le n \\ \frac{92}{75}(-7.2+20\frac{i}{n})^2, & \text{if } 0.4n < i \le 0.5n \\ \frac{92}{75}(12.8-20\frac{i}{n})^2, & \text{if } 0.5n < i < 0.6n. \end{cases}$$

So the scedasis function is defined as

$$c(s) = \begin{cases} \frac{75}{92}0.8^2, & \text{if } 0 \le s \le 0.4 \vee 0.6 \le s \le 1 \\ \frac{75}{92}(-7.2+20s)^2, & \text{if } 0.4 < s \le 0.5 \\ \frac{75}{92}(12.8-20s)^2, & \text{if } 0.5 < s < 0.6s, \end{cases}$$

with $\int_0^1 c(s) = 1$.

B Figures and Tables

B.1 Simulations

Table 13: Total violations of the $\text{VaR}_{0.98}$ predictions produced by each of the sample fraction selection methods by CEVT

DGP	1		2		3		4	
	Violations	z (p)	Violations	z (p)	Violations	z (p)	Violations	z (p)
Eye	10	1.00 (1.00)	19	2.87 (0)	15	1.60 (0.11)	5	-1.60 (0.11)
Hall	12	0.64 (0.52)	19	2.87 (0)	14	1.28 (0.20)	6	-1.28 (0.20)
KS	12	0.64 (0.52)	19	2.87 (0)	14	1.28 (0.20)	6	-1.28 (0.20)

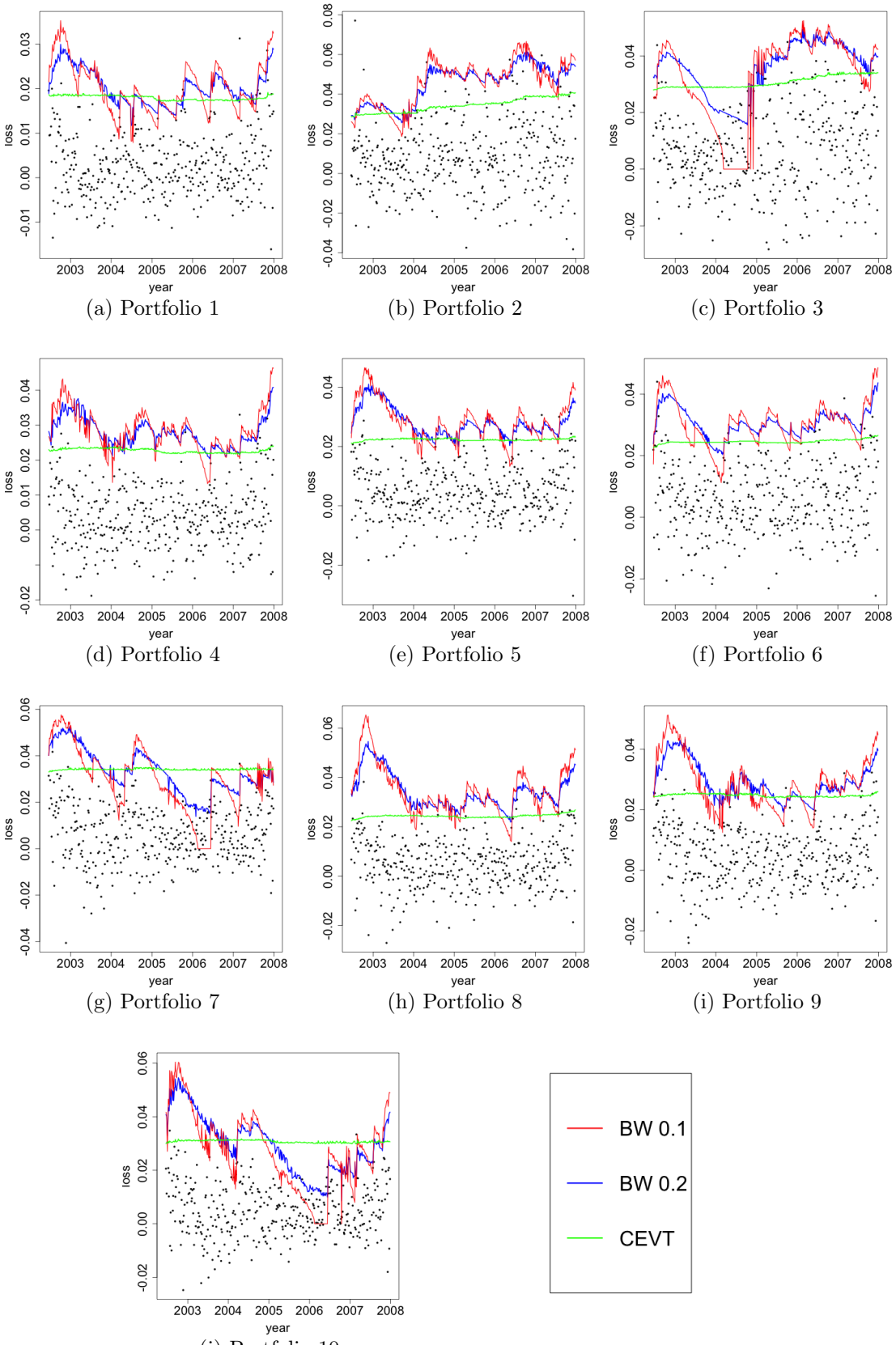
B.2 Empirical application

Table 14: Industry names corresponding to the portfolio numbers

Portfolio Number	Industry Name
1	Food
2	Mines
3	Oil
4	Clothes
5	Consumer Durables
6	Fabricated Products
7	Machinery & Business Equipment
8	Transportation
9	Retail Stores
10	Other

Table 15: p-values of the equal tail indices test statistics for each pair of portfolios

Portfolio	2	3	4	5	6	7	8	9	10
1	0.97	0.95	0.52	0.66	0.47	0.40	0.82	0.90	0.90
2	-	0.87	0.53	0.67	0.47	0.37	0.85	0.89	0.88
3	-	-	0.42	0.72	0.50	0.40	0.81	0.96	0.94
4	-	-	-	0.40	0.33	0.18	0.96	0.47	0.52
5	-	-	-	-	0.64	0.67	0.58	0.67	0.74
6	-	-	-	-	-	0.81	0.27	0.57	0.36
7	-	-	-	-	-	-	0.47	0.38	0.29
8	-	-	-	-	-	-	-	0.74	0.73
9	-	-	-	-	-	-	-	-	0.97



(a) Portfolio 1 (b) Portfolio 2 (c) Portfolio 3

(d) Portfolio 4 (e) Portfolio 5 (f) Portfolio 6

(g) Portfolio 7 (h) Portfolio 8 (i) Portfolio 9

(j) Portfolio 10

Figure 4: Loss observations and $\text{VaR}_{0.98}$ forecasts over each portfolio test sample by different bandwidth sizes and CEVT

References

Acerbi, C., & Szekely, B. (2014). Back-testing expected shortfall. *Risk*, *27*, 76–81.

Ancona-Navarrete, M. A., & Tawn, J. A. (2000). A comparison of methods for estimating the extremal index. *Extremes*, *3*(1), 5–38.

Berghaus, B., & Bücher, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *The Annals of Statistics*, *46*(5), 2307–2335.

Bickel, P. J., & Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 967–985.

Cont, R. (2007). Volatility clustering in financial markets: Empirical facts and agent-based models. In G. Teyssière & A. P. Kirman (Eds.), *Long memory in economics* (pp. 289–309).

Danielsson, J., De Haan, L., Peng, L., & De Vries, C. (2000). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, *76*, 226–248.

Danielsson, J., Ergun, L., De Haan, L., & De Vries, C. (2019). Tail index estimation: Quantile-driven threshold selection. *Working Paper*.

De Haan, L., Tank, A., & Neves, C. (2015). On tail trend detection: Modeling relative risk. *Extremes*, 141–178.

Drees, H., & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, *75*, 149–172.

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

Einmahl, J., Ferreira, A., De Haan, L., Neves, C., & Zhou, C. (2022). Spatial dependence and space-time trend in extreme events. *Annals of Statistics*, *50*, 30–52.

Fernandez, V. (2005). Risk management under extreme events. *International Review of Financial Analysis*, *14*, 113–148.

Garside, M. (2021). Mining land area share worldwide by country 2020. https://www.statista.com/statistics/1255879/distribution-of-mining-land-area-worldwide-by-country/

Granger, C. W. J., & Ding, Z. (1995). Some properties of absolute return: An alternative measure of risk. *Annales d'Économie et de Statistique*, (40), 67–91.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, *32*, 177–203.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annuals of Statistics*, *3*, 1163–1174.

Hyung, N., & De Vries, C. (2007). Portfolio selection with heavy tails. *Journal of Empirical Finance*, *14*, 383–400.

Jacobsen, B., & Dannenburg, D. (2003). Volatility clustering in monthly stock returns. *Journal of Empirical Finance*, *10*(4), 479–503.

Jansen, D., Koedijk, K., & De Vries, C. (2000). Portfolio selection with limited down-side risk. *Journal of Empirical Finance*, *7*, 247–269.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jones, M. C., & Foster, P. J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, *6*(4), 1005–1013.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and computing*, *3*(3), 135–146.

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

Leadbetter, M. R., & Nandagopalan, S. (1989). On exceedance point processes for stationary sequences under mild oscillation restrictions. *Extreme Value Theory*, 69–80.

Macaraeg, D. (2021). *Selecting optimal sample fraction for tail index estimation in a safety first portfolio problem* (Bachelor's thesis). Erasmus University Rotterdam.

Marron, J. S., & Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(4), 653–671.

Mason, D. (1982). Laws of large numbers for sums of extreme values. *Institute of Mathematical Statistics*, *10*, 754–764.

Moloney, N. R., Faranda, D., & Sato, Y. (2019). An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(2), 022101.

Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of financial economics*, *22*(1), 27–59.

Robert, C. Y., Segers, J., & Ferro, C. A. T. (2008). A sliding blocks estimator for the extremal index.

Shirazi, A. (2014). Value at risk (var) backtesting techniques and p-value risk decomposition analysis. *Available at SSRN 2413702*.

Smith, R. L., & Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(3), 515–528.

Stupfler, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes*, *22*.

Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. *Behavioral decision making* (pp. 25–41). Springer.