ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS QUANTITATIVE FINANCE

Driving forces of Scedasis of negated stock returns in the cross-section and over time

Name student: Job Hölscher

Student ID number: 449506jh

Supervisor: prof. dr. C. Zhou

Second assessor: dr. M Khismatullina

Date: 24-03-2023

Abstract

Einmahl et al. (2016) propose a univariate extreme value statistic called 'Scedasis', which can be easily interpreted as the frequency of extremes. In Einmahl et al. (2022), they investigate the multivariate case and apply their research on a rainfall dataset. In this paper, we apply their research on a financial dataset and explore two Fama-French factors (Fama and French, 1993) as possible driving factors of Scedasis in the financial setting. For the latter, we use a panel regression. We find Scedasis of negated stock returns to be invariant in the cross-section of stocks but varying over time. In the panel regression, we discover Book-To-Market ratio to be a predictive factor of Scedasis in stock returns.

Contents

1 Introduction

Extreme values are arguably the most important feature of time series to understand and model or even predict. This modelling could for example decide the minimum height of dikes due to extreme flooding (Papalexiou and Montanari, 2019), the strength of tall buildings (Laogan and Elnashai, 1999) and, in a financial environment, the potential losses of an investment portfolio (Olsen, 1997).

Specifically, in financial analysis, extremes in time series such as stock returns are modelled for several reasons. One of these reasons is risk mitigation for investors, by means of Value At Risk (Duffie and Pan, 1997) and Extreme Downside Risk (Huang et al., 2012) amongst many other statistical measures. Another reason is its indication of the state of the economy; a high frequency of extreme stock returns indicates high volatility which in turn is inherent to economic downturn, for example during the Great Recession (Rabbani et al., 2017). Moreover, Black (1976) report evidence that suggests a negative shock in stock returns generates more volatility than a positive shock of equal magnitude. Hence, especially negative extreme stock returns are important to model.

Due to the mitigation of risk for investors and the impact negative shocks have on volatility, we study the frequency of negative extremes in stock returns. Einmahl et al. (2016) propose an extreme value statistic called 'Scedasis', which can be interpreted as a frequency of extremes. They propose the multivariate case in Einmahl et al. (2022) with an application on rainfall data. In this paper, we apply the multivariate analysis on a cross-section of negated stock returns. Fama and French (1995) find the Book-To-Market ratio and Size factors to capture a major part of the cross-section of stock returns. If we find variance of Scedasis over time and in the cross-section of stock returns, we aim to discover a possible driving force through these two factors using a panel regression. The research question is: do Size and Book-To-Market ratio drive Scedasis of negated stock returns in the cross section and over time? Due to the findings in Fama and French (1995), we expect both Size and Book-To-Market ratio to affect the Scedasis of negated stock returns.

We will first summarize the analysis of Einmahl et al. (2016) and its extension to the multivariate case in Einmahl et al. (2022). The latter introduces two tests on Scedasis: one for variance over time and one for difference in the cross-section. If we find a varying Scedasis in at least one dimension, over time or in the cross-section, we apply a panel regression of yearly Scedases on yearly values of Size and Book-To-Market ratio. Two dataset will be needed for the full analysis: the Scedasis analysis of Einmahl et al. (2022) will be applied on a dataset containing 20 years of daily negated returns of 50 stocks of the S&P 500. Using the daily negated returns, we construct yearly Scedasis values for each stock. In the panel regression, we utilize a second dataset, which contains yearly values of Book-To-Market ratios and Size along with the constructed yearly Scedasis series. For the application to be viable, we need to pre-process the daily multivariate stock returns such that it does not exhibit serial dependence and ensure the extreme value index (EVI), a well known extreme value statistic (Gnedenko, 1943), is constant for all considered stock return series.

In the application on negated stock returns we reject the hypothesis of equality of Scedasis over time. However, we do no reject equality of Scedasis in the cross-section. Using normalized yearly Scedasis values we find the demeaned yearly log difference of Book-To-Market ratio to be a significant driver of next year's Scedasis values. Hence, we find Book-To-Market ratio to be a predictive factor of yearly Scedasis.

We define the model for scedasis in Section 2.1 following Einmahl et al. (2022). We summarize estimators and test statistics for the scedasis analysis in Section 2.2. Furthermore, in the final subsection of Section 2.2, we propose the panel regression on yearly scedasis values. We then apply the research to the financial datasets, which we pre-process in Section 2.3. Section 3 summarizes both daily and yearly datasets. We present the results of the data pre-processing in Section 4.1. Consecutively, we present the results of the tests proposed by Einmahl et al. (2022) in the financial application in Section 4.2 and the results of the panel regression in Section 4.3. Finally, we discuss the results in section 5 and end with a conclusion in Section 6. We attach an extra proof, supplementary figures, tables and the used Matlab code in the appendix in Section 7. This paper contributes to current literature by applying the research in Einmahl et al. (2022) on a financial dataset and proposing a method of investigating driving factors in two dimensions through a panel regression.

2 Methodology

We aim to investigate the frequency of extremes of multivariate stock series, that are possibly cross-sectionally dependent, but serially independent. As way of describing a frequency of extremes in the univariate context with independent observations, Einmahl et al. (2016) propose a concept called 'scedasis'. In Einmahl et al. (2022) they build on that research by adapting to the multivariate case with possibly dependent time series. In their research, they use rainfall data. The aim is to apply their research to a financial dataset. We summarize the construction of scedasis in section 2.1 and the testing on scedasis in 2.2.1 and 2.2.2. We propose a panel regression in section 2.2.3 and finally apply the analysis on a financial dataset in section 2.3.

2.1 Scedasis

Consider a random non-stationary multivariate negated stock return time series $\{\xi_{i,j}\}$ with possible dependence in the cross section and independence over time. The multivariate time series of negated stock returns consists of n timepoints and m stocks. For each random variable $\xi_{i,j}$ ($i = 1,...,n$ and $j = 1,...,m$), $F_{i,j}$ is the distribution function. We denote x^* as the right endpoint of a continuous distribution function $F_0(x)$. We assume that all $F_{i,j}$ share the same endpoint x^*. We define that $F_{i,j}$ has a 'scedasis' $c\left(\frac{i}{n}, j\right)$ with respect to F_0 if

$$c\left(\frac{i}{n}, j\right) = \lim_{x \uparrow x^*} \frac{1 - F_{i,j}(x)}{1 - F_0(x)} \qquad i = 1,..,n \quad j = 1,..,m, \tag{1}$$

where $c(\cdot, j)$ is a positive continuous function on [0,1] for each stock j. The function $c\left(\frac{i}{n}, j\right)$ can be interpreted as the relative frequency of extremes at timepoint i for stock j. If, for any j, the scedasis is constant over time, the stock j has homoscedastic extremes. One could compare this to homoscedasticity when considering variance of a time series, only now specified for extremes. The situation in which c changes over time is referred to as heteroscedastic extremes (Einmahl et al., 2016). If we integrate over time from 0 up to t, we construct the 'integrated scedasis' C:

$$C_j(t) = \frac{1}{m}\int_0^t c(x,j)dx \qquad j = 1,..,m. \tag{2}$$

As described in Einmahl et al. (2022), the integrated scedasis can be used for testing equality over time and in the cross section using appropriate test statistics. In order for c to be uniquely defined, we impose the condition that the sum of the integrated scedasis for all stocks combined over the full time span equals one, i.e. $\sum_{j=1}^{m} C_j(1) = 1$. This way, the integrated scedasis represents the relative frequency of extremes of each stock. We provide an estimator for $C_j(t)$ in section 2.3.

In addition, in equation (14), we assume that $F_0(x)$ is in the maximum domain of attraction of a generalized extreme value distribution G_γ, i.e. the normalized maximum of a sample generated by F_0 converges to G_γ for some $\gamma \in \mathbb{R}$:

$$\lim_{n\to\infty} F_0^n(a_n x + b_n) = G_\gamma(x) := \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}, \tag{3}$$

for some sequences of constants $a_n > 0$ and b_n and all x for which $1+\gamma x > 0$. If the above holds, we write $F_0 \in \mathscr{D}(G_\gamma)$ (de Haan et al., 2015). Equation (14) and (3) together implies that that for all $F_{i,j}$, $F_{i,j} \in \mathscr{D}(G_\gamma)$, where γ is the common extreme value index by definition. Note that the parameter γ is constant for all stocks (j).

2.2 Testing on Scedasis of Stock Returns

In this subsection, we use the integrated scedasis to test for equality of scedasis, both in the cross-section and over time. The estimator of C_j is the relative number of exceedances over the threshold (Einmahl et al., 2022). We therefore define the estimator as follows:

$$\hat{C}_j(t) := \frac{1}{k}\sum_{i=1}^{nt} \mathbb{1}_{\{\xi_{i,j}>u\}} \qquad 0 \le t \le 1 \quad j = 1,..,m, \tag{4}$$

where k (see section 2.3) is the amount of observations of all stocks above a certain threshold u. For robustness, we perform each test in the financial application with varying thresholds. We test the equality of scedasis in the cross-section over the full time span in section 2.2.1 and time-invariance of scedasis for all stocks combined in section 2.2.2. If we find different scedases either in the cross-section or over time, we finally apply a panel regression to discover a possible driving force in section 2.2.3.

2.2.1 Equality of scedasis in the cross-section

If we find the integrated scedases of all stocks to be the same, each one would equal $1/m$ due to the restriction of the sum over all stocks equaling 1. We therefore test $H_0 : C_j(1) = \frac{1}{m}$ for all $j = 1,\ldots,m$. Einmahl et al. (2022) propose the test by checking whether the limit vector of

$$\left(\sqrt{k}\left(\widehat{C}_1(1)-\frac{1}{m}\right),\sqrt{k}\left(\widehat{C}_2(1)-\frac{1}{m}\right),\ldots,\sqrt{k}\left(\widehat{C}_m(1)-\frac{1}{m}\right)\right) \tag{5}$$

has mean zero. It has a covariance matrix defined by Σ. To express Σ, we assume the existence of the following tail copula function (Einmahl et al., 2022):

$$R_{j_1,j_2}(x,y)=\lim_{t\downarrow 0}\frac{1}{t}P\left(1-F_{i,j_1}\left(X_{i,j_1}\right)\leq tx, 1-F_{i,j_2}\left(X_{i,j_2}\right)\leq ty\right), \tag{6}$$

where $(x,y)\in[0,\infty]^2\backslash\{(\infty,\infty)\}$. If $R_{j_1,j_2}(x,y)=0$, we have complete tail independence and for $R_{j_1,j_2}(x,y)=x\wedge y$, we have complete tail dependence. The theoretical value of each entry of the 'true' covariance matrix Σ is then given by:

$$\sigma_{j_1,j_2}=\frac{1}{m}\int_0^1 R_{j_1,j_2}\left(c\left(u,j_1\right),c\left(u,j_2\right)\right)du. \tag{7}$$

The following estimator is a consistent estimator of σ_{j_1,j_2} for $j_1\neq j_2$, as $n\to\infty$ (Einmahl et al., 2022):

$$\hat{\sigma}_{j_1,j_2}=\frac{1}{k}\sum_{i=1}^{n}\mathbb{1}_{\{X_{i,j_1}>u,X_{i,j_2}>u\}}, \tag{8}$$

where u is the common threshold. Together, the entries form the consistent estimator $\hat{\Sigma}$ of Σ. Define $\mathbb{1}_m$ as the m-unit vector and I_m the m-dimensional identity matrix. From Theorem 2.3 in Einmahl et al. (2022) under H_0, we get the following corresponding test statistic for our null hypothesis:

$$T_n := D_{m-1}'\left(\left(M\hat{\Sigma}M'\right)_{m-1}\right)^{-1}D_{m-1}, \tag{9}$$

where the subscript $m-1$ means it consists of the first $m-1$ rows and the first $m-1$ columns only. Furthermore:

$$D=\sqrt{k}\left(\widehat{C}_1(1)-\frac{1}{m},\ldots,\widehat{C}_m(1)-\frac{1}{m}\right)' \quad \text{and} \tag{10}$$

$$M=I_m-\frac{1}{m}\mathbb{1}_m\mathbb{1}_m'. \tag{11}$$

If we assume invertability of $\hat{\Sigma}$, then, under H_0: $T_n\xrightarrow{d}\chi^2_{m-1}$, as $n\to\infty$.

2.2.2 Constant scedasis over time

To test whether scedasis is constant over time, we use $t\in[0,1]$ to be the relative timepoint of the full timespan. If we find, for a stock $j\in\{1,\ldots,m\}$ and for any $0\leq t\leq 1$, the integrated scedasis to be t times the total integrated scedasis, the scedasis $c(\cdot,j)$ is constant over time. Hence, we test $H_{0,j}:C_j(t)=tC_j(1)$ for each stock j. We define a Brownian bridge as B. Following the method in Einmahl et al. (2022), we use the following process, which holds under $H_{0,j}$:

$$Q_j(t)=\sqrt{k\widehat{C}_j(1)}\left(\frac{\widehat{C}_j(t)}{\widehat{C}_j(1)}-t\right)\xrightarrow{d}B(t) \qquad 0\leq t\leq 1. \tag{12}$$

Using Q_j, we construct the following test statistic, similar to the method in Einmahl et al. (2016):

$$T_j=\sup_{0\leq t\leq 1}|Q_j(t)|, \tag{13}$$

which, under $H_{0,j}$, follows the Kolmogorov distribution by construction (Kolmogorov, 1933). Therefore, we can use the p-values of the Kolmogorov-Smirnov test statistic (Smirnov, 1939).

2.2.3 Panel Regression

If we find either inequality of scedasis in the cross-section or over time, we aim to investigate the potential driving factor behind this difference in scedasis. The application in this paper is on stock returns, hence we look for driving factors of extreme (negative) stock returns. Fama and French (1995) find that Book-To-Market ratio (BM) and Market Equity (Size) together capture a large part of the cross-section of average stock returns. We therefore expect these factors to drive extreme values of stock returns in the cross-section. Since integrated scedasis is estimated using a certain time span, we estimate the scedasis value for each year ($y = 1,...T$) in the dataset. Consider the date of each observation $\xi_{i,j}$. We estimate the scedasis yearly for each company separately. Since the dataset consists of 20 years ($T = 20$), we consider $C_{j,y}$ in a similar way to model (2), only in subsets with length one year. We define $n_y = t_y - t_{y-1} + 1$ as the amount of trading days in year y, where t_y is the index in the dataset of the final stock date of year y with $t_0 = 0$ and $t_T = n$. Since some years contain many and other years few extremes, the yearly Scedases can vary greatly over time. In order to investigate the effects of the company specific factors on Scedasis, we therefore normalize the yearly Scedasis values. To do so, we consider the following yearly scedasis function:

$$c_y\left(\frac{i}{n_y}, j\right) = \lim_{x \uparrow x^*} \frac{1 - F_{i,j}(x)}{1 - F_0(x)} \qquad i = t_{y-1}, .., t_y \quad j = 1, .., m. \tag{14}$$

Consequently, the integrated yearly Scedasis value for each company becomes

$$C_{j,y}(t) = \frac{1}{m} \int_0^t c_y(x, j) dx \qquad j = 1, .., m, \qquad y = 1, .., T, \tag{15}$$

with $\sum_{j=1}^{m} C_{j,y}(1) = 1$ for each year y. The estimates of the yearly scedases become

$$\hat{C}_{j,y}(t) = \frac{1}{k_y} \sum_{i=t_{y-1}+1}^{t_{y-1}+t \cdot n_y} \mathbb{1}_{\{\xi_{i,j} > u\}} \qquad j = 1, .., m, \qquad y = 1, .., T, \tag{16}$$

where k_y is the amount of observations above the common threshold u in year y. For each year y, we restrict $\sum_{j=1}^{m} C_{j,y}(1) = 1$ and note that $\sum_{j=1}^{m} \hat{C}_{j,y}(1) = 1$.

We are interested in the effect of Size and Book-to-Market ratio on scedasis over time and in the cross-section, hence we consider a panel regression. The created dataset consists of maximally $m = 50$ stocks and $T = 22$ years. The total amount of observations is: $n_C = m \times T = 50 \times 22 = 1100$. We consider the Error components model, first proposed by Balestra and Nerlove (1966), because it performs well for both a small m and T (Mátyás, 1992). Using this model, one extracts information on the effects of factors on scedasis from variation in the cross-section, variation over time and variation of the factors (Maddala, 1971). With the Fama and French factors, the error components model is in our case given by:

$$\hat{C}_{j,y} = \alpha + \beta_{size} X_{size,j,y} + \beta_{BM} X_{BM,j,y} + \varepsilon_{j,y} \qquad j = 1, .., m, \quad y = 1, .., T, \tag{17}$$

where α is a constant and β_p the effect of factor p on scedasis. Furthermore, $X_{p,j,y}$ is the value of factor p of firm j at year t. We denote $\beta' = [\beta_{size}, \beta_{BM}]$ and $X'_{j,y} = [X_{size,j,y}, X_{BM,j,y}]$. We define the error term as in Wallace and Hussain (1969): $\varepsilon_{j,y} = u_j + v_y + w_{j,y}$ with each term normally distributed with mean zero and variance terms σ_u^2, σ_v^2, and σ_w^2

respectively. We assume independence between each of the three terms. For convenience, we formulate the model in matrix form:

$$\hat{C} = (1, X)\begin{pmatrix}\alpha \\ \beta\end{pmatrix} + \varepsilon, \tag{18}$$

with $\hat{C}$ a $mT \times 1$ vector consisting of all vectors of scedasis values per firm stacked and ordered chronologically: $\hat{C} = (\hat{C}_{1,1}, \hat{C}_{1,2}, .., \hat{C}_{1,T}, \hat{C}_{2,1}, .., \hat{C}_{m,T})$. Stacked similarly, X is a $mT \times 2$ matrix and ε is a vector of length mT. Following Wallace and Hussain (1969), the best linear unbiased estimate of β is given by:

$$\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}(X'\hat{\Sigma}^{-1}\hat{C}), \tag{19}$$

with variance $\hat{\Sigma}_{\beta} = (X'\Sigma^{-1}X)^{-1}$.

For Σ we use the Aitken estimator below, which has a smaller asymptotic variance than regular covariance estimators (Wallace and Hussain, 1969). The general idea is to divide the variance into components belonging to variance in the cross-section (σ_u^2), variance over time (σ_v^2) and combined variance (σ_w^2):

$$\hat{\Sigma} = \hat{\sigma}_w^2 I_{mT} + \hat{\sigma}_u^2 \mathbf{A} + \hat{\sigma}_v^2 \mathbf{B}, \tag{20}$$

where I_{mT} is an $mT \times mT$ identity matrix and $\mathbf{A}$ and $\mathbf{B}$ are defined as follows:

$$\mathbf{A} = \begin{bmatrix} J_T, & 0, \ldots, 0 \\ 0, & J_T, \ldots, 0 \\ \vdots & \\ 0, & 0, \ldots, J_T \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} I_T, & I_T, \ldots, I_T \\ I_T, & I_T, \ldots, I_T \\ \vdots & \\ I_T, & I_T, \ldots, I_T \end{bmatrix}. \tag{21}$$

Here, J_T is a $T \times T$ matrix of ones and both $\mathbf{A}$ and $\mathbf{B}$ consist of m rows and columns of block matrices. Following Wallace and Hussain (1969), the empirical estimates for the variance terms are given by:

$$\begin{aligned} \hat{\sigma}_w^2 &= \frac{1}{(m-1)(T-1)} \sum_j \sum_y \left(\hat{\varepsilon}_{jy} - \frac{1}{T}\hat{\varepsilon}_{j.} - \frac{1}{m}\hat{\varepsilon}_{.y} \right)^2, \\ \hat{\sigma}_u^2 &= \frac{1}{T}\left[\sum_{j=1}^{m} \frac{\hat{\varepsilon}_{j.}^2}{T(m-1)} - \hat{\sigma}_w^2 \right], \\ \hat{\sigma}_v^2 &= \frac{1}{m}\left[\sum_{y=1}^{T} \frac{\hat{\varepsilon}_{.y}^2}{m(T-1)} - \hat{\sigma}_w^2 \right], \end{aligned} \tag{22}$$

with $\hat{\varepsilon}_{jy}$ the estimated OLS error term in model (18). The dot in a subscript represents the sum over the corresponding subscript. Hence, $\hat{\varepsilon}_{j.}$ is the sum over the years y of all errors corresponding to firm j. For computational convenience we use the closed-form term for the inverse of $\hat{\Sigma}$ given in the Appendix (Wallace and Hussain, 1969).

2.3 Application to Financial Datasets

We apply the method of estimating scedasis to a financial dataset. We consider negated daily nominal returns of 50 chosen stocks in the S&P500. For the analysis, we require a 'universal threshold' of extremes for the full dataset.

Furthermore, for our multivariate time series to be viable for the estimation of scedasis and the methodology in section 2.2 we need serial independence. Stock returns, however, are known to exhibit serial dependence. Finally, in Section 2.1, we assume a constant extreme value index (γ) for all considered stocks. Hence, we only choose stocks with an equal extreme value index. We set a universal threshold for extremes in section 2.3.1, address the removal of possible serial dependence in section 2.3.2 and the equality of the extreme value index for all stocks in 2.3.3.

2.3.1 Universal Extreme value threshold

For all sections in this paper, we are concerned with tail analysis. This analysis is applied to 'extreme values', which are the values belonging to the tail of the distribution. Above a certain threshold, observations are considered extreme values. For the comparison between scedasis of different time series, we need equal thresholds for all returns of stocks. Therefore, we choose one 'common threshold' above which all observations are considered extreme values. For the construction of the common threshold, we take all observations of the multivariate series together to get the univariate series ξ_r ($r = 1,.,,N$), where $N = n \times m$. Let $\xi_{1:N} \leq \xi_{2:N} ... \leq \xi_{N:N}$ be the ordered observations of the series ξ_r. We set $u = \xi_{N-k:N}$, the k-th largest value, as the threshold above which we consider the observations to be extreme values.

The intuition for the right choice of k is as follows: we are estimating the observations that are part of the 'tail region'. If we choose a k too high, the observations get too close to the central part of the distribution, where tail function analysis is not viable. In the extreme value index context, this would result in a large bias of the estimate of the extreme value index. On the other hand, if k is too small, the variance of the estimate may grow too large. This is an example of the classical bias-variance trade-off (Dominicy et al., 2017). For appropriate choices of k, we use a so-called Hill plot (Drees et al., 2000), which is the 'Hill estimator' of the extreme value index plotted against k. An appropriate k is a value in the 'stable region' in the plot, where the trade-off between variance and bias is optimal. Within this 'stable region', we vary the chosen k for robustness. We use the Hill Estimator (Hill, 1975) for estimating the extreme value index, which relies on these fat tails in estimation:

$$\hat{\gamma} = \frac{1}{k}\sum_{i=0}^{k-1} log(\xi_{N-i:N}) - log(\xi_{N-k:N}) \qquad (k < N). \tag{23}$$

Quintos et al. (2001) show Hill estimators for the extreme value index to be constant in their application on stock returns, hence we discard possible time variance of γ.

We use the initial 50 stocks combined in estimating $\hat{\gamma}$ for different values of k. We choose the common threshold $u = \xi_{N-k:N}$ for the chosen k in the 'stable region' in the Hill plot. Throughout one run of the analysis, the threshold value u is fixed. N and k, however, will be (slightly) different after the removal of observations and stocks in section 2.3.2 and 2.3.3. Throughout each run of the analysis, we use one universal threshold to ensure the set of extremes after data selection is in accordance with the observations used in the analysis. The value $\hat{\gamma}$ is used as a constant estimate for the γ parameter in the scedasis analysis. Henceforth, we refer to this value when we write γ. For robustness, we conduct the full analysis for a varying threshold within the 'stable region' of the Hill plot.

2.3.2 Serial Dependence

In this subsection we explain how we remove potential serial dependence in the extremes of the stock return series. In the analysis, we only consider extreme values of stock returns. Hence, we aim to remove the serial dependence in the extremes. If we can determine how many extremes are part of the same 'cluster of extremes', we can eliminate extreme values that belong to the same cluster. This way, maximally one value per cluster is selected and serial dependence is kept to the minimum. For the purpose of cluster size estimation, we examine the 'Extremal index', first described in Cartwright (1958). Leadbetter (1982) show that the extremal index can be interpreted as the reciprocal of the mean cluster size. The extremal index is defined as follows:

We consider a univariate stationary sequence. For notational convenience, we write ζ_i in the univariate case (instead of $\xi_{i,j}$ in the multivariate case). The univariate stationary sequence ζ_i has distribution function F with the finite endpoint $x^* = \sup\{x : F(x) < 1\}$. We use a common threshold u above which an observation is considered as extreme, and define $T(u)$ the time between extremes:

$$T(u) = \min\{y \geq 1 : \zeta_{y+1} \geq u\}. \tag{24}$$

We denote $M_{k,l} = \max\{\xi_i : i = k+1, ..., l\}$. The distribution of the interexeedance time, as shown in Ferro and Segers (2003) is then given by:

$$P\{T(u) > y\} = P(M_{1,y+1} \leq u \mid X_1 > u). \tag{25}$$

Ferro and Segers (2003) prove that, under some mild conditions, as $n \to \infty$:

$$(1 - F(u))T(u) \xrightarrow{d} T_\theta \qquad \text{as} \quad u \uparrow x^* \quad \text{and} \tag{26}$$

$$T_\theta \sim (1-\theta)\varepsilon_0 + \theta\mu_\theta, \tag{27}$$

where ε_0 is the degenerate distribution at 0 and $\mu_\theta \sim exp(\theta)$. The extremal index is the θ parameter in this framework. We note that the value of θ is independent of the threshold, for a high threshold u. For the estimation of the Extremal index, Ferro and Segers (2003) uses the number of observations above the common threshold (u). We further denote S_i as the $i-$th exceedance time. Following this definition, we have that the observed interexceedance times are $T_i = S_{i+1} - S_i$ for $i = 1, ..., k-1$. The T_i are the observed interexceedance times and $T(u)$ is the estimated interexceedance time given a high threshold u. In their paper, Ferro and Segers (2003) propose several estimators for θ. We use the 'intervals estimator', as it is shown to perform well for both low and high thresholds in Ferro and Segers (2003). The estimator is defined as follows:

$$\theta_n(u) = \begin{cases} 1 \wedge \hat{\theta}_n^*(u), & \text{if} \quad \max\{T_i : 1 \leq i \leq k-1\} > 2, \\ 1 \wedge \hat{\theta}_n(u), & \text{if} \quad \max\{T_i : 1 \leq i \leq k-1\} \leq 2, \end{cases}$$

where $\hat{\theta}_n^*(u)$ and $\hat{\theta}_n(u)$ are given by:

$$\hat{\theta}_n^*(u) = \frac{2(\sum_{i=1}^{k-1}(T_i - 1))^2}{(k-1)\sum_{i=1}^{k-1}(T_i-1)(T_i-2)}, \tag{28}$$

$$\hat{\theta}_n(u) = \frac{2(\sum_{i=1}^{k-1} T_i)^2}{(k-1)\sum_{i=1}^{k-1} T_i^2}. \tag{29}$$

We perform estimation of the extremal index for each considered stock separately. Each stock is a univariate stationary sequence $\xi_{i,j}$ for fixed j. We estimate the extremal index to generate estimates $\hat{\theta}_{n,j}$ for $j = 1, .., m$.

In order for serial dependence in extremes to be removed, we remove extremes that belong to the same cluster. We set $\theta_{min} = \min\{\hat{\theta}_{n,j}; j = 1, .., m\}$. The largest average cluster size for the chosen stocks is equal to $1/\theta_{min}$. In order to avoid creating any missing datapoints, we only remove timepoints from the dataset. In other words, if we remove an instance, we remove the instance for all stocks simultaneously. We use the largest average cluster size as minimum 'distance' between extreme values to make sure all possible clusters are dealt with. We once again consider the multivariate time series $\xi_{i,j}$ ($i = 1, .., n$ and $j = 1, .., m$) and the common extreme value threshold u. To make sure no extremes from the same cluster are kept, we apply the following iterative procedure for for all stocks together, where the largest extreme is kept for extremes in the same cluster:

Algorithm 1 Removal of serial dependence in extremes

procedure CONSIDER THE MULTIVARIATE SEQUENCE $\xi_{i,j}$
  for $j = 1 : m$ **do**
    for $i = 1 : n - \lceil 1/\theta_{min} \rceil$ **do**
      if $\xi_{i,j} \geq u$ **then**
        for $r = 1 : \lceil 1/\theta_{min} \rceil$ **do**
          if $\xi_{i+r,j} \geq u$ and $\xi_{i+r,j} > \xi_{i,j}$ **then**
            $\xi_{i,j} := -\infty$
          if $\xi_{i+r,j} \geq u$ and $\xi_{i+r,j} \leq \xi_{i,j}$ **then**
            $\xi_{i+r,j} := -\infty$
          if $\xi_{i+r,j} < u$ **then**
            Continue
Remove all rows i with at least one entry equal to $-\infty$

Consequently, for each considered stock j, for all extremes possibly belonging to the same cluster, only the day with the largest value occurring is retained[1]. By deleting timepoints, we have retained a dataset without 'holes'. Henceforth, *after* the removal of serial dependence, n^* is the length of the dataset and k^* the amount of extreme values. Consequently, we have $N^* = n^* \times m$.

[1] After the removal of observations, we find no evidence of remaining serial dependence in partial autocorrelations

2.3.3 Equality of Extreme Value Index

We consider the same method of estimating the extreme value index as we did in Section 2.1. However, we use each stock separately to get an estimate of the extreme value index. Furthermore, we consider the dataset after the removal of serial dependence, which is of length n^*. Similar to estimate (23) in Section 2.1, the estimate consequently becomes the following:

$$\hat{\gamma}_j = \frac{1}{k_j} \sum_{i=0}^{k_j-1} log(\xi_{n^*-i:n^*}) - log(\xi_{n^*-k_j:n^*}) \qquad (k_j < n^*) \qquad j = 1, .., m, \tag{30}$$

where the value of k_j for each stock is given by:

$$k_j = \sum_{i=1}^{n^*} \mathbb{1}_{\xi_{i,j} \geq u} \qquad j = 1, .., m, \tag{31}$$

where $\mathbb{1}_x$ is an indicator function equaling one if the subscript is true and zero otherwise. By construction, $\sum_{j=1}^{m} k_j = k^*$. Gomes et al. (2007) show that when the time series $\xi_{i,j}$ is heavy-tailed, the Hill estimator is consistent for the estimation of the extreme value index. They show that $\sqrt{k_j}(\hat{\gamma}_j - \gamma)$ weakly converges to $N(0, \gamma^2)$. Hence, for the estimate of the variance of the Hill estimators we use the squared value of the estimate itself divided by k_j, $(\hat{\gamma}_j^2 / k_j)$, for each $j = 1, .., m$. We then test whether the estimate is significantly different from the extreme value index with all observations combined, $\hat{\gamma}$, as derived in Section 2.1 using a regular t-test with test statistic $t_j = \dfrac{|\hat{\gamma}_j - \hat{\gamma}|}{\hat{\gamma}_j / \sqrt{k_j}}$. We select only the stocks for which t_j does not exceed the critical value of 1.96 (at a confidence level of 95%). After this selection of stocks we have m^* stocks and N^* becomes $N^{**} = n^* \times m^*$. In the scedasis analysis, we use $k = k^*$, $n = n^*$, $m = m^*$ and $N = N^{**}$.

3 Data

In this Section, we display summary statistics of the datasets used in our application. All data is acquired through Wharton Research Data Services [2]. We display a summary of negated nominal returns of five SP 500 stocks between January 3rd 2000 until 30th of March 2022 in Table 1,where we consider the original dataset of 50 stocks. In Table 1 we see the negated returns are slightly negative on average, hence on average returns are slightly positive. The standard deviation of each stock is much higher than its mean.

[2] Wharton Research Data Services URL

Company	Mean ($\times 10^{-3}$)	St. Dev.	Min	Max
FISERV INC	-0.702	0.022	-0.769	0.195
GENERAL DYNAMICS CORP	-0.472	0.016	-0.117	0.124
PEPSICO	-0.347	0.013	-0.207	0.119
AMGEN	-0.339	0.020	-0.315	0.134
SCHLUMBERGER	-0.416	0.024	-0.199	0.274

Table 1: Summary statistics of five S&P 500 companies' daily negated nominal returns from January 3rd 2000 until March 30, 2022 (consisting of 5598 trading days). All fifty companies are considered, hence $m = 50$. Consequently, $N = 50 \times 5598 = 279900$. Mean, standard deviation, minimum and maximum value are displayed.

Company	$\hat{C}_{j,y}$	BM	Size
CMS ENERGY CORP	0.018 (0.037)	0.823 (0.521)	6,771 (5,881)
TEXTRON INC	0.048 (0.061)	0.496 (0.225)	9,969 (5,220)
LILLY ELI & CO	0.019 (0.031)	0.140 (0.074)	92,623 (44,110)
WEYERHAEUSER CO	0.022 (0.033)	1.106 (0.790)	11,149 (8,903)
EQUIFAX INC	0.027 (0.044)	0.212 (0.109)	9,144 (7,683)

Table 2: Overview of 5 S&P 500 companies' normalized yearly Scedasis ($\hat{C}_{j,y}$), Book-to-Market ratio (BM) and Size. The years considered are 2000 until 2021, hence $T = 22$ We once again consider 50 companies, hence $n_C = m \times T = 50 \times 22 = 1100$. The mean values per company are given with standard deviation within parentheses.

Table 2 shows the mean values and standard deviation of the yearly Scedasis, Book-to-Market ratio and Size of 5 random S&P 500 companies in the second dataset. The set of companies used for the daily and yearly datasets is equal. The average yearly Scedasis values are positive, yet small. This is due to $\sum_{j=1}^{m} \hat{C}_{j,y} = 1$ for each year y. Table 2 further shows positive Book-to-Market ratios, yet only one company with an average Book-to-Market ratio higher than one. We further note a considerable difference in size both between companies as per company over time, given the standard deviations.

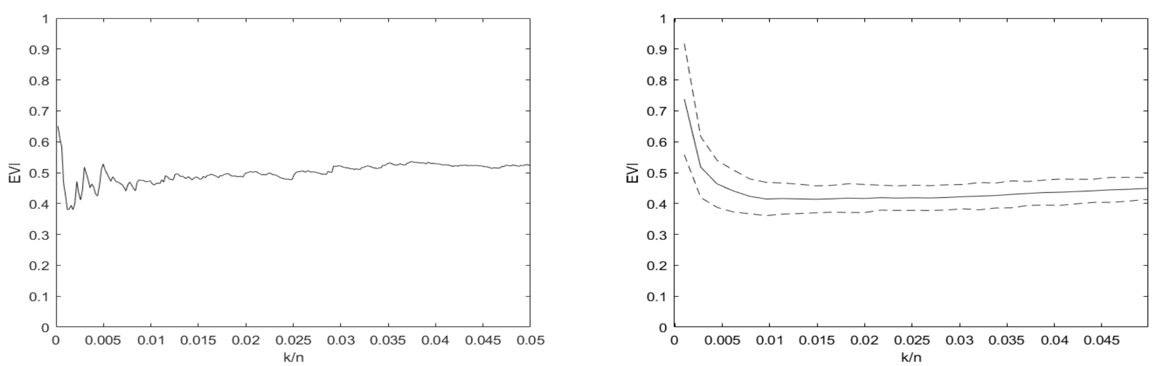
4 Results

The application of the methodology on a financial dataset yields results in three parts. First, we present results of the Extreme Value Index and the removal of serial dependence in Section 4.1. Secondly, in Section 4.2, we provide results of the tests for constant Scedasis over time and in the cross-section. Finally, we present results of the panel regression in Section 4.3.

4.1 Extreme Value Index and Serial Dependence

For the application of the analysis on negated stock returns, we require a choice of extreme value threshold. Figure 1(a) shows an example of a 'Hill plot' for one specific stock return series and Figure 1(b) the Hill estimates using all available data. Figure 1(b) shows that the variation of the Hill estimate decreases with a lower threshold. However, starting at a certain point, the bias slightly increases. Based on Figure 1(b), we choose the value $k/n = 0.015$ to be the ideal choice of extremes in the bias-variance trade-off. $k/n = 0.015$ corresponds to a threshold return of 5.2%, which occurs on average once every 67 trading days ($1/0.015 = 67$). Hence, we consider a negated daily nominal return of at least 5.2% to be an extreme occurrence.

We determine the extremal index for the removal of Serial Dependence in extremes using the chosen threshold: $\hat{\theta}_{min}^{-1} = 8.165$. Hence, we consider two extremes within 8 trading days of one another to belong to the same cluster of extremes. Algorithm 1 yields the declustered dataset. We show the figures containing partial autocorrelations before and after the removal of serial dependence in Appendix 7.2. The figures in the Appendix show the considered dataset contains significant partial autocorrelations up to at least 17 lagged observations. The figures in the Appendix further shows that the application of Algorithm 1 removes autocorrelation accordingly.



(a)

(b)

Figure 1: Figure (a) and (b) show extreme value index estimates for different thresholds. In (a) the Hill estimate of the stock returns of 'Masco Corporation' is shown for different thresholds. In (b), 50 stocks are considered as one series. The expected value and 95% confidence intervals are constructed in the following manner: we choose 50 stocks z times. For each selection, a new extreme value index is estimated. Figure (b) shows the empirical mean and 95% confidence bounds of the z estimates ($z = 100$).

4.2 Test results

Figure 2 shows the p-values of the test for equality of Scedasis in the cross section for decreasing thresholds. For the considered threshold of $k/n = 0.015$, we do not reject equality of Scedasis in the cross-section. This result holds true for all values of k/n up to 0.075. The difference is significant for higher k/n, which means for very low thresholds only. Hence, we find no evidence of differences in integrated Scedasis across companies.

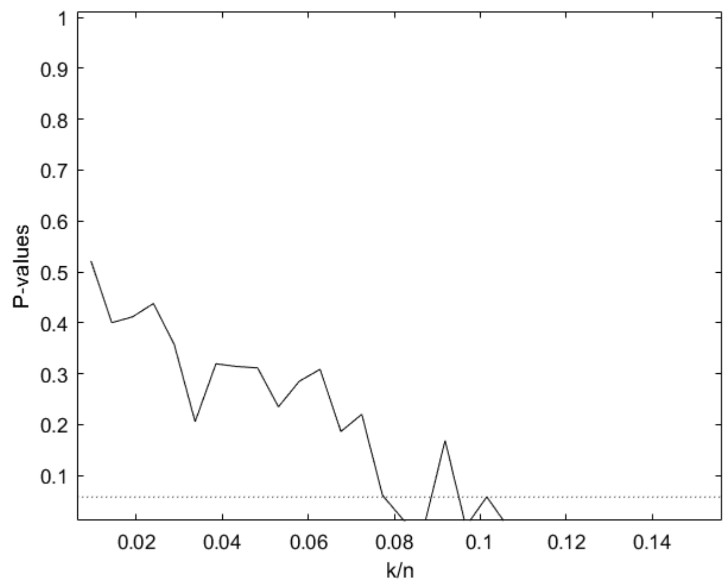


Figure 2: P-values of the test of equality of Scedasis in the cross section for different thresholds. The P-values correspond to the Chi-squared distribution of the test statistic for the null hypothesis in (9). The horizontal dotted line corresponds to the critical value of rejection at a rejection rate of $\alpha = 5\%$

Figure 3 shows the Rejection Rate of test statistic (13) for all stocks. When the Rejection Rate equals 1, all 50 stocks reject equality of Scedasis over time. Equality of Scedasis over time is rejected in the majority of considered stocks for k/n up to 0.12. For $k/n = 0.015$, we find approximately 95% of stocks to reject the hypothesis at a significance level of $\alpha = 5\%$. The stocks that do not reject equality of Scedasis over time for the considered threshold are 'Textron Inc' and 'Pinnacle West Capital Corp'. If $k/n = 0.015$, we have an extreme occurrence every 67 tradingdays on average. During crises, however, extreme occurrences are more frequent, and in economically stable periods extreme negative returns are less frequent. For the considered threshold, this phenomenon is significant for all stocks in the period 2000 until 2020, except for 'Textron Inc' and 'Pinnacle West Capital Corp'. If stocks have a constant Scedasis over time, the difference between the frequency of extremes in economically stable periods and during crises is not significant. Further research could investigate whether this indicates that these two stocks are less susceptible for economic downturn.

However, since we find 95% of stocks to reject equality of Scedasis over time for $k/n = 0.015$, we do reject the general hypothesis that Scedasis is equal over time for the considered stock return application. With these results, we find the hypothesis of equal Scedasis rejected in at least one dimension, hence we conduct the panel regression to find possible driving factors.

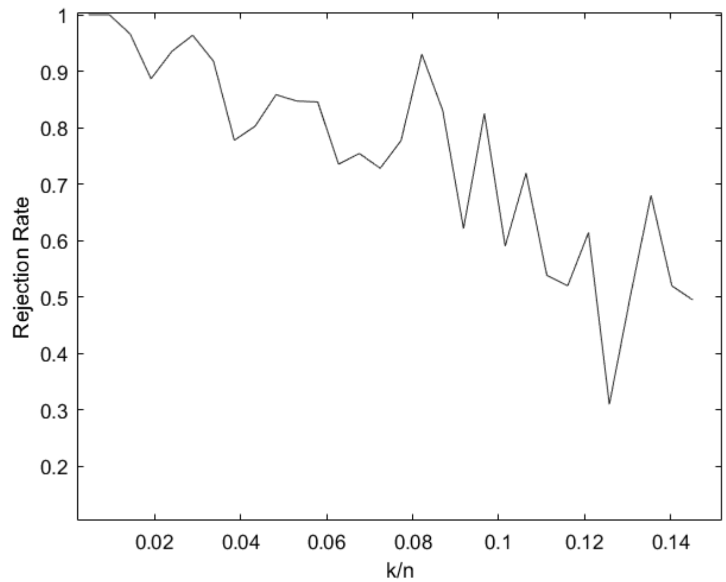


Figure 3: Rejection rate of stocks in the simultaneous testing on all stocks for equality of Scedasis over time for different thresholds. The rejection of a single stock is the rejection of test statistic (13) at a rejection rate of $\alpha = 5\%$. If the rejection rate is one, all stocks simultaneously reject equality of Scedasis over time. If the rejection rate is 0.5, half of the stocks reject, etc.

4.3 Panel Regression results

We aim to investigate possible company-specific driving factors of yearly Scedasis. Figure 4 shows so-called Boxplots of the normalized yearly Scedasis values of a set of randomly selected stocks in the dataset. At first glance, these frequencies of extremes do appear to vary in the cross-section. For example, for 'Hasbro Inc.' (HAS) and 'Colgate-Palmolive' (CL), the 25th percentile is higher than the 75th percentile of 'CMS Energy Corporation' (CMS) and 'Sempra Energy' (SRE). This suggests differences in normalized yearly Scedasis n the cross-section. We want to find the driving factor behind these differences in the frequencies of extremes.

Table 3 shows the results of panel regressions of possible driving factors on yearly Scedases. It shows 7 panel regressions each using a different combination of the explanatory variables. All Scedasis is normalized, hence it represents the relative difference in Scedasis across companies. Therefore, if the Scedasis of one company increases, the relative frequency of extremes of this specific stock increases relative to other stocks. We therefore aim to find differences between companies. For this purpose we demean the explanatory variables before analysis. Hence we consider the yearly demeaned log difference in Book-To-Market ratio and Size along with their lagged values and all considered values squared. We include lagged values because they could potentially indicate predictive potential of the factors on future Scedasis and squared values to assess the change in the slope of the relationship of the regular value on Scedasis.

We find the lagged value of Book-To-Market ratio to be significant and positive in all combinations of explanatory variables in which it is considered. This means we find predictive power of Book-To-Market ratio on integrated Scedasis. When its squared value is included, the coefficient of the quadratic value is significant and negative. Therefore, an increase in the difference between the Book-To-Market ratio of last year and 2 years ago has a positive effect on the current integrated Scedasis value. Due to the negative coefficient of the squared value, the marginal effect decreases with the level of the difference. For other coefficients we find no consistently significant results. Consequently, in contrast with our hypothesis, we find no evidence that Size is a driving factor of Scedasis.

Moreover, we make a specific choice by setting $k/n = 0.015$. We base this decision on the Hill plot in Figure 1. Since that choice is somewhat subjective, we conduct a robustness check for $k/n = 0.010$ (Table 4 in Appendix 7.3) and another check using $k/n = 0.020$ (Table 5 in Appendix 7.3). In the robustness check with a lower threshold in Appendix Table 4 we find fewer significant results. The robustness check using the higher threshold in Appendix Table 5 yields similar results to the results in Table 3. Together, the results of the robustness checks show that the significance of the effect of the lagged Book-To-Market ratio and its quadratic term on normalized integrated Scedasis is not dependent on the specific choice of k/n, as long as it lies in the stable and least biased region of the Hill plot.

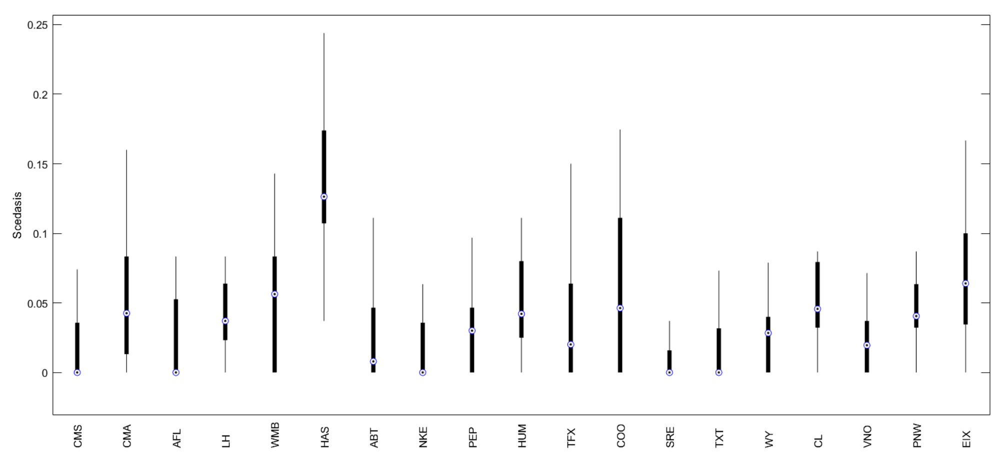


Figure 4: Plot of the distribution of Normalized yearly Scedasis values for a subset of random stocks in the dataset, where the stocks are given by their tickers. The maximum and minimum values, the median and 25th and 75th percentile of constructed yearly Scedasis values are displayed.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0334*	0.0325*	0.0327*	0.0320*	0.0322*	0.0321*	0.0333*
	(0.0017)	(0.0016)	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
BM_Diff	0.0134*	0.0133*	0.0138*	0.0015	0.0015		
	(0.0026)	(0.0025)	(0.0025)	(0.0009)	(0.0009)		
Size_Diff	0.0000	0.0008	0.0011	0.0009	0.0009		
	(0.0032)	(0.0032)	(0.0032)	(0.0008)	(0.0008)		
BM_Diff_sq	-0.0004*	-0.0004*	-0.0004*				
	(0.0001)	(0.0001)	(0.0001)				
Size_Diff_sq	0.0000	0.0000	0.0000				
	(0.0001)	(0.0001)	(0.0001)				
BM_Diff_Lag	0.0120*	0.0024*		0.0027*		0.0027*	0.0110*
	(0.0025)	(0.0009)		(0.0009)		(0.0009)	(0.0025)
Size_Diff_Lag	0.0098*	-0.0002		-0.0001		-0.0001	0.0128*
	(0.0036)	(0.0008)		(0.0008)		(0.0008)	(0.0037)
BM_Diff_Lag_sq	-0.0003*						-0.0003*
	(0.0001)						(0.0001)
Size_Diff_Lag_sq	-0.0002						-0.0003*
	(0.0001)						(0.0001)

Table 3: The estimated $\hat{\beta}$ parameter is given for the constant and explanatory variables created by the two factors in 7 different panel regressions. Book-To-Market ratio and Size are considered as their demeaned yearly log differences. The demeaning is done by taking the average of all yearly log differences of the 50 companies and subtracting the corresponding value from each instance. The standard deviation, the square root of the specific diagonal element of $\hat{\Sigma}_{\beta}$ in (19), is given within parentheses and an asterisk indicates significance at a level of $\alpha = 5\%$. Note: each column contains a specific panel regression.

5 Discussion

We find Scedasis to be varying over time, while invariant in the cross-section. Schwert (2011) state that high volatility is linked to economic downturn and crises. High volatility means more (negative) extremes and hence a higher Scedasis. Thus, a varying Scedasis over time is in line with current literature. The invariance in the cross-section could potentially be due to the few extremes that are left after declustering, which means only few datapoints are left to conduct the test on. Further research could indicate pairwise dependence and other possible relations of Scedasis in the cross-section.

When normalizing yearly integrated Scedasis values, we eliminate the common 'time' factor of all companies together in the analysis. Only the relative Scedasis between companies in each year is represented. In that case we do find an indication of differences in relative frequencies of extremes per company. Fama and French (1993) argue that a high Book-To-Market ratio is positively related to low earnings on assets, hence low stock returns. This indicates a stock will be under-performing and will be more likely to have (negative) extremes. This result holds true up to five years before and after the measurement of equity. An increase in relative yearly Scedasis value indicates relatively more frequent negative extremes compared to other stocks. In our analysis, we find an increase in Book-To-Market ratio to predict more frequent negative extremes one year in advance. Hence, the results are in line with the research in Fama and French (1993). Further research could indicate whether the results in this paper extend to more lags in the Book-To-Market ratio.

Furthermore, using the coefficient of the level and the quadratic term in Table 3, regression (7), we determine a maximum yearly Scedasis value at a difference in Book-To-Market ratio of $-0.0110/2*(-0.0003) = 1.833$. Therefore, if the difference in Book-To-Market ratio from one year to the next is 1.8, the increase in the relative frequency of extremes compared to other S&P 500 stocks reaches its maximum. The range of Book-To-Market ratio in the considered dataset is $[-0.6, 8.1]$, hence a difference value of 1.8 is feasible.

Fama and French (1993) further state that Size, when controlling for Book-To-Market ratio, only has a small effect on stock returns. We find no consistent evidence of any effect of Size on Scedasis. This could be due to the fact that we exclusively consider S&P 500 stocks, which correspond to large companies by construction. A possibility for further research is to conduct the same panel regression using stocks of companies of different sizes. Specifically, stocks in the SMB (Small minus Big) portfolio used for the Size factor in Fama and French (1993) would be adequate.

Finally, the research in this paper is limited to two possible driving factors of Scedasis over time and in the cross-section of stock returns. Further research could indicate whether other Fama-French factors, such as Investments or Operating profitability, proposed as factors in Fama and French (2015), are drivers of the frequency of extremes of stock returns. Clearly we are not limited to Fama and French factors when assessing company-specific factors and their effect on Scedasis. Hence, in further research, virtually any company-specific predictor can be explored.

6 Conclusion

In this study, we investigate the Scedasis of negated stock returns both in the cross-section and over time. The concept 'Scedasis' was first proposed in Einmahl et al. (2016) and elaborated further in Einmahl et al. (2022). The concept can easily be interpreted as the frequency of extremes. We summarize the propositions and statistical tests of Einmahl et al. (2022) and propose a panel regression to investigate two possible Fama-French factors as driving factors of Scedasis. We find Scedasis not to be varying in the cross section but varying over time. We find no evidence of the size of a company to be related to Scedasis. However, we find the yearly difference in Book-To-Market ratio to be a (predictive) driver of company-specific relative Scedasis. This is in line with our hypothesis. A higher Scedasis corresponds to more frequent extremes. Hence, if a companies' Book-To-Market ratio increases relative to others, only a risk-loving person should invest.

References

Balestra, P. and Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica: Journal of the Econometric society*, 34(3):585–612.

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American statistical association bisiness and economic statistics section*, pages 177–181.

Cartwright, D. E. (1958). On estimating the mean energy of sea waves from the highest waves in a record. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 247(1248):22–48.

de Haan, L., Tank, A. K., and Neves, C. (2015). On tail trend detection: modeling relative risk. *Extremes*, 18(2):141–178.

Dominicy, Y., Ilmonen, P., and Veredas, D. (2017). Multivariate hill estimators. *International Statistical Review*, 85(1):108–142.

Drees, H., Resnick, S., and de Haan, L. (2000). How to make a hill plot. *Annals of Statistics*, 28(1):254–274.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *Journal of derivatives*, 4(3):7–49.

Einmahl, J. H., de Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):31–51.

Einmahl, J. H., Ferreira, A., de Haan, L., Neves, C., and Zhou, C. (2022). Spatial dependence and space–time trend in extreme events. *Annals of Statistics*, 50(1):30–52.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, 50(1):131–155.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22.

Ferro, C. A. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453.

Gomes, M. I., Martins, M. J., and Neves, M. (2007). Improving second order reduced bias extreme value index estimation. *Revstat*, 5(2):177–207.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174.

Huang, W., Liu, Q., Rhee, S. G., and Wu, F. (2012). Extreme downside risk and expected stock returns. *Journal of Banking & Finance*, 36(5):1492–1502.

Kolmogorov (1933). Sulla determinazione empirica di una legge di distribuzione. In *Giornale dell'Istituto Italiano degli Attuari*, volume 4, pages 83–91.

Laogan, B. T. and Elnashai, A. S. (1999). Structural performance and economics of tall high strength rc buildings in seismic regions. *The Structural Design of Tall Buildings*, 8(3):171–204.

Leadbetter, M. R. (1982). Extremes and local dependence in stationary sequences. Technical report, NORTH CAROLINA UNIV AT CHAPEL HILL, DEPT OF STATISTICS.

Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica: Journal of the Econometric Society*, 39(2):341–358.

Mátyás, L. (1992). Error components models. In *The Econometrics of Panel data*, volume 39, pages 46–71. Springer.

Olsen, R. A. (1997). Investment risk: The experts' perspective. *Financial Analysts Journal*, 53(2):62–66.

Papalexiou, S. M. and Montanari, A. (2019). Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, 55(6):4901–4914.

Quintos, C., Fan, Z., and Phillips, P. C. (2001). Structural change tests in tail behaviour and the asian crisis. *The Review of Economic Studies*, 68(3):633–663.

Rabbani, A. G., Grable, J. E., Heo, W., Nobre, L., and Kuzniak, S. (2017). Stock market volatility and changes in financial risk tolerance during the great recession. *Journal of Financial Counseling and Planning*, 28(1):140–154.

Schwert, G. W. (2011). Stock volatility during the recent financial crisis. *European Financial Management*, 17(5):789–805.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14.

Wallace, T. D. and Hussain, A. (1969). The use of error components models in combining cross section with time series data. *Econometrica: Journal of the Econometric Society*, 37(1):55–72.

7 Appendix

7.1

In order to avoid any numerical calculation of the inverse of a big covariance matrix, the following closed-form solution of the inverse of the covariance matrix is used in model (19) (Wallace and Hussain, 1969):

$$\Sigma^{-1} = \frac{1}{\sigma_w^2}\left(I_{mT} - \gamma_1\mathbf{A} - \gamma_2\mathbf{B} + \gamma_3 J_{mT}\right), \tag{32}$$

with I_{mT} a $mT \times mT$ identity matrix, J_{mT}, $\mathbf{A}$ and $\mathbf{B}$ as defined in (21) and:

$$\gamma_1 = \frac{\sigma_u^2}{(\sigma_w^2 + T\sigma_u^2)}, \tag{33}$$

$$\gamma_2 = \frac{\sigma_v^2}{(\sigma_w^2 + m\sigma_v^2)}, \tag{34}$$

$$\gamma_3 = \frac{\sigma_u^2\sigma_v^2}{(\sigma_w^2 + T\sigma_u^2)(\sigma_w^2 + m\sigma_v^2)}\left[\frac{2\sigma_w^2 + m\sigma_v^2 + T\sigma_u^2}{\sigma_w^2 + m\sigma_v^2 + T\sigma_u^2}\right], \tag{35}$$

where all variance terms are estimated using the empirical estimates in (22).

7.2

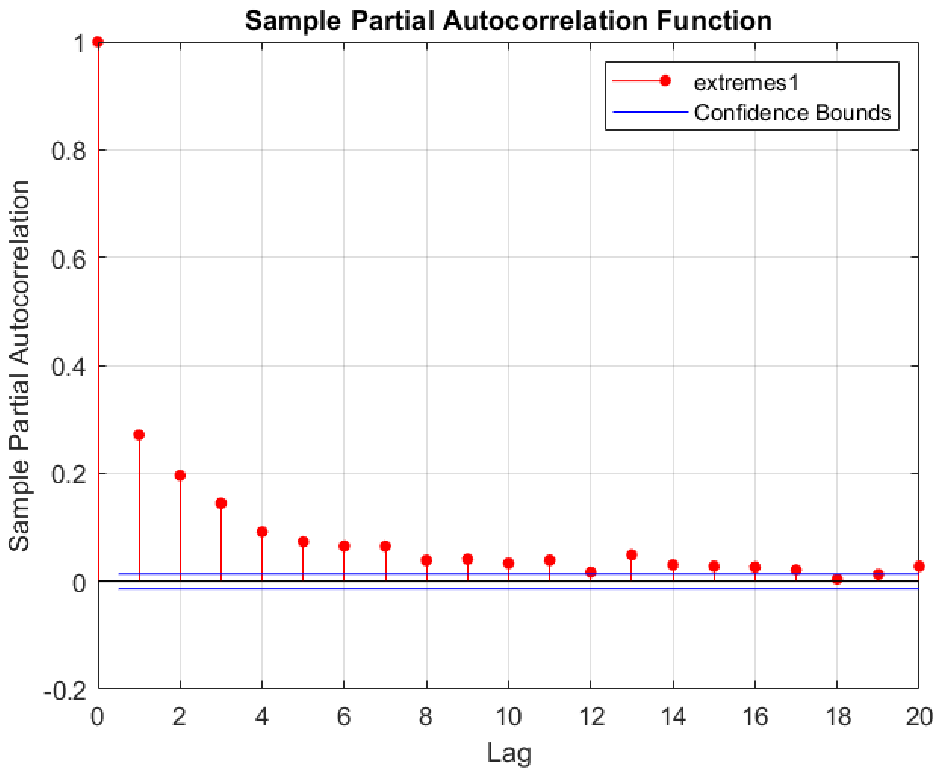


Figure 5: Autocorrelation before the removal of serial dependence. The blue lines represent 95% confidence bounds. If a red line with red dot exceeds these bars, the lag is significant. 19 out of 20 are expected to stay within the blue lines. 2 out of 20 lags are within the 95% confidence bounds in this Figure.

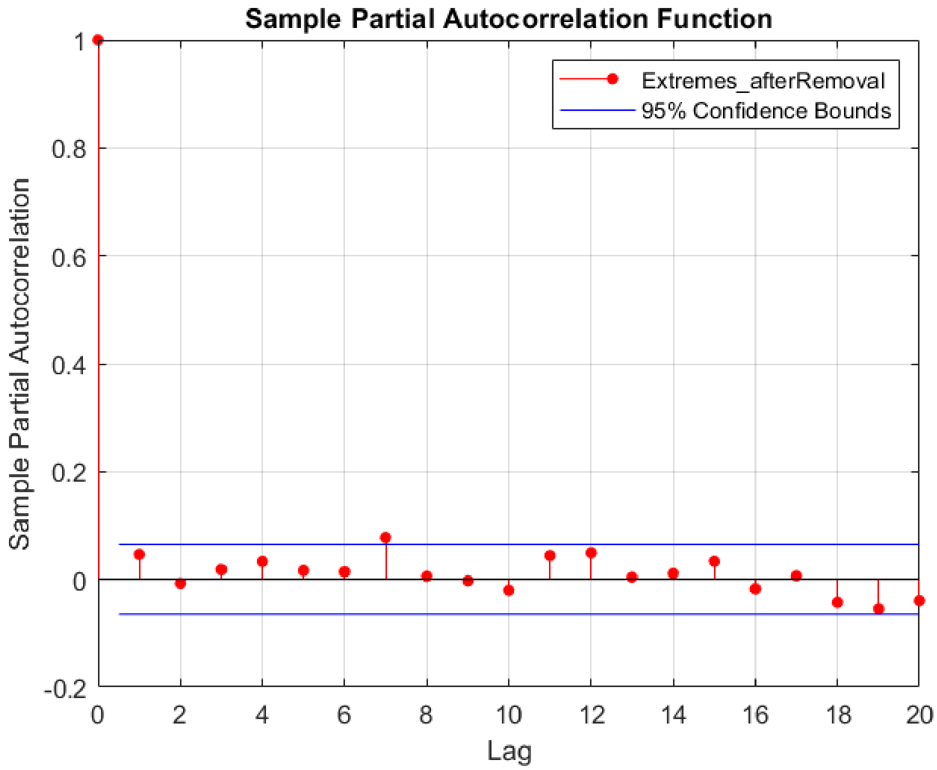


Figure 6: Autocorrelation after the removal of serial dependence. The blue lines represent 95% confidence bounds. If a red line with red dot exceeds these bars, the lag is significant. 19 out of 20 are expected to stay within the blue lines. 19 out of 20 are within the 95% confidence bounds in this Figure. Hence, serial dependence is successfully removed.

7.3

k/n = 0.010	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0311*	0.0301*	0.0301*	0.0294*	0.0294*	0.0294*	0.0304*
	(0.0018)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0018)
BM_Diff	0.0149*	0.0144*	0.0143*	0.0011	0.0011		
	(0.0034)	(0.0034)	(0.0034)	(0.0011)	(0.0011)		
Size_Diff	-0.0037	-0.0033	-0.0034	-0.0011	-0.0011		
	(0.0037)	(0.0037)	(0.0037)	(0.0009)	(0.0009)		
BM_Diff_sq	-0.0004*	-0.0004*	-0.0004*				
	(0.0001)	(0.0001)	(0.0001)				
Size_Diff_sq	0.0001	0.0001	0.0001				
	(0.0001)	(0.0001)	(0.0001)				
BM_Diff_Lag	0.0082*	-0.0005		0		0.0001	0.0072*
	(0.0034)	(0.0011)		(0.0011)		(0.0011)	(0.0034)
Size_Diff_Lag	0.0062	-0.0004		-0.0004		-0.0004	0.007
	(0.0043)	(0.0009)		(0.0009)		(0.0009)	(0.0044)
BM_Diff_Lag_sq	-0.0003*						-0.0002
	(0.0001)						(0.0001)
Size_Diff_Lag_sq	-0.0001						-0.0002
	(0.0001)						(0.0001)

Table 4: Robustness check for $k/n = 0.010$. The estimated $\hat{\beta}$ parameter is given for the constant and explanatory variables created by the two factors in 7 different panel regressions. Book-To-Market ratio and Size are considered as their demeaned yearly log differences. The demeaning is done by taking the average of all yearly log differences of the 50 companies and subtracting the corresponding value from each instance. The standard deviation, the square root of the specific diagonal element of $\hat{\Sigma}_\beta$ in (19), is given within parentheses and an asterisk indicates significance at a level of α = 5%. Note: each column contains a specific panel regression.

k/n = 0.020	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0409*	0.0396*	0.0399*	0.0395*	0.0398*	0.0397*	0.0412*
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)
BM_Diff	0.011*	0.0113*	0.012*	0.0022	0.0022		
	(0.0031)	(0.0031)	(0.0031)	(0.0011)	(0.0011)		
Size_Diff	-0.0003	-0.0001	0.0001	0.0007	0.0007		
	(0.0041)	(0.0042)	(0.0042)	(0.0010)	(0.0010)		
BM_Diff_sq	-0.0003*	-0.0003*	-0.0003*				
	(0.0001)	(0.0001)	(0.0001)				
Size_Diff_sq	0	0	0				
	(0.0001)	(0.0001)	(0.0001)				
BM_Diff_Lag	0.0129*	0.0032*		0.0035*		0.0035*	0.0118*
	(0.0033)	(0.0011)		(0.0011)		(0.0011)	(0.0033)
Size_Diff_Lag	0.0127*	-0.0003*		-0.0002		-0.0002	0.0156*
	(0.0047)	(0.001)		(0.001)		(0.001)	(0.0047)
BM_Diff_Lag_sq	-0.0003*						-0.0002
	(0.0001)						(0.0001)
Size_Diff_Lag_sq	-0.0003*						-0.0003*
	(0.0001)						(0.0001)

Table 5: Robustness check for $k/n = 0.020$. The estimated $\hat{\beta}$ parameter is given for the constant and explanatory variables created by the two factors in 7 different panel regressions. Book-To-Market ratio and Size are considered as their demeaned yearly log differences. The demeaning is done by taking the average of all yearly log differences of the 50 companies and subtracting the corresponding value from each instance. The standard deviation, the square root of the specific diagonal element of $\hat{\Sigma}_\beta$ in (19), is given within parentheses and an asterisk indicates significance at a level of $\alpha = 5\%$. Note: each column contains a specific panel regression.