Analysis of Heavy-Tailed Time Series

XIAOLEI XIE

PhD Thesis

This thesis has been submitted to the PhD School of the Faculty of Science, University of Copenhagen

August 21, 2017

DEPARTMENT OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Xiaolei Xie
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 Copenhagen
Denmark
xie@math.ku.dk
xie.xiaolei@gmail.com

Supervisor:	Prof. Thomas Mikosch University of Copenhagen
Co-supervisor:	Associate Prof. Jeffrey Collamore University of Copenhagen
Co-supervisor:	Prof. Rolf Poulsen University of Copenhagen
Assessment committee:	Associate Prof. Jesper Lund Pedersen (chairman) University of Copenhagen
	Prof. Catalin Stărică University of Neuchatel
	Prof. Olivier Wintenberger University Marie & Pierre Curie, Paris

Abstract

This thesis is about analysis of heavy-tailed time series. We discuss tail properties of real-world equity return series and investigate the possibility that a single tail index is shared by all return series of actively traded equities in a market. Conditions for this hypothesis to be true are identified.

We study the eigenvalues and eigenvectors of sample covariance and sample auto-covariance matrices of multivariate heavy-tailed time series, and particularly for time series with very high dimensions. Asymptotic approximations of the eigenvalues and eigenvectors of such matrices are found and expressed in terms of the parameters of the dependence structure, among others.

Furthermore, we study an importance sampling method for estimating rare-event probabilities of multivariate heavy-tailed time series generated by matrix recursion. We show that the proposed algorithm is efficient in the sense that its relative error remains bounded as the probability of interest tends to zero. We make use of exponential twisting of the transition kernel of an *Markov additive process*, and take advantage of asymptotic theories on products of positive random matrices.

Resumé

Denne afhandling handler om en analyse af tidsrækker med tunge haler. Vi diskuterer haleegenskaber af real-world equity return series og undersøger muligheden for at et enkelt haleindeks deles af alle return seriers af aktivt handlede aktier i et marked. Betingelser for denne hypotese til være sande er identificeret.

Vi studerer egenværdier og egenvektorer af observerede kovarians- og auto-kovarians-matricer af multivariate tung-halede tidsrækker, og især til tidsrækker med meget høje dimensioner. Asymptotiske tilnærmelser af egenværdierne og egenvektorer af sådanne matricer findes og udtrykkes i afhængighedens parameterstruktur.

Vi studerer også en importance sampling metode til estimering sandsynligheder af sjældne begivenheder for multivariate tung-halede tidsrækker genereret af matrice rekursioner. Vi viser, at den foreslåede algoritme er effektiv i den forstand, at dens relative fejl forbliver begrænset, när sandsynligheden af interesse konvergerer til nul. Vi gør brug af eksponentiel vridning af overgangskernen til en *Markov additive process* og tager fordel af asymptotiske teorier om produkter af positive tilfældige matricer.

Acknowledgments

I owe my gratitude to many people and especially to my supervisors Prof. Thomas Mikosch and Prof. Jeffrey Collamore. Their insight and guidance are indispensable throughout the scientific endeavor that culminates in this thesis. Meanwhile, I cannot be more grateful to Prof. Mogens Steffensen – without his financial support from the HIPERFIT program, my PhD project would not have been feasible.

I also feel deeply indebted to my co-authors Prof. Richard A. Davis, Prof. Casper de Vries, Prof. Anja Janßen and Dr. Johannes Heiny. Collaboration with them is not only fruitful but enjoyable. Those moments of relaxed discussions are to be remembered forever. Dr. Johannes Heiny was also my office-mate. I truly enjoyed those casual conversations with him about maths and life in general.

To the professors and PhD students in the Mathematics Department of Copenhagen University and especially to Prof. Jostein Paulsen, I owe a big "thank you" too. I really enjoyed working with him as a teaching assistant and at the same time learned a lot of fantastic maths from him and his book manuscript.

This is also the opportunity for me to thank Dr. Dan Cheng and Dr. Xuan Leng, whom I first met at EPFL Lausanne and later at TU Delft. Discussions with them have been greatly helpful and inspiring to me.

Last but not least, I would like to thank the PhD Committee for spending their precious time to review and assess this thesis.

Xiaolei Xie *Copenhagen, September* 2017.

Summary

This PhD thesis provides results from analysis of heavy-tailed time series. Its main part consists of the following research papers written between October 2014 and August 2017.

[P1] Mikosch, T., de Vries, C. and Xie, X. Do return series have power-law tails with the same index? *Technical Report*

[P2] Collamore, J., Vidyashankar, A. and Xie, X. Rare event simulation for GARCH(p,q) processes. *Technical Report.*

[P3] Janssen, A., Mikosch, T., Mohsen, R. and Xie, X. Limit theory for the singular values of the sample auto-covariance matrix function of multivariate time series. *Bernoulli, to appear.*

[P4] Davis, R. A., Heiny, J., Mikosch, T., and Xie, X. Extreme value analysis for the sample auto-covariance matrices of heavy-tailed multivariate time series. *Extremes 19*, 3 (2016), 517–547. [pdf]

Contents

Chapter 1

Introduction

The phenomenon of heavy-tailedness is widely observed in many disciplines of science, for example, phase transition of matter and black body radiation as studied in physics, neuronal avalanches in biology, claim sizes of insurance mathematics and stock returns in finance. The last application is indeed the focus of this thesis. To discuss the phenomena in precise terms, we introduce the concept of regular variation.

1.1 Regular variation

The concept of regular variation is defined by the following scaling property: if a function f on $(0, \infty)$ satisfies

$$\lim_{x\to\infty} \frac{f(cx)}{f(x)} = c^{\alpha} \quad \forall c > 0,$$

then we say f is regularly varying with index α. f can be written in the form $f(x) = \frac{L(x)}{x^{\alpha}}$, where $L(x)$ is a slowly varying function, i.e. $\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1$, $\forall c > 0$. We call a random variable X regularly varying with index $\alpha \geq 0$ if it satisfies the tail balance condition in the limit $x \to \infty$.

$$\mathbb{P}(X > x) \sim p_{+}\frac{L(x)}{x^{\alpha}}, \quad \mathbb{P}(X < -x) \sim p_{-}\frac{L(x)}{x^{\alpha}}, \quad \text{for some } p_{+}, p_{-} \geq 0,\ p_{+} + p_{-} = 1$$

When expanded to multiple dimensions, the scaling property of regular variation is better described in terms of vague convergence to a spectral measure μ_{α}: if a random vector X satisfies

$$\frac{\mathbb{P}(X/|X| \in \cdot, |X| > cx)}{\mathbb{P}(|X| > x)} \xrightarrow{v} c^{-\alpha}\mu_{\alpha}(\cdot), \quad x \to \infty, \forall c > 0,$$

then we say X is regularly varying with index α. Here μ_{α} is a probability measure on the unit sphere [29]. It is called the spectral measure of X and α is again the tail index. If X is regularly varying with index α, then each component and each linear combination of its components are regularly varying with the same index α. This follows from Feller [61], p. 278. Cf. also Jessen and Mikosch [81], lemma 3.1, and Embrechts et. al. [58], lemma 1.3.1.

Clearly, estimating the tail index α of a sequence $X_1, X_2, \ldots$ of regularly varying variables is particularly important for understanding the behavior of a heavy-tailed series. A standard method proposed for this purpose is due to Hill [71]:

$$\hat{\alpha}_H = \left[\frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{X_{(i)}}{X_{(k+1)}}\right)\right]^{-1}, \tag{1.1}$$

where $X_1, X_2, \ldots, X_n$ is a sample whose tail index is the subject of estimation, and $X_{(i)}$ is the i-th upper order statistic of the sample. Several authors have contributed to showing the weak consistency and asymptotic normality of the estimator $\hat{\alpha}_H$, under the assumptions $k \to \infty, k/n \to 0$ as $n \to \infty$. See Theorem 6.4.6 of Embrechts et. al. [58] for details.

Figure 1.1 shows the Hill estimates of the tail indices of daily stock return series from 3 sectors of the *Standard & Poor's 500* index[1]. The 2.5% and 97.5% quantiles of the asymptotic normal distribution of the estimates are also given. One can see the confidence bands are fairly large compared with the estimated values. This certainly raises the question of how similar they really are and if/how their variations can be explained by economic arguments.

Random variables with regularly varying tails have some very nice features: if X_1 and X_2 are both positive and regularly varying with indices α_1 and α_2, respectively, then $a_1X_1+a_2X_2$, for $a_1, a_2 > 0$, is regularly varying with index $\min\{\alpha_1, \alpha_2\}$ (cf. Mikosch and Jessen [81]). Moreover, if X_1, X_2 are iid, $\mathbb{P}(a_1X_1+a_2X_2 > u) \sim \mathbb{P}(a_1X_1 > u)+\mathbb{P}(a_2X_2 > u)$.

Now consider p return series $X_{i,t},\ i = 1, 2, \ldots, p; t = 1, \ldots, n$. Suppose each of these series is a linear combination of K factors $Y_{j,t},\ j = 1, 2, \ldots, K$, the j-th of which is regularly varying with index α_j. Then by the summation property, each and every $\{X_{i,t}\}$ is regularly varying with index $\min_{1\le j\le K} \alpha_j$. In practice a factor $Y_{i,t}$ is estimated as $\hat{Y}_{j,t} = \sum_{i=1}^{p} a_{j,i}X_{i,t}$, where $(a_{j,1}, a_{j,2}, \ldots, a_{j,p})'$ is the j-th eigenvector of the sample covariance matrix of $\{X_{i,t}\}, i = 1, \ldots, p; t = 1, \ldots, n$, i.e. the eigenvector associated with the j-th largest eigenvalue. For this reason, it is important to understand the eigen system of the sample covariance matrix of $\{X_{i,t}\}$. This topic is addressed in chapters 4 and 5 of this thesis.

When a product of independent positive random variables, say X_1X_2, involves one with regularly varying tails, a useful result is that of Breiman [27]: assume X_1 is regularly varying with index α and there exists $\epsilon > 0$ such that $\mathbb{E}X_2^{\alpha+\epsilon} < \infty$. Then $\mathbb{P}(X_1X_2 > x) \sim \mathbb{E}X_2^{\alpha}\mathbb{P}(X_1 > x)$. More generally, if X_1, X_2 are regularly varying with the same tail index α or if $\mathbb{P}(X_2 > x) = o(\mathbb{P}(X_1 > x))$, then X_1X_2 is regularly varying with index α.

In addition to Breiman's result, the following is also well-known, assuming X_1, X_2 are positive independent random variables and X_1 is regularly varying with index α:

1. if X_2 is regularly varying with index α or $\lim_{x\to\infty} \frac{P(X_2>x)}{\mathbb{P}(X_1>x)} = 0$, then X_1X_2 is regularly varying with index α.
2. if X_1, X_2 are iid and $\mathbb{E}X_1^{\alpha} = \infty$, then $\frac{\mathbb{P}(X_1X_2>x)}{\mathbb{P}(X_1>x)} \to \infty$
3. if X_1, X_2 are iid and $\mathbb{E}X_1^{\alpha} < \infty$, then the only possible limit of $\frac{\mathbb{P}(X_1X_2>x)}{\mathbb{P}(X_1>x)}$ is $2\mathbb{E}X_1^{\alpha}$.

For an extensive summary of the regular variation properties of functions of regularly varying random variables, see Mikosch and Jessen [81].

The notion of a regularly varying strictly stationary process is also of interest. Originally introduced by Davis and Hsing [39] for univariate processes, the concept was extended to multivariate processes by Davis and Mikosch [42]: an $\mathbb{R}^d$-valued strictly stationary process X_t is said to be regularly varying with index α, if for each $h \ge 0$,

$$\frac{\mathbb{P}\left[x^{-1}(X_0, X_1, \ldots, X_h) \in \cdot\right]}{\mathbb{P}(|X_0| > x)} \xrightarrow{v} \nu_{\alpha}^{(h)}(\cdot), \quad x \to \infty$$

[1] An American stock index comprising around 500 companies.

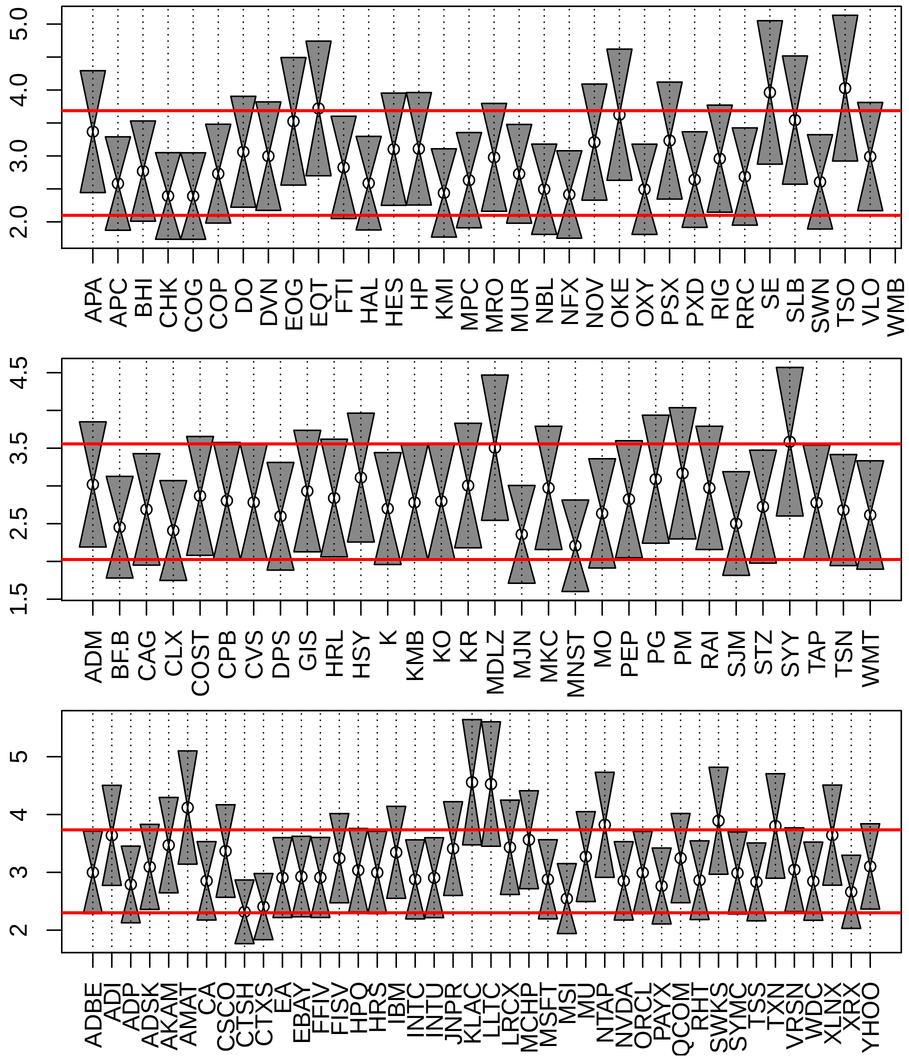


Figure 1.1: Hill estimates $\hat{\alpha}_{50}$ of the lower tail-indices α of daily return series in sectors of the S&P 500 index. The data span from 1 January 2010 to 31 December 2014 and comprise $n = 1304$ observations. The graphs from top to bottom correspond to the "Energy", "Consumer Staples" and "Information Technology" sectors. Each circle corresponds to a Hill estimate $\hat{\alpha}_{50}$; the gray triangles above and below it mark the 97.5% and 2.5% quantiles of its approximate normal distribution; see (2.4) and the discussion following it for an interpretation. The lower and upper red lines mark the medians of the 2.5% and 97.5% quantiles, respectively, evaluated from all stocks in the sector. The data are taken from *Yahoo Finance*; the labels on the horizontal axes are Yahoo symbols of the stocks.

where $\nu_\alpha^{(h)}$ a non-null Radon measure on $\mathbb{R}^{d(h+1)} \setminus \{0\}$ that is homogeneous of order $-\alpha$, i.e. $\forall c > 0, \forall S \subset \mathbb{R}^{d(h+1)} \setminus \{0\}, \nu_\alpha^{(h)}(cS) = c^{-\alpha}\nu_\alpha^{(h)}(S)$. $|\cdot|$ could be any given vector norm.

Basrak and Segers [15] proved an equivalence condition for regular variation of a strictly stationary process: an $\mathbb{R}^d$-valued strictly stationary process $\{X_t\}$ is regularly varying with index α if and only if there exists an $\mathbb{R}^d$-valued sequence $\{\Theta_i\}_{i=0,1,\dots}$ and a Pareto(α) random variable Z, i.e. $\mathbb{P}(Z > x) = x^{-\alpha}, \forall x \geq 1$, independent of $\{\Theta_i\}$

$$\mathbb{P}\left[x^{-1}(X_0, X_1, \dots, X_h) \in \cdot \,\middle|\, |X_0| > x\right] \xrightarrow{w} \mathbb{P}(Z(\Theta_1, \dots, \Theta_h) \in \cdot), \quad x \to \infty, \forall h \geq 0$$

The sequences $\{\Theta_i\}_{i=0,1,\dots}$ and $\{Z\Theta_i\}_{i=0,1,\dots}$ are called the spectral tail process and the tail process of $\{X_t\}$, respectively.

A recent remarkable result from Mikosch and Wintenberger [90] regarding strictly stationary regularly varying processes is quoted below:

Theorem 1.1. *Let (Y_t) be an $\mathbb{R}^d$-valued strictly stationary sequence , $S_n = Y_1 + \cdots + Y_n$ and (a_n) be such that $n\,\mathbb{P}(|Y| > a_n) \to 1$. Also write for $\varepsilon > 0$, $\overline{Y}_t = Y_t \mathbf{1}(|Y_t| \leq \varepsilon a_n)$, $\underline{Y}_t = Y_t - \overline{Y}_t$ and*

$$\overline{S}_{l,n} = \sum_{t=1}^{l} \overline{Y}_t \qquad \underline{S}_{l,n} = \sum_{t=1}^{l} \underline{Y}_t\,.$$

Assume the following conditions:

1. *(Y_t) is regularly varying with index $\alpha \in (0,2) \setminus \{1\}$ and spectral tail process (Θ_j).*
2. *A mixing condition holds: there exists an integer sequence $m_n \to \infty$ $k_n = [n/m_n] \to \infty$ and*
$$\mathbb{E}\mathrm{e}^{\,it'\underline{S}_n/a_n} - \left(\mathbb{E}\mathrm{e}^{\,it'\underline{S}_{m_n,n}/a_n}\right)^{k_n} \to 0\,, \qquad n \to \infty\,, \qquad t \in \mathbb{R}^d\,. \tag{1.2}$$
3. *An anti-clustering condition holds:*
$$\lim_{l\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\max_{t=l,\dots,m_n} |Y_t| > \delta a_n \mid |Y_0| > \delta a_n\big) = 0\,, \qquad \delta > 0 \tag{1.3}$$
for the same sequence (m_n) as in 2.
4. *If $\alpha \in (1,2)$, in addition $\mathbb{E}[Y] = 0$ and the vanishing small values condition holds:*
$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\big(a_n^{-1}|\overline{S}_n - \mathbb{E}[\overline{S}_n]| > \delta\big) = 0\,, \qquad \delta > 0 \tag{1.4}$$
and $\sum_{i=1}^{\infty} \mathbb{E}[|\Theta_i|] < \infty$.

Then $a_n^{-1} S_n \xrightarrow{\mathrm{d}} \xi_\alpha$ for an α-stable $\mathbb{R}^d$-valued vector ξ_α with log characteristic function

$$\int_0^\infty \mathbb{E}\left[\exp\left(i\,y\,t' \sum_{j=0}^{\infty} \Theta_j\right) - \exp\left(i\,y\,t' \sum_{j=1}^{\infty} \Theta_j\right) - i\,y\,t'\mathbf{1}_{(1,2)}(\alpha)\right] d(-y^\alpha)\,, \qquad t \in \mathbb{R}^d\,. \tag{1.5}$$

Remark 1.2. *If we additionally assume that Y is symmetric, which implies $\mathbb{E}[\overline{Y}] = \mathbf{0}$, then the statement of the theorem also holds for $\alpha = 1$.*

Using theorem 1.1 we prove in chapter 4 joint convergence in distribution of eigenvalues of a sample covariance matrix of stochastic volatility processes (see §1.4) to $\alpha/2$-stable random variables.

1.2 Stochastic recurrence equation

One of the most important dynamical mechanisms that lead to regularly varying random vectors is a stochastic recursion of the following form:

$$X_t = A_t X_{t-1} + B_t, \quad t \in \mathbb{Z} \tag{1.6}$$

where X_t is a d-dimensional random vector, A_t is a $d \times d$ random matrix and B_t is a d-dimensional vector, random or deterministic. The sequence $\{(A_t, B_t)\}_{t\in\mathbb{Z}}$ is iid. The stationary solution to (1.6) satisfies the fixed-point equation $X \stackrel{d}{=} AX + B$, where X and (A, B) are generic elements of the $\{X_t\}_{t\in\mathbb{Z}}$ and $\{(A_t, B_t)\}_{t\in\mathbb{Z}}$ sequences.

Kesten [84] showed that, when A_t is almost surely non-negative, has no row or column of only zeros, and B_t is almost surely non-negative and is not equal to the null vector with probability 1, then the solution X to the equation $X \stackrel{d}{=} AX + B$ is regularly varying with a positive index α, assuming the following conditions (M) and (A):

- Condition (M)
  1. The top Lyapunov exponent
     $$\gamma = \inf_{n\geq 1} \frac{1}{n} \mathbb{E} \log \|A_n \cdots A_1\|$$
     is negative.
  2. There exists $\alpha > 0$ such that
     $$1 = \lambda(\alpha) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\|A_n \cdots A_1\|^\alpha$$
  3. $\mathbb{E}(\|A_1\|^\alpha \log^+ \|A_1\|) < \infty$
  4. $\mathbb{E}|B_1|^\alpha < \infty$
- Condition (A) : the group generated by
  $$\{\log \rho(s) : s = A_n \cdots A_1 \text{ for some } n \geq 1\}$$
  is dense in $\mathbb{R}$, where $\rho(s)$ denotes the spectral radius of matrix s.

Upon these conditions, Kesten's theorem gives

$$u^\alpha \mathbb{P}(u^{-1}X \in \cdot) \xrightarrow{v} \nu_\alpha^{(h)}(\cdot) \tag{1.7}$$

where $\nu_\alpha^{(h)}$ is a non-null Radon measure on $\mathbb{R}_+^d \setminus \{0\}$ with the property $\forall a > 0, \nu_\alpha^{(h)}(aA) = a^{-\alpha}\nu_\alpha^{(h)}(A)$.

In addition to non-negative matrices, two other classes of random matrices have been shown to lead to power-law tails via the recurrence relation (3.2). Alsmeyer and Mentemeier [3] considered invertible matrices whose distribution has a density. Let $M(d, \mathbb{R})$ denote a metric space of $d \times d$ matrices with real entries that are invertible with probability 1. They replaced Kesten's condition $\mathbb{E}(\|A\|^\alpha \log^+ \|A\|) < \infty$ with a stronger counterpart $\mathbb{E}[\|A\|^\alpha(\log^+ \|A\| + \log \|A^{-1}\|)] < \infty$, and lifted the condition (A). In addition, they assumed

1. For any open set $U \subset \mathbb{S}^{d-1}$ and any $u \in \mathbb{S}^{d-1}$, $\exists n \geq 1$ such that

$$\mathbb{P}\left(\frac{\prod_{i=1} A_i u}{|\prod_{i=1} A_i u|} \in U\right) > 0.$$

2. There exist $N \geq 1$, $c, \epsilon > 0$ and an invertible matrix $\bar{A} \in M(d, \mathbb{R})$ such that for any set $C \subset M(d, \mathbb{R})$, it holds true $\mathbb{P}(A_N \cdots A_1 \in C) \geq c|B_\epsilon(\bar{A}) \cap C|$, where $|\cdot|$ denotes the Lebesgue measure and $B_\epsilon(\bar{A})$ is the ball with radius ϵ centered at $\bar{A}$.

These assumptions are termed conditions (id). Furthermore, they assumed that there was no point in $\mathbb{R}^d$ such that the recurrence equation (3.2) was stuck at this point with probability 1: $\mathbb{P}(Ax + B = x) < 1$ for all $x \in \mathbb{R}^d$. With these assumptions, they showed

$$\lim_{u \to \infty} u^\alpha \mathbb{P}(\langle x, X \rangle > u) = e_\alpha(x)$$

where $x \in \mathbb{S}^{d-1}$ and $e_\alpha(\cdot)$ is a continuous function on $\mathbb{S}^{d-1}$.

The second of the (id) conditions, which is satisfied when the distribution of A has a Lebesgue density, can actually be lifted if stronger moment conditions are imposed on A and B, and in addition, a proximity condition is satisfied by the support of A. This is the result of Guivarc'h and Le Page [65]. Let G_A denote the semi-group generated by $\{\Pi_n : \Pi_n = A_n \cdots A_1, A_i \in M(d, \mathbb{R})\}$. The authors assumed

1. There is no finite union W of proper sub-spaces of $\mathbb{R}^d$ that satisfies $\forall a \in G_A, aW = W$.

2. G_A contains a proximal element, i.e. an element a whose largest singular value is an algebraically simple eigenvalue of a.

These two assumptions are termed (ip) conditions. Replacing the (id) conditions of Alsmeyer and Mentemeier with (ip) and the moment conditions of the former with

$$\mathbb{E}(\|A\|^{\alpha+\delta}) < \infty, \quad \mathbb{E}(\|A\|^\alpha \|A^{-1}\|^\delta) < \infty, \quad \mathbb{E}(|B|^{\alpha+\delta}) < \infty \quad \text{for some } \delta > 0,$$

Guivarc'h and Le Page proved the vague convergence result of (1.7).

1.3 GARCH models

Introduced by Bollerslev [24] in 1986, *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) models have been hugely popular for modeling volatility of financial time series and have inspired numerous variants. A GARCH(p, q) model is a stationary process $\{X_t\}_{t \in \mathbb{Z}}$ satisfying

$$X_t = \sigma_t Z_t, \tag{1.8}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1.9}$$

where $\{X_t\}$ is a model for return series of stock prices, foreign exchange rates, interest rates, etc; $\{Z_t\}$ is an iid, mean 0, unit-variance sequence, σ_t^2 is the variance of the distribution of X_t conditional on $\{(X_i,\sigma_i^2)\}_{i<t}$; ω, $\{\alpha_i\}_{i=1}^p$, $\{\beta_i\}_{i=1}^q$ are non-negative parameters of the model. Written in matrix form, as shown in (3.13), the GARCH(p,q) recurrence equation is of the form of (1.6). With appropriate conditions,

$$(\sigma_t^2,\dots,\sigma_{t-q+1}^2,X_{t-1}^2,\dots,X_{t-p+1}^2)$$

is shown to be a positive Harris recurrent Markov chain (cf. Bollerslev [24] and Buraczewski et al. [29]), whose stationary distribution has regularly varying tails. The tail index α is given by

$$\Lambda(\alpha)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\|A_n\cdots A_1\|^\alpha=0, \tag{1.10}$$

where $\{A_i\}_{i\in\mathbb{Z}}$ are iid matrices whose entries are functions of $\{\alpha_i\}_{i=1}^p$, $\{\beta_i\}_{i=1}^q$ and $\{Z_t^2\}$:

$$A_t=\begin{pmatrix}
\alpha_1Z_{t-1}^2+\beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p\\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0\\
Z_{t-1}^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0\\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0\\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix} \tag{1.11}$$

While GARCH models have been very successful for modeling financial time series, they do have their drawbacks. For example, the tail index is very sensitive to the model parameters $\{\alpha_i\}_{i=1}^p$ and $\{\beta_i\}_{i=1}^q$. In applications, these parameters need to be estimated from a sample and are always uncertain to some extent. For this reason, there can be a significant discrepancy between the tail index obtained via (1.10) with the model parameters substituted for their sample estimates and the Hill estimate (1.1).

There exist various extensions of the univariate GARCH model to the multivariate case. The most notable one is perhaps the *constant conditional correlation* (CCC) model of Bollerslev [25] and Jeantheau [80]. In the bivariate case, CCC is the model

$$\mathbf{X}_t=\begin{pmatrix}X_{1,t}\\X_{2,t}\end{pmatrix}=\begin{pmatrix}\sigma_{1,t}&0\\0&\sigma_{2,t}\end{pmatrix}\begin{pmatrix}Z_{1,t}\\Z_{2,t}\end{pmatrix}=\Sigma_t\,\mathbf{Z}_t\,,\qquad t\in\mathbb{Z}\,.$$

Thus both return components $X_{i,t}$ have the form of a univariate stochastic volatility model $X_{i,t}=\sigma_{i,t}Z_{i,t}$ with non-negative volatility $\sigma_{i,t}$ and an iid bivariate noise sequence $(\mathbf{Z}_t)$ with zero mean and unit variance components. We also have the specification

$$\begin{aligned}
\mathbf{Y}_t=\begin{pmatrix}\sigma_{1,t}^2\\\sigma_{2,t}^2\end{pmatrix} &= \begin{pmatrix}\alpha_{01}\\\alpha_{02}\end{pmatrix}+\begin{pmatrix}\alpha_{11}&\alpha_{12}\\\alpha_{21}&\alpha_{22}\end{pmatrix}\begin{pmatrix}X_{1,t-1}^2\\X_{2,t-1}^2\end{pmatrix}+\begin{pmatrix}\beta_{11}&\beta_{12}\\\beta_{21}&\beta_{22}\end{pmatrix}\begin{pmatrix}\sigma_{1,t-1}^2\\\sigma_{2,t-1}^2\end{pmatrix}\\
&= \begin{pmatrix}\alpha_{01}\\\alpha_{02}\end{pmatrix}+\begin{pmatrix}\alpha_{11}Z_{1,t-1}^2+\beta_{11}&\alpha_{12}Z_{2,t-1}^2+\beta_{12}\\\alpha_{21}Z_{1,t-1}^2+\beta_{21}&\alpha_{22}Z_{2,t-1}^2+\beta_{22}\end{pmatrix}\begin{pmatrix}\sigma_{1,t-1}^2\\\sigma_{2,t-1}^2\end{pmatrix},
\end{aligned} \tag{1.12}$$

for positive α_{0i} and suitable non-negative α_{ij},β_{ij}, $i,j=1,2$. Writing

$$\mathbf{B}_t=\begin{pmatrix}\alpha_{01}\\\alpha_{02}\end{pmatrix}\quad\text{and}\quad\mathbf{A}_t=\begin{pmatrix}\alpha_{11}Z_{1,t-1}^2+\beta_{11}&\alpha_{12}Z_{2,t-1}^2+\beta_{12}\\\alpha_{21}Z_{1,t-1}^2+\beta_{21}&\alpha_{22}Z_{2,t-1}^2+\beta_{22}\end{pmatrix},$$

we see that we are again in the framework of a stochastic recurrence equation but this time for vector-valued $\mathbf{B}_t$ and matrix-valued $\mathbf{A}_t$:

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t , \qquad t \in \mathbb{Z} . \tag{1.13}$$

Kesten [84] also provided the corresponding theory for stationarity and tails in this case. Stărică [107] dealt with the corresponding problems for CCC-GARCH processes, making use of the theory in Kesten [84], Bougerol and Picard [26] and its specification to the tails of GARCH models in Basrak et al. [14]. Stărică [107] assumed the Kesten conditions for the matrices $\mathbf{A}_t$. These conditions ensure that the product matrices $\mathbf{A}_1 \cdots \mathbf{A}_n$ have positive entries for sufficiently large n. Then Kesten's theory implies that all components of the vector $\mathbf{X}_t$ have power-law tails with the same index α and also that the finite-dimensional distributions of the process $(\mathbf{X}_t)$ are regularly varying with index α.

Various GARCH modifications are derived by considering linear combinations of CCC-GARCH models. The property of multivariate regular variation of multivariate GARCH ensures that, after linear transformations, the new process in all components has again power-law tails with the same index as the original GARCH process; see Basrak et al. [14]. Models which are constructed in this way are the Orthogonal GARCH model of Alexander and Chibumba [2], its generalization GO-GARCH by van der Weide [113], the Full Factor GARCH model of Vrontos et al. [116] and the Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen [87]. These models are characterized by their treatment of each series as a linear combination of factors, and each of the factors is modeled as a GARCH process; see Silvennoinen and Teräsvirtä [104].

Not all choices of α- and β-parameters in the model (1.12) allow for an application of the Kesten theory. For example, assume that only the diagonal elements α_{ii} and β_{ii} are positive. Then $\mathbf{A}_t$ is diagonal and, hence, the condition that $\mathbf{A}_1 \cdots \mathbf{A}_n$ have positive entries for sufficiently large n cannot be satisfied. In the latter situation, both $(X_{1,t})$ and $(X_{2,t})$ are univariate GARCH processes. Assuming the conditions of the univariate Kesten-Goldie theorem for each component process, $(X_{1,t})$ and $(X_{2,t})$ have power-law tails with indices α_1 and α_2, respectively, given by the solutions to the equations $\mathbb{E}[(\alpha_{ii} Z_{i,t}^2 + \beta_{ii})^{\alpha_i/2}] = 1$, $i = 1, 2$. In this model, one can introduce dependence between the two component series $(X_{1,t})$ and $(X_{2,t})$ by assuming dependence between the noise variables $Z_{1,t}$ and $Z_{2,t}$. Another situation when the Kesten theory fails appears when $\mathbf{A}_t$ is an upper or lower triangle matrix: then the products $\mathbf{A}_1 \cdots \mathbf{A}_n$ are always of the same triangular type. Similar remarks apply when one considers a CCC model in general dimension. Of course, one may argue that the latter models are not natural: they are degenerate since they do not allow for a linear relationship between all squared volatilities on a given day.

1.4 Stochastic volatility models

With the availability of high-frequency data, a different approach than that of GARCH has been popularized for modeling volatility of financial time series and has led to greatly improved accuracy of prediction. This is the approach of stochastic volatility models.

In the pioneering work of Clark [33], the author modeled the logarithmic price Y_t as a subordinated stochastic process: $Y_t = V_{\tau_t}, t \geq 0$, where V_i is a Brownian motion. τ_t is a real-valued, non-negative, non-decreasing sequence with $\tau_0 = 0$. It models a time change. As pointed out by Shephard and Andersen [101], the log-price process $\{Y_t\}$ is serially uncorrelated although potentially dependent, provided that V_t and τ_t are independent.

Later authors, e.g. Back [8] chose to model the log-price process as a semi-martingale, with increments of the martingale component modeled as a product process:

$$Y_t = Y_0 + A_t + M_t,$$

where $\{A_t\}$ is a finite-variation process and $\{M_t\}$ is a martingale and hence $\{Y_t\}$ is a semi-martingale. $\{\sigma_t\}$ is non-negative and $\{Z_t\}$ is an iid process with zero mean and unit variance. $\{\sigma_t\}$ and $\{Z_t\}$ are independent of each other.

For discrete time models, Taylor [110] was the first to propose a product process for modeling the martingale part M_t of Y_t:

$$M_t - M_{t-1} = \sigma_t Z_t, \quad t \in \mathbb{Z}$$

A convenient choice of σ_t is

$$\log \sigma_t = \sum_{l \in \mathbb{Z}} \psi_l \eta_{t-l}, \quad t \in \mathbb{Z} \tag{1.14}$$

where $\{\psi_l\}_{l\in\mathbb{Z}}$ is a sequence of real numbers with at least one non-zero element, $\{\eta_t\}_{t\in\mathbb{Z}}$ is an iid sequence of random variables with zero-mean and finite variance. By Kolmogorov's 3-series theorem, the infinite series above converges if and only if $\sum_{l\in\mathbb{Z}} \psi_l^2 < \infty$, and in this case $\{\sigma_t\}$ is stationary. In particular, if $\{\eta_i\}_{i\in\mathbb{Z}}$ is normally distributed with zero mean, $\{\sigma_t\}_{t\in\mathbb{Z}}$ has log-normal marginal distributions.

If, however, the sequences $\{Z_t\}_{t\in\mathbb{Z}}$ or $\{\sigma_t\}_{t\in\mathbb{Z}}$ are regularly varying with index α and some additional conditions are satisfied, $\{X_t\}$ is also regularly varying with the same index. Specifically, if $\{Z_t\}$ is regularly varying and $\{\sigma_t\}$ has a lighter tail, the conclusion follows from Breiman's lemma [27]. See §4.3 of Janßen et al. [79], i.e. chapter 4 of this thesis for more details.

There are a few advantages in using stochastic volatility models. They are among the simplest models allowing for conditional heteroscedasticity (cf. Andersen et al. [4]); nevertheless, they greatly improve the accuracy of predicting future volatilities by taking advantage of high frequency data. Specifically, for the semi-martingale $\{Y_t\}$ we have

$$[Y_\delta]_t = \sum_{j=1}^{\lfloor t/\delta \rfloor} (Y_{j\delta} - Y_{(j-1)\delta})^2 \xrightarrow{P} [Y]_t, \quad \delta \to 0. \tag{1.15}$$

The object $[Y_\delta]_t$ is called the realized quadratic variation in time series literature. The convergence in probability follows directly from the definition of quadratic variation $[Y]_t$:

$$[Y]_t = \lim_{\Delta t \to 0} \sum_{i=1}^{n} (Y_{t_i} - Y_{t_{i-1}})^2$$

where $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$ and $\Delta t = \sup_{1\le i\le n}(t_i - t_{i-1})$. Furthermore, Jacod [76] and Barndorff-Nielsen and Shephard [12] proved a central limit theorem:

$$\frac{[Y_\delta]_t - [Y]_t}{\sqrt{2\delta \int_0^t \sigma_s^2 ds}} \xrightarrow{d} N(0,1), \quad \delta \to 0.$$

If the process of expected return has continuous sample path and M_t is a stochastic volatility process, it has been shown that $[Y]_t = [M]_t$ and $[Y_\delta]_t \xrightarrow{P} \int_0^t \sigma_s^2 ds$. Meanwhile, Itô's formula gives, for a semi-martingale $\{Y_t\}$,

$$\begin{aligned} Y_t^2 &= [Y]_t + 2\int_0^t Y_s dY_s \\ &= [Y]_t + 2\int_0^t Y_s dA_s + 2\int_0^t Y_s dM_s, \\ \mathbb{E}Y_t^2 &= \mathbb{E}[Y]_t + 2\mathbb{E}\left(\int_0^t Y_s dA_s\right) \\ &\approx \mathbb{E}[Y]_t, \quad \text{when } t \text{ is small.} \end{aligned} \tag{1.16}$$

(1.16) and (1.15) show that $\mathbb{E}Y_t^2$, or in other words, the forecast of future squared return, can be obtained as $\mathbb{E}[Y_\delta]_t$, i.e. the forecast of future realized quadratic variation.

1.5 Contributions of this thesis

In this section we summarize our results from the research papers.

1.5.1 Tail parameters of equity return series

In chapter 2 we consider a minimal market where a riskless bond and an equity are the only assets available to investors. We model the investor's preference of the equity with *Generalized Disappointment Aversion (GDA)*, an idea envisaged by Gul [66] and generalized by Routledge and Zin [99]. Specifically, in the GDA theory a rational investor's behavior is characterized by his attempt to maximize the GDA functional $U(F)$:

$$U(F) = \mathbb{E}_F[u(C)] - b\mathbb{E}_F[u(\delta) - u(C)\mathbf{1}_{\{C<\delta\}}]$$

where C is the investor's wealth evaluated at a fixed date, δ is the level of disappointment – if C falls below δ, the investor becomes disappointed. b parametrizes the growth of his disappointment. F is the distribution function of the return of the investor's portfolio. In the aforementioned minimal market, F is the distribution function of the equity's return. The subscript F of $\mathbb{E}_F$ reminds us that the expectation is taken with respect to the distribution function F.

We have established that, in the case of an equity return series with two-sided, functionally independent Pareto tails, GDA preference functionals are monotone increasing/decreasing with the tail index/scale parameters. Thus in a market dominated by such equities, the investors would pursue the largest tail index in the market, leading to a shared common tail index for all equities.

The empirical results presented in section 2.2 suggest this may well be the case for the "Consumer Staples" sector of S&P 500, given the Hill estimates of tail indices shown in figure 1.1 and the largely positive results of tests for equal tail indices shown in figure 2.3.

On the other hand, we have also seen that, when the left and the right tails have the same indices, investor preference over the equity has more sophisticated variations in the parameter space including the tail parameters of the equity, the interest rate, the investor's risk appetite as captured by his utility function, and his threshold of disappointment.

We also acknowledge that our model of the market and the investor is a simple one, not accounting for the dependence between equities, nor the categorization of investors and their interactions. These are potential topics of future work.

1.5.2 Importance sampling estimator of GARCH(p,q) rare event probability

In §1.2 we have seen how power-law tails can arise e.g. by Kesten's theorem in the stationary distribution of a Markov chain $\{V_t\}_{t\geq 0}$ described by a stochastic recurrence equation. In §1.3 we have introduced GARCH(p,q) processes as examples of such Markov chains. By Kesten's theorem, the stationary distribution of a GARCH(p,q) process has power-law tails asymptotically, i.e. $\mathbb{P}(|V|>x)=cx^{-\alpha}+o(x^{-\alpha}), c>0, \alpha>0$ as $x\to\infty$. While a nice result, this theorem does not allow us to compute, in precision, the probability $\mathbb{P}(|V|>x)$. A numerical procedure is needed for this purpose. A straightforward approach is of course direct Monte Carlo: we simulate the first n steps of $\{V_t\}_{t\geq 0}$ and approximate $\mathbb{P}(|V|>x)$ as

$$\mathbb{P}(|V|>x)\approx\frac{1}{n-K}\sum_{t=n-K+1}^{n}\mathbf{1}_{\{|V_t|>x\}} \tag{1.17}$$

where we discard the first K steps of the simulated sample path so that the empirical distribution of $\{V_t\}_{t\geq n-K+1}$ is closer to the stationary distribution π of $\{V_t\}$.

The difficulty of this naive approach is that, when x is large, the event $\{|V|>x\}$ happens very rarely, making the variance of the estimate too big to be of any use. A method to overcome this difficulty is importance sampling: we introduce a Markov additive process $\{(Y_t,S_t)\}_{t\geq 0}, |Y_t|=1$ on the unit sphere:

$$\begin{aligned}
B &= (\omega,0,\cdots,0)',\\
V_t &= (\sigma_t^2,\sigma_{t-1}^2,\cdots,\sigma_{t-q+1}^2,X_{t-1}^2,\cdots,X_{t-p+1}^2)'=A_tV_{t-1}+B,\\
Y_t &= \frac{A_t\cdots A_1V_0}{|A_t\cdots A_1V_0|}=\frac{A_tY_{t-1}}{|A_tY_{t-1}|},\\
l_t &= \log|A_tY_{t-1}|,\\
S_t &= \log|A_t\cdots A_1V_0|=\sum_{i=1}^{t}l_i+\log|V_0|,
\end{aligned}$$

where $X_t=\sigma_tZ_t$ (see (1.8)). For convergence we also use the notation

$$A_t\cdot Y_{t-1}:=\frac{A_tY_{t-1}}{|A_tY_{t-1}|}.$$

The matrices $\{A_t\}_{t\geq 1}$ are defined by (1.11). For more details, see §3.2. V_t, our object of main interest, is a function of the path $\{(Y_i,l_i)\}_{1\leq i\leq t}$. To increase the chance of observing $\{|V_t|>x\}$, we adopt a dual change of the transition kernel of $\{(Y_t,l_t)\}_{t\geq 0}$. Before the first occurrence of $\{|V_t|>x\}$ we simulate (Y_t,l_t) according to a shifted transition kernel, whose induced probability measure is denoted $\mathbb{P}^\alpha$:

$$\begin{aligned}
&\mathbb{P}^\alpha\left[(Y_t,l_t)\in dy\times dl|Y_{t-1}=w\right]\\
=\ &e^{\alpha l}\frac{r_\alpha(y)}{r_\alpha(w)}\mathbb{P}\left[(Y_t,l_t)\in dy\times dl|Y_{t-1}=w\right],
\end{aligned} \tag{1.18}$$

where $\mathbb{P}$ denotes the probability with respect to the probability measure π, the stationary probability measure of $\{V_t\}$. r_α is a right eigenfunction of the operator $\mathscr{P}^\alpha$, which is defined by its action on a test function $g : \mathbb{S}^{d-1} \to \mathbb{R}_+$:

$$\mathscr{P}^\alpha g(x) = \int_{\mathrm{dom}(Z)} |A(z^2)x|^\alpha g(A(z^2) \cdot x) f_Z(z) dz.$$

r_α is the right eigenfunction of $\mathscr{P}^\alpha$ associated with the eigenvalue $\lambda(\alpha) = e^{\Lambda(\alpha)} = 1$, i.e. $\mathscr{P}^\alpha r_\alpha(x) = r_\alpha(x)$. The function $\Lambda(\alpha)$ is defined by (1.10), f_Z is the density function of Z and $A(z^2)$ is the matrix (1.11) with Z^2_{t-1} substituted for z^2.

Conditional on (Y_{t-1}, l_{t-1}), the only source of randomness to Y_t comes from Z_{t-1}. Hence the shift of conditional probability distribution shown in (1.18) is equivalent to shifting the distribution of Z_{t-1}. We have

$$\mathbb{P}^\alpha(Z_{t-1} \in dz | Y_{t-1} = w) = |A(z^2)w|^\alpha \frac{r_\alpha(A(z^2) \cdot w)}{r_\alpha(w)} \mathbb{P}(Z_{t-1} \in dz).$$

Note that $\{Z_t\}$ is an iid sequence in the original measure. An expected value with respect to $\mathbb{P}^\alpha$ is related to its counterpart with respect to $\mathbb{P}$ via

$$\mathbb{E}[g(Y_t, l_t)] = \mathbb{E}^\alpha \left[g(Y_t, l_t) e^{-\alpha l_t} \frac{r_\alpha(Y_{t-1})}{r_\alpha(Y_t)} \right]. \tag{1.19}$$

Let us define $T_x = \min\{t \geq 1 : |V_t| > x\}$. Once the first excursion of $|V_t|$ above x has occurred, i.e. $t > T_x$, we change the transition kernel back to the original and continue the simulation until the process returns to a designated set $\mathcal{C} = \{v : |v| \leq M\}$, where the positive number M is chosen in accordance with the function Λ. We denote the successive times of $\{V_t\}$ returning to the set $\mathcal{C}$ as $0 = K_0 < K_1 < K_2 < \cdots$. It can be shown that $\{(V_{K_{m+1}}, \sum_{i=K_m}^{K_{m+1}-1} \mathbf{1}_{\{|V_i|>x\}})\}_{m \geq 0}$ is a positive Harris recurrent Markov chain for all $x \geq 0$. Let $N_x = \sum_{i=0}^{K_1-1} \mathbf{1}_{\{|V_i|>x\}}$. We show by the law of large numbers for Markov chains

$$\mathbb{P}(|V| > x) = \pi(\mathcal{C}) \mathbb{E}_\eta(N_x)$$

where η is π restricted to the set $\mathcal{C}$, i.e. $\forall S \subseteq \mathcal{C}, \eta(S) = \pi(S)/\pi(\mathcal{C})$. $\mathbb{E}_\eta$ means the expectation is taken only on condition $V_0 \sim \eta$. Finally by (1.19),

$$\mathbb{P}(|V| > x) = \pi(\mathcal{C}) \mathbb{E}_\eta N_x = \mathbb{E}^{\mathcal{D}}_\eta \mathcal{E}_x,$$

where

$$\mathcal{E}_x = \pi(\mathcal{C}) N_x |A_{T_x} \cdots A_1 V_0|^{-\alpha} \frac{r_\alpha(Y_0)}{r_\alpha(Y_{T_x})} \mathbf{1}_{\{T_x < K_1\}}$$

is our importance sampling estimator. $\mathbb{E}^{\mathcal{D}}_\eta$, $\mathcal{D}$ for "dual", is to remind us that the expectation is taken with respect to the shifted transition kernel as given by (1.19), and then with respect to the original transition kernel. In §3.5 we show

$$\limsup_{x \to \infty} \frac{\mathbb{E}^{\mathcal{D}}_\eta \mathcal{E}^2_x}{[\mathbb{P}(|V| > x)]^2} < \infty$$

In plain words, the estimator $\mathcal{E}_x$ has bounded relative error. See Asmussen and Glynn [6], §1 for definition of *bounded relative error*. The method presented in this paper is a multivariate generalization of Collamore et al. [36]. The reader is referred to it for the treatment of the one-dimensional problem.

1.5.3 Eigenvalues of the sample covariance matrix of a stochastic volatility model

In §1.4 we have introduced stochastic volatility models and discussed their connection to realized quadratic variation and their improved predictive power derived thereby. But there we have only discussed univariate models. In fact, the generalization to multivariate models is rather straightforward. In chapter 4 we adopt the following multivariate model:

$$\begin{aligned} X_{i,t} &= \sigma_{i,t} Z_{i,t} \quad 1 \le i \le p,\ t \in \mathbb{Z}, \\ \log \sigma_{i,t} &= \sum_{k,l \in \mathbb{Z}} \psi_{k,l} \eta_{i-k,t-l} \quad 1 \le i \le p,\ t \in \mathbb{Z}, \end{aligned}$$

where $\{Z_{i,t}\}$ and $\{\eta_{i,t}\}$ are iid fields of random variables. They are independent of each other. The distribution of η, a generic element of $\{\eta_{i,t}\}$, satisfies $\mathbb{P}(e^\eta > x) \sim x^{-\alpha} L(x)$, where $\alpha > 0$ and $L(x)$ is a slowly varying function. The coefficients $\{\psi_{k,l}\}$ are real and satisfy $\sum_{k,l \in \mathbb{Z}} |\psi_{k,l}| < \infty$.

Depending on the tails of $\{\sigma_{i,t}\}$ and $\{Z_{i,t}\}$, two situations can arise. When $\{Z_{i,t}\}_{t \in \mathbb{Z}}$ is a regularly varying sequence with index $\alpha \in (0,4)$ and dominates the tail, we show that each of the sequences $\{X_{i,t}\}_{t \in \mathbb{Z}}, 1 \le i \le p$ and each of $\{X_{i,t}X_{j,t}\}_{t \in \mathbb{Z}}, 1 \le i < j \le p$ are regularly varying with index α, assuming suitable conditions on $\{\sigma_{i,t}\}$.

Define the matrix $\mathbf{X}$

$$\mathbf{X} = \{X_{i,t}\}_{1 \le i \le p,\ 1 \le t \le n}$$

and let $\mathbf{X}'$ denote the transpose of $\mathbf{X}$. Then

$$\mathbf{X}\mathbf{X}' = \left\{ \sum_{t=1}^n X_{i,t} X_{j,t} \right\}_{1 \le i,j \le p}.$$

Using theorem 1.1 of Mikosch and Wintenberger [90] (see theorem 1.1 of this thesis), we prove

$$a_n^{-2} \left(\sum_{t=1}^n X_{i,t}^2 - n \mathbf{1}_{(2,4)}(\alpha) \mathbb{E} X^2 \right)_{1 \le i \le p} \xrightarrow{d} (\xi_{i,\alpha/2})_{1 \le i \le p}$$

and

$$a_n^{-2} \sum_{t=1}^n X_{i,t} X_{j,t} \xrightarrow{P} 0 \quad \text{for } i \ne j.$$

where $\{a_n\}_{n \ge 1}$ is such that $n\mathbb{P}(|X| > a_n) \to 1$ as $n \to \infty$, and $\{\xi_{i,\alpha/2}\}_{1 \le i \le p}$ is an iid sequence of $\alpha/2$-stable random variable. See chapter 4 for details. Built on this result, we show that $\mathbf{X}\mathbf{X}'$ is approximated by its diagonal:

$$a_n^{-2} \|\mathbf{X}\mathbf{X}' - \operatorname{diag}(\mathbf{X}\mathbf{X}')\| \xrightarrow{P} 0. \tag{1.20}$$

where $\|\cdot\|$ denotes the spectral norm. Following (1.20), we have

$$a_n^{-2}(\lambda_{(1)}, \ldots, \lambda_{(p)}) \xrightarrow{d} (\xi_{(1),\alpha/2}, \ldots, \xi_{(p),\alpha/2})$$

where $\lambda_{(i)}$ is the i-th upper order statistic of the eigenvalues of the matrix $\mathbf{X}\mathbf{X}'$ and $\xi_{(i),\alpha/2}$ is the i-th upper order statistic of the iid sequence $\{\xi_{i,\alpha/2}\}_{1 \le i \le p}$.

When $\sigma_{i,t}$ dominates the tail of $X_{i,t} = \sigma_{i,t} Z_{i,t}$ and satisfies a few more technical conditions, we show that each of the sequences $\{\sigma_{i,t}\}_{t \in \mathbb{Z},\ 1 \le i \le p}$ is regularly varying with index α. Moreover, in contrast to the previous case, we show that the sequence $\{\sigma_{i,t}\sigma_{j,t}\}_{t\in\mathbb{Z},1\le i,j\le p}$ is regularly varying with index $\alpha/\psi^{i,j}$, where $\psi^{i,j} = \max_{k,l}(\psi_{k,l} + \psi_{k+i-j,l})$. For $d \ge 1$, the d-variate sequence $\{(\sigma_{i_k,t}\sigma_{j_k,t})_{1\le k\le d}\}_{t\in\mathbb{Z}}$ is regularly varying with index $\alpha/(\max_{1\le k\le d}\psi^{i_k,j_k})$.

This result then allows us to approximate the matrix of $\mathbf{XX}'$ by $\{\tilde{X}_{i,j}\}_{1\le i,j\le p}$, where

$$\tilde{X}_{i,j} = \sum_{t=1}^{n} X_{i,t}X_{j,t}\mathbf{1}_{\{1\le i,j\le p,\psi^{i,j}=2\}}$$

A notable difference from the previous case is that the matrix $a_n^{-2}\{\tilde{X}_{i,j}\}_{1\le i,j\le p}$ can have non-vanishing values on its off-diagonal entries in the limit $n \to \infty$, implying its eigenvalues in this limit may not be solely determined by its diagonal entries.

1.5.4 Extreme value analysis for the sample auto-covariance matrices of time series

Janßen et al. [79] investigated the sample covariance matrix of the time series $\{X_{i,t}\}_{t\in\mathbb{Z}}$, $1 \le i \le p$ for *stochastic volatility models* assuming the dimension of the matrix p is fixed. It is also of interest to look into the sample covariance and the sample auto-covariance matrix when the dimension p tends to infinity at some rate as the number of observations n tends to infinity. This is the subject of Davis et al. [38] which we summarize in this section.

We are interested in the model

$$X_{i,t} \quad = \sum_{k,l\in\mathbb{Z}} h_{k,l}Z_{i-k,t-l}, \quad i,t \in \mathbb{Z} \tag{1.21}$$

where $\{Z_{i,t}\}_{i,t\in\mathbb{Z}}$ is a field of iid random variables and $h_{k,l}$ is an array of real coefficients. We assume $\{Z_{i,t}\}$ is regularly varying with index $\alpha \in (0,4)$ and

$$\sum_{k,l\in\mathbb{Z}} |h_{k,l}|^{\delta} < \infty$$

for some $\delta \in (0, \min\{\alpha/2, 1\})$. This condition ensures that the infinite series in (1.21) is almost surely absolutely convergent. Since each $\{X_{i,t}\}_{t\in\mathbb{Z}}$ is a linear combination of the sequences $\{Z_{i,t}\}_{t\in\mathbb{Z}}$ that are regularly varying with index α, each of $\{X_{i,t}\}_{t\in\mathbb{Z}}$ is regularly varying with index α.

Define matrices

$$\mathbf{X}(s) = \{X_{i,t+s}\}_{1\le i\le p;\ 1\le t\le n} \quad s \ge 0,$$

and assume that

$$\frac{p}{n^{\beta}\ell(n)} \to \infty, n \to \infty,$$

where $\beta \ge 0$ and ℓ is a slowly varying function. If $\beta = 0$, we assume in addition $\ell(n) \to \infty, n \to \infty$. We intend to understand the behavior of the singular values of the matrix $\mathbf{X}(0)\mathbf{X}(s)$, i.e. the square roots of the eigenvalues of $\mathbf{X}(0)\mathbf{X}(s)\mathbf{X}'(s)\mathbf{X}'(0)$.

Corresponding to the time-lagged matrices $\mathbf{X}(s)$, we also introduce the time-lagged coefficient matrices. Define

$$\mathbf{H}(s) = \{h_{k,l+s}\}_{k,l\in\mathbb{Z}}$$

and denote by $v_1(s) \geq v_2(s) \geq \cdots$ the singular values of the matrix $\mathbf{H}(0)\mathbf{H}(s) = \{\sum_{l\in\mathbb{Z}} h_{i,l}h_{j,l+s}\}_{i,j\in\mathbb{Z}}$. We have established the following point process convergence result upon appropriate conditions (see theorem 5.10 of chapter 5):

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(\lambda_{(i)}(0),\ldots,\lambda_{(i)}(s))} \xrightarrow{d} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}(v_j(0),\ldots,v_j(s))}.$$

where $\{a_{np}\}_{n,p\in\mathbb{Z}}$ is a sequence of positive real numbers such that $np\mathbb{P}(Z > a_{np}) \to 1, n \to \infty$. ε_x is the Dirac measure with unit mass at x and $\Gamma_i = \sum_{k=1}^{i} E_k$, where $\{E_k\}_{1\leq k\leq i}$ is an iid sequence of Exp(1) random variables. In particular, we see $a_{np}^{-2}\lambda_{(1)}(0) \xrightarrow{d} E_1^{-2/\alpha}v_1(0)$, i.e. asymptotically $a_{np}^{-2}\lambda_{(1)}(0)$ has a scaled Fréchet($\alpha/2$) distribution.

These results extend and generalize previous work by Soshnikov [105] and Auffinger et al. [7] who deal with the case of iid $X_{i,t}$, assuming regular variation with $\alpha \in (0, 4)$. This work also generalizes Davis et al. [45] as well as Davis et al. [44] who deal with the linear model (1.21) under suboptimal conditions on the growth rate of $p \to \infty$.

In this paper, large deviation results for sums of regularly varying random variables are consequently used; see Nagaev [92]. This is in contrast to the papers by Soshnikov [105] and Auffinger et al. [7] who only deal with the case when $p/n \to \gamma \in (0, \infty)$. The large deviation approach allows one to determine the dominating parts of the sample covariance matrix $\mathbf{XX}'$ and the sample auto-covariance matrices. Typically these parts are given by functionals of $\{Z_{i,t}^2\}$.

In the case of a finite 4th moment of $\{X_{i,t}\}$, the theory changes completely. A typical result in this case is proved by Johnstone [83]: if $p/n \to \gamma \in (0, \infty)$, then

$$\frac{\lambda_{(1)} - \mu_{np}}{\sigma_{np}} \xrightarrow{d} W_1 \sim F_1,$$

where the constants μ_{np} and σ_{np} are given by

$$\begin{aligned} \mu_{np} &= (\sqrt{n-1} + \sqrt{p})^2, \\ \sigma_{np} &= (\sqrt{n-1} + \sqrt{p})\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}. \end{aligned}$$

W_1 is a Tracy-Widom random variable with distribution function F_1 defined by

$$F_1(s) = \exp\left\{-\frac{1}{2}\int_s^\infty [q(x) + (x-s)q^2(x)]dx\right\}.$$

q is defined as the solution to the Painlevé II differential equation

$$\begin{aligned} q''(x) &= xq(x) + 2q^3(x), \\ q(x) &\sim \mathrm{Ai}(x) \quad \text{as } x \to \infty. \end{aligned}$$

where Ai(x) is the Airy function. In the heavy-tailed case considered in this thesis, asymptotic results about the eigenvalues are easier to derive than in the light-tailed case, i.e. when 4th moments are finite.

Chapter 2

Do return series have power-law tails with the same index?

Thomas Mikosch, Casper de Vries & Xiaolei Xie
technical report

Abstract

We consider an investor with preferences that accord with Generalized Disappointment Aversion (GDA). Such an investor cares about downside risk and we assume he recognizes the heavy tail feature of asset return distributions. We argue that when a market is dominated by rational investors of this kind, the return distributions of equities that are actively traded in this market may have nearly equal tail-indices due to monotonicity of the GDA preference with respect to the tail index. We give conditions upon which the GDA preference is monotone and hence suggests an equal tail index for all actively traded stocks.

We also estimate tail indices and scale parameters of S&P 500 stocks and test the hypothesis that two given stock return series have the same tail index. The results vary across different sectors of the index.

2.1 Introduction

It is one of the stylized facts of financial econometrics that returns of speculative prices are *heavy-tailed*. There is no agreement in the literature about how heavy these tails really are. For example, Barndorff-Nielsen and Shephard [11] and Eberlein [54] favor "semi-heavy" tails which are comparable with those of a gamma distribution. On the other hand, tails of returns have been studied in great detail in the extreme value community. Among extreme value specialists there is general agreement that returns X_t have tails of power-law-type, i.e.,

$$\mathbb{P}(X_t > x) \sim c_+ \, x^{-\alpha_{\text{up}}} \quad \text{and} \quad \mathbb{P}(X_t < -x) \sim c_- \, x^{-\alpha_{\text{low}}} \,, \quad x \to \infty \,, \tag{2.1}$$

where $c_\pm$, α_{up} and α_{low} are positive constants.[1] See for example, Embrechts et al. [58], Jansen and de Vries [77], Mikosch [88], Resnick [97]. In the extreme value literature it is common to replace the constants $c_\pm$ by suitable *slowly varying* functions; cf. Embrechts et al. [58], Chapter 3. In this paper, for the sake of argument, we stick to the condition (2.1).

There are some good theoretical reasons for the appearance of power-law tails in situations where certain moments of data are believed to be infinite. Tails of type (2.1) describe the maximum domain of attraction of the Fréchet distribution $\Phi_{\alpha_{\text{up}}}(x) = \exp(-x^{-\alpha_{\text{up}}})$

[1] Here and in what follows, $f(x) \sim g(x)$ for positive functions f and g means that $f(x)/g(x) \to 1$ as $x \to \infty$.

for $x > 0$, i.e., scaled maxima of an iid sequence (X_t) with upper tail described in (2.1) converge in distribution to $\Phi_{\alpha_{\rm up}}$. Equivalently, power-law tails are prescribed by the generalized Pareto distribution which is the limit distribution of the excesses of X_t above high thresholds, i.e., for a suitable positive scaling function $a(u)$,

$$\mathbb{P}((X_t - u)/a(u) > x \mid X_t > x) \to (1 + x/\alpha_{\rm up})^{-\alpha_{\rm up}}, \qquad u \to \infty.$$

The aforementioned results are considered very natural for iid and weakly dependent strictly stationary sequences of random variables (X_t); in the world of extremes they are the analogs of the central limit theorem from the world of sums.

In the literature on extremes for return data one finds the statement that *estimated* values $\hat{\alpha}_{\rm up}$ and $\hat{\alpha}_{\rm low}$ of the tail-indices $\alpha_{\rm up}$ and $\alpha_{\rm low}$, respectively, typically have the tendency that $\hat{\alpha}_{\rm up} > \hat{\alpha}_{\rm low}$. This observation is often explained by the fact that investors are more prone to negative than to positive news in the market. Moreover, in the literature the *estimated* tail-indices $\hat{\alpha}$ (both in the left and right tails) are typically found in the range $(2, 4)$. For an illustration, see Figure 2.1 where estimates $\hat{\alpha}_{\rm low}$ in three sectors of the Standard & Poors 500 index are shown. The estimates are based on 1304 observations of daily return data from 4 January 2010 to 31 December 2014.

When looking at Figure 2.1 one might ask the following questions:

- In view of the wide asymptotic confidence bands for the estimators of tail-indices, are the tail-indices from different series really distinct?
- Are there some *theoretical* reasons supporting the fact that the tail-indices from different series are *not* distinct?

In this paper, we try to find some answers to these questions.

The estimator of the tail-index $\alpha > 0$ in the model

$$\mathbb{P}(X_t > x) \sim c\, x^{-\alpha}, \qquad x \to \infty,$$

favored in the literature is the *Hill estimator*; the graphs in Figure 2.1 are based on this estimator. We introduce this estimator in Section 2.2 and discuss some of its virtues and vices. In addition to tail-index estimation we also discuss the related problem of estimation of the scale parameters in the tail (these are the constants c_+ and c_- in (2.1)).

In Section 2.3 we discuss the theoretical problem of appearance of power-law tails in models for daily or, more generally, low-frequency return data. In particular, in Section 2.3.1 we address the power-law tails of univariate and multivariate GARCH models as potential models for a set of return data from distinct assets. As a matter of fact, under mild conditions, the aforementioned models have power-law tails due to their relation with so-called *stochastic recurrence equations*. Moreover, some of the *standard* multivariate GARCH models as the CCC ensure that the component-wise marginal distributions have power-law tails with the same index.

In Section 2.3.2 we discuss an economic argument for the fact that return data of similar assets (such as return series in a given sector of the S&P 500 index) may have tail-indices which are close to each other. We argue based on a utility function approach. We explicitly recognize the behavioral concern for downside risk in an investor's evaluation of a portfolio using the framework of Generalized Disappointment Aversion (GDA) introduced by Routledge and Zin [99]. GDA is an extension of the concept of Disappointment Aversion (DA) of Gul [66] who derived DA from first principles (axiomatic).

In Section 2.4 we summarize the discussion of the previous sections.

2.2 Power-law tails of return series: some empirical results

In this section, we assume the model (2.1) for the tails of the marginal distribution of a univariate return series (X_t). For the sake of argument, we assume that this series constitutes a strictly stationary sequence. In what follows, we focus on the left tail of the distribution, i.e., on the losses. However, it is common to present the tail-index estimators for positive data. Therefore we will multiply the losses X_t by minus one, swapping the negative with the positive values. For simplicity, we also suppress subscripts in the notation:

$$\mathbb{P}(-X_t > x) \sim c\, x^{-\alpha}\,, \qquad x \to \infty\,, \tag{2.2}$$

where we assume that the two parameters – the tail-index α and the scale parameter c – are positive. They play crucial roles for the understanding of the risk hidden in the data, hence for asset allocation and risk management. These parameters are market characteristics and provide a simple but useful description of the risk, for example in terms of high quantiles such as Value-at-Risk. Alternatively, these parameters can be used for model building of the equities in the market such as the GARCH model; see Section 2.3.

2.2.1 Hill estimates of lower tail-indices

Various estimators of the tail-index α in the model (2.2) have been proposed in the literature; see Embrechts et al. [58], de Haan and Ferreira [67], Resnick [97]. The most popular among them was introduced by Hill [71]. Given a sample $-X_1, \dots, -X_n$ whose marginal distribution satisfies (2.2), calculate the order statistics $X_{(1)} \le \cdots \le X_{(n)}$ and construct the *Hill estimator*:

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

Here k is the number of upper order statistics in the sample used for the estimation. The estimator $\hat{\alpha}_k$ is a maximum-likelihood estimator of α based on the k upper order statistics in the pure Pareto model (recall that we multiplied the data by minus one)

$$\mathbb{P}(-X_t > x) = \frac{K^\alpha}{x^\alpha}\,, \qquad x > K\,, \tag{2.3}$$

under the hypothesis that we do not know the (high) threshold value K. The estimator has "good" theoretical properties such as asymptotic consistency and asymptotic normality. These properties hold under strict stationarity assumptions on the data; Drees and Rootzén [52] give perhaps most general conditions for dependent sequences and de Haan and Ferreira [67] provide a complete asymptotic theory in the iid case.

A major problem for Hill estimation is the choice of the number k of upper order statistics. As a matter of fact, if k is too large the order statistics are too close to the center of the distribution of the $-X_t$, leading to a bias of the estimator. On the other hand, by construction, $\hat{\alpha}_k$ is an average of k log-differences of the data. Therefore, the variance of the estimator is the larger the smaller k. For these reasons, asymptotic theory requires to choose $k = k_n$ $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. This fact does not make the estimation of α an easy matter: one has to choose a "small" value k which is not "too large". For practical purposes, a so-called *Hill plot* is recommended where $\hat{\alpha}_k$ is plotted for a variety of k-values, corresponding to some high quantile $X_{(n-k)}$ of the data. Then k is chosen from a region in the plot where $\hat{\alpha}_k$ is relatively stable. For example, in Figure 2.1 we have chosen $k = 50$ from a sample of size $n = 1304$, corresponding to the 96%-quantile of the data. In general, the estimation of the tail-index is an art and requires some expertise; for some guidance see Embrechts et al. [58], Resnick [97] and Drees et al. [53].

In Figure 2.1, we exhibit 95% asymptotic confidence bands derived from the central limit theorem

$$\sqrt{k}\left(\hat{\alpha}_k - \alpha\right) \xrightarrow{\mathrm{d}} N(0, \alpha^2)\,, \tag{2.4}$$

i.e., $\hat{\alpha}_k$ is asymptotically unbiased and has variance α^2/k. Since $k/n \to 0$ this means that the confidence bands are significantly larger than the classical $1/\sqrt{n}$-rates. This fact is one explanation for the fact that it is difficult to say something meaningful about the true value of α. There exist various other reasons why one should not have 100% trust in the confidence bands shown in Figure 2.1. Indeed, (2.4) holds under rather subtle *second order conditions* on the tail $\mathbb{P}(X_t > x)$ which cannot be verified on data. However, given a theoretical model such as the GARCH, these conditions can be verified based on the theoretical properties of the model. If they are not satisfied the Hill estimator may exhibit significant bias; see Embrechts et al. [58] and Resnick [96] for illustrations of this fact leading to so-called "Hill horror plots". Moreover, the Hill estimator is rather sensitive to non-stationarity of the data and to dependence. For example, results in Drees [51], and Drees and Rootzén [52] show that the asymptotic variance of the Hill estimator can be significantly larger than in the iid case. Since return data are dependent, the asymptotic confidence bands should be even wider than exhibited in Figure 2.1. Again, only under he assumption of a concrete model like GARCH these confidence bands can be evaluated and therefore the bands shown in Figure 2.1 just show some benchmark which holds in the iid case and under additional conditions on the tail asymptotics.

In Figure 2.1 we see significant overlap of the confidence intervals of the Hill estimates of the losses in the "Energy" and "Consumer Staples" sectors of the S&P 500 index, as well as those of a large portion of losses in the "Information Technology" sector. This fact indicates that the returns in each sector may have comparable tail-indices.

Hoga's [72] test about the change of extreme quantiles in a sample may provide some further insight about how similar these tail-indices are. Different tail-indices are likely to result in different extreme quantiles. Nevertheless, changes in the extreme quantiles may also result from changing scale parameters in the tail. Therefore we first investigate the scale parameters of daily stock returns in the same sectors of S&P 500 before we apply the test.

2.2.2 Hill estimates of lower-tail scale parameters

We assume the pure Pareto model (2.3) with scale parameter $K > 0$. Hill [71] proposed the maximum-likelihood estimator of K derived from the joint distribution of k upper order statistics in the sample; cf. Embrechts et al. [58], p. 334. It is given by

$$\hat{K}_k = \left(\frac{k}{n}\right)^{1/\hat{\alpha}_k} X_{(n-k)}\,.$$

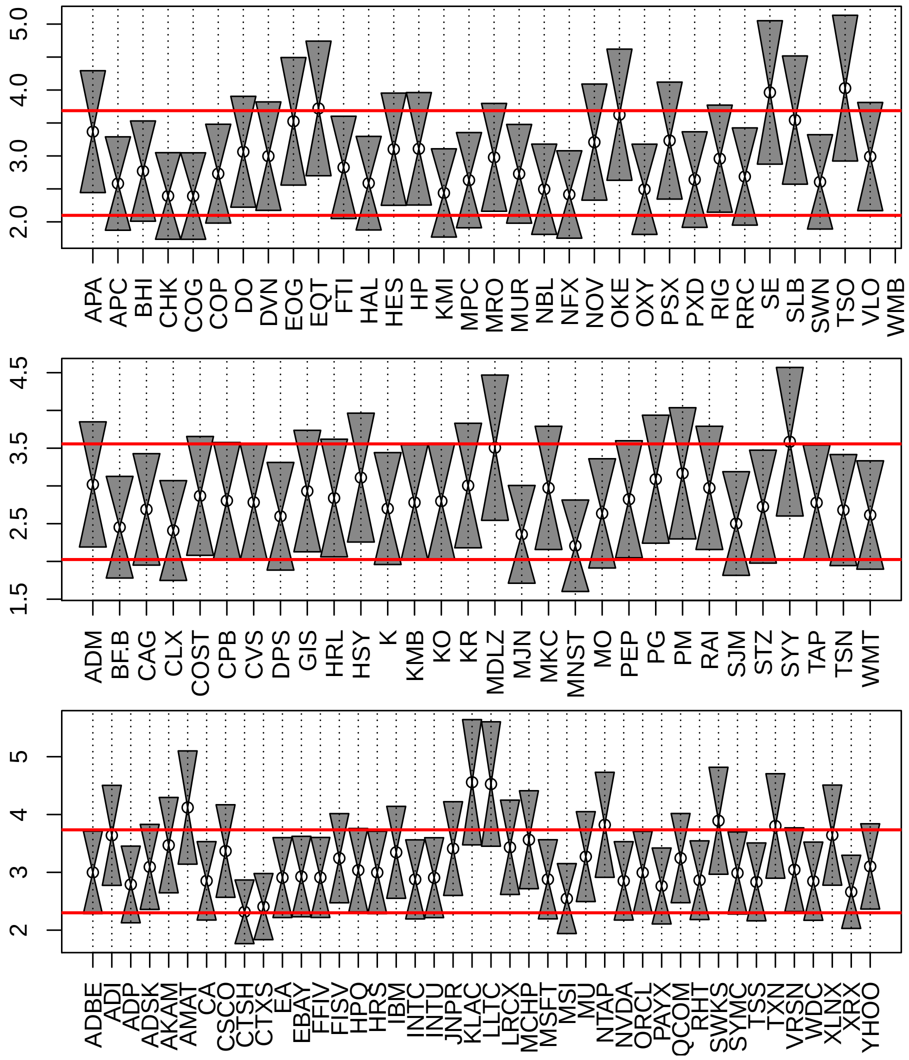


Figure 2.1: Hill estimates $\hat{\alpha}_{50}$ of the lower tail-indices α of daily return series in sectors of the S&P 500 index. The data span from 1 January 2010 to 31 December 2014 and comprise $n = 1304$ observations. The graphs from top to bottom correspond to the "Energy", "Consumer Staples" and "Information Technology" sectors. Each circle corresponds to a Hill estimate $\hat{\alpha}_{50}$; the gray triangles above and below it mark the 97.5% and 2.5% quantiles of its approximate normal distribution; see (2.4) and the discussion following it for an interpretation. The lower and upper red lines mark the medians of the 2.5% and 97.5% quantiles, respectively, evaluated from all stocks in the sector. The data are taken from *Yahoo Finance*; the labels on the horizontal axes are Yahoo symbols of the stocks.

Using the asymptotic normality property of upper order statistics (cf. de Haan and Ferreira [67], Theorem 2.2.1), one can show

$$\sqrt{k}\,(\hat{K}_k - K) \quad \overset{\mathrm{d}}{\to} \quad N\big(0, (K/\alpha)^2\big) \quad \text{and} \quad \sqrt{k}\,(\hat{K}_k^{\alpha} - K^{\alpha}) \overset{d}{\to} N\big(0, K^{2\alpha}\big)\,,$$

where the tail-index α is regarded as known. From the above asymptotic normality property, confidence bands of $\hat{K}_k$ and $\hat{K}_k^{\hat{\alpha}}$ can be constructed. Estimates $\hat{K}_k$ in the "Energy", "Consumer Staples" and "Information Technology" sectors of the S&P 500 index are computed using this method. The results are shown in Figure 2.2. One can see

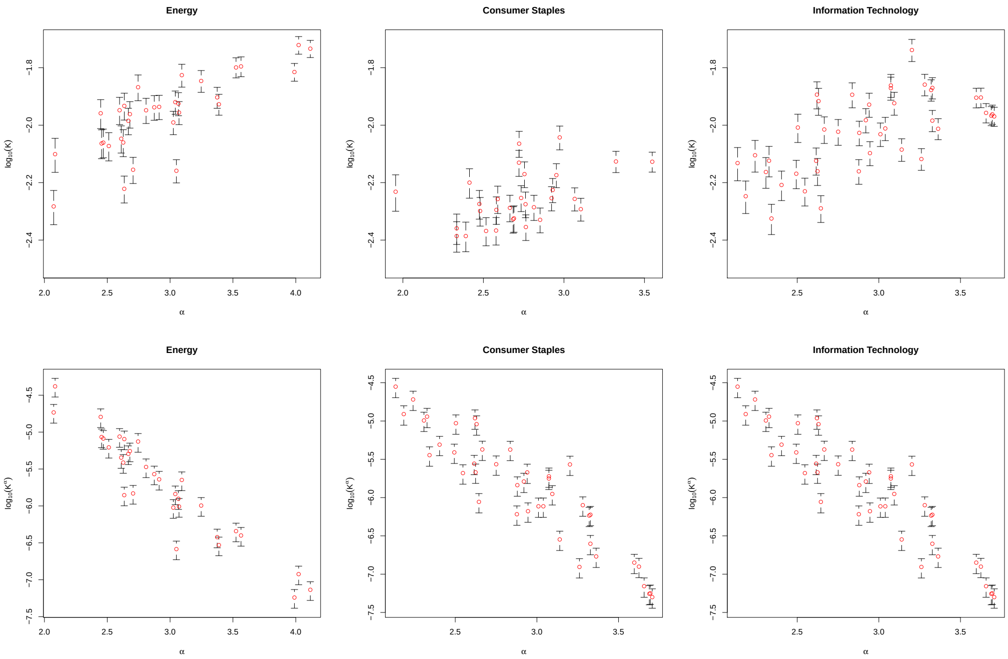


Figure 2.2: Estimates of $\hat{K}_k$ (top) and $\hat{K}_k^{\hat{\alpha}}$ (bottom) on $\log_{10}$-scale of stocks in sectors of the S&P 500 index. The estimates are ordered according to the corresponding estimated α-values. The points are the estimated values, the bars the asymptotic 95%-confidence intervals; the confidence bands of the corresponding Hill estimates $\hat{\alpha}_k$ of these sectors are shown in Figure 2.1.

that K generally takes a rather small value. For the more volatile sectors of "Energy" and "Information Technology", the average value of K is around 0.01, while for the more stable sector of "Consumer Staples", the average value is around 0.005. Due to the smallness of K, mild variations of α would lead to huge variations of K^α, as shown in the 2nd row of Figure 2.2.

Secondly, it appears that there is positive dependence between the values of α and K. As argued in Section 2.3.3, this is consistent with the assumption that the return series have Pareto tails on both sides with the tail parameters on each side independent of those on the other.

Thirdly, Figure 2.2 shows that, on average, the values of K in the "Energy" and "Information Technology" sectors are larger than those in the "Consumer Staples" sector. For a given loss probability, a larger value of K implies that large losses are more probable. Thus one can conclude that these two sectors are considerably riskier than the "Consumer Staples" sector. This is of course a confirmation of one's economic instinct.

Yet another indication from Figure 2.2 is that, while the "Energy" and the "Information Technology" sectors are similar in riskiness, the dependence between α and K is stronger in "Energy". As discussed in Section 2.3.2 below, when moving along a curve of equal preference in the direction of increasing α, the parameter K also increases. So the strong positive dependence seen in the "Energy" sector suggests that these stocks might have very similar investor preferences. This in turn may be attributed to stronger business relations between the energy enterprises. While two IT companies may provide a variety of products and services and do not depend on each other, two energy companies are more likely to depend on each other via relations of supplier and customer or otherwise to compete with each other if they are on the same link of the chain of energy production and distribution.

2.2.3 A test for equal tail-indices based on Hill estimation

Suppose we have two independent strictly stationary positive series $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ with corresponding distribution functions F_X and F_Y that have power-law tails with indices α_X and α_Y, respectively. From (2.4) one can deduce

$$\sqrt{k}\begin{pmatrix}\hat{\alpha}_X - \alpha_X\\ \hat{\alpha}_Y - \alpha_Y\end{pmatrix} \xrightarrow{d} \begin{pmatrix}Z_X\\ Z_Y\end{pmatrix} \sim N\left(0, \operatorname{diag}(\alpha_X^2, \alpha_Y^2)\right) \tag{2.5}$$

where $\hat{\alpha}_X$ and $\hat{\alpha}_Y$ are Hill estimators of α_X and α_Y; we suppress their dependence on k. Then it follows from the continuous mapping theorem

$$\sqrt{k}[(\hat{\alpha}_X - \alpha_X) - (\hat{\alpha}_Y - \alpha_Y)] \xrightarrow{d} Z_X - Z_Y \sim N(0, \alpha_X^2 + \alpha_Y^2)\,. \tag{2.6}$$

This relation allows one to construct an asymptotic test under the null hypothesis $\alpha_X = \alpha_Y$ and with test statistic $\hat{\alpha}_X - \hat{\alpha}_Y$. We apply this test to the equities in the "Energy", "Consumer Staples" and "Information Technology" sectors of the S&P 500 index. The results are shown in the top row of Figure 2.3. They indicate that tail-indices of equities in the "Energy" or "Information Technology" sectors are more variable than in the "Consumer Staples" sector, as the null hypothesis is rejected more often for members of these two former sectors. Moreover, these figures suggest that the test based on the Hill estimator is quite powerful in distinguishing between tail-indices. In contrast to the test presented in Section 2.2.4 the present test results in more rejections for the "Energy" and the "Information Technology" sectors.

As a caution, one should bear in mind that (2.5) is valid on condition that the X- and Y- series are independent of each other (or weakly dependent on each other), which is generally untrue for two return series in the same market.

2.2.4 A test for a change in the extreme tail

Here we apply a test from a recent paper by Hoga [72]. This test has been developed for a different kind of problem. Given a strictly stationary time series $X_1, \dots, X_n$ with a marginal distribution F with right power-law tail, the goal is to test whether there is a structural break of the *extreme quantiles* $F^{-1}(1-p)$ for values p very close to zero. If the tail-index *or* the scale parameter in a distribution of type (2.3) change inside a sample, then it is likely that the extreme quantiles change as well. We will test for a change of tail-index or scale parameter in this indirect way.

The null hypothesis of the test in [72] is that there is no change of the extreme quantiles $F^{-1}(1-p)$ for $p = p_n \to 0$ in any subsample with indices $t \in (n\,t_0, n(1-t_0))$ where t_0 is a fixed number in $(0, 0.5)$. Writing $\hat{x}_p(a, b)$ for an estimator of the extreme

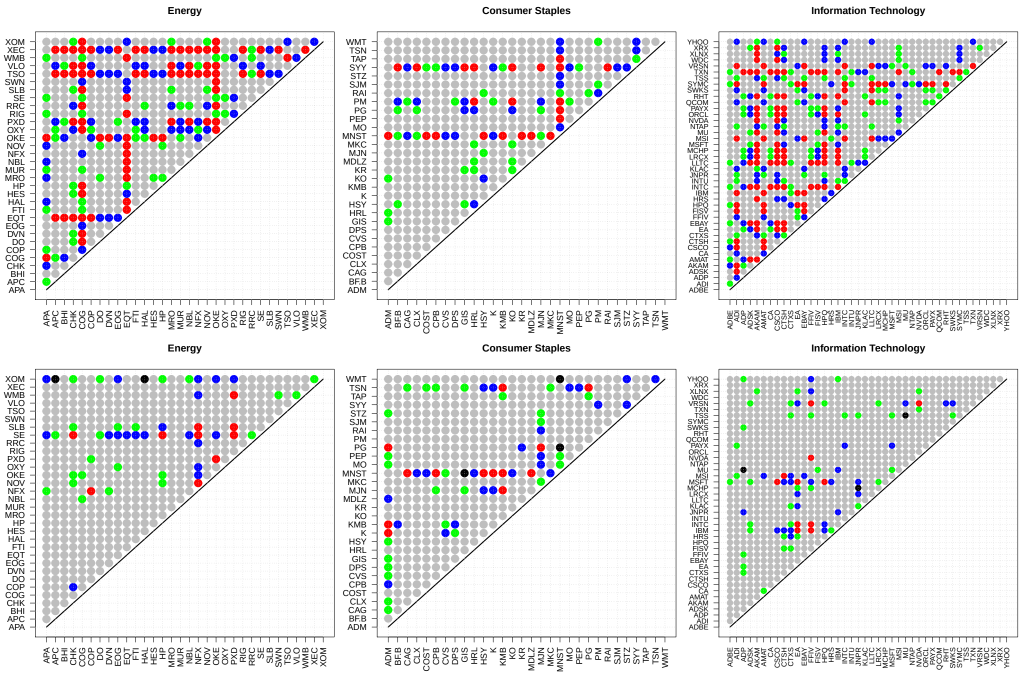


Figure 2.3: *Top row*: Test for pairwise equality of tail-indices of losses in the "Energy", "Consumer Staples" and "Information Technology" sectors of S&P 500. The test statistic in $\hat{\alpha}_X - \hat{\alpha}_Y$ is based on Hill estimates of α_X and α_Y. The green, blue and red points correspond to pairs of stock in a sector when the test statistic is outside the intervals $[q_{0.075}, q_{0.925}]$, $[q_{0.05}, q_{0.95}]$, $[q_{0.025}, q_{0.975}]$, respectively, where q_p is the p-quantile of the limiting $N(0, \alpha_X^2 + \alpha_Y^2)$-distribution of the test statistic in (2.6). Grey points stand for pairs for which the test statistic is inside $[q_{0.075}, q_{0.925}]$. *Bottom row:* Test for changing tail-index or scale parameter of losses using Hoga's test based on concatenated series of pairs of stocks. The green, blue and red points correspond to pairs of stock in a sector when the test statistic T_n exceeds the 85%-, 90%-, 95%-quantile of the limit distribution. Grey points stand for pairs for which the test statistic is below the asymptotic 85%-quantile. Black points represent pairs for which the computation of T_n fails for given precision requirements and time limits. The same number (50) of upper order statistics is used for both tests.

$(1-p)$-quantile based on the subsample with indices $t \in (na, nb)$, the test statistic is given by

$$T_n = \sup_{s \in [t_0, 1-t_0]} \frac{\left[s(1-s)\log\left(\hat{x}_p(0,s)/\hat{x}_p(s,1)\right)\right]^2}{\int_{t_0}^{s}\left[r\log\left(\hat{x}_p(0,r)/\hat{x}_p(0,s)\right)\right]^2 dr + \int_{s}^{1-t_0}\left[(1-r)\log\left(\hat{x}_p(r,1)/\hat{x}_p(s,1)\right)\right]^2 dr} \tag{2.7}$$

Under the null hypothesis, (T_n) converges to a complicated functional of Brownian motion on $[0,1]$; the asymptotic quantiles need to be evaluated by simulation.

When applied to our problem we would like to test whether there is a change of the tail-index *or* scale parameter in (2.3) in each of the S&P 500 series in the distinct sectors. We also want to get some indication about a possible change of tail-index or scale parameter from one series to another within a given sector. For this reason, we choose any pair of series within a sector and concatenate each of the paired series. Then we run the test on the concatenated series. Of course, despite the fact that we test changes of tail-index or scale parameter *in a very indirect way* – there may be many other reasons for the change of extreme quantiles in a sample – we also concatenate two rather distinct series. Even if we assume that the two series come from related models (such as GARCH), the parameters of these models will in general not be the same. Moreover, the concatenation of two strictly stationary time series is in general not strictly stationary. Therefore we have to be careful with interpretations of the results of the tests.

In Figure 2.4 we show the values of the test statistic T_n (horizontal bars) for $t_0 = 0.1$ and daily return series of stock in the "Energy" and "Consumer Staples" sectors of the S&P 500 index. The "null hypothesis" is that the tail-index and scale parameter remain the same throughout the selected period of time. For most stocks, the hypothesis cannot be rejected even at the 85% level. This fact may be an indication that the distribution inside a series is rather homogeneous. Alternatively, it may show that the power of the test is very low. A possible reason for this suspicion is that the convergence rate of (T_n) to its limit is very slow, i.e., the asymptotic distribution is not representative for the distribution of T_n for the chosen n; for some simulation evidence, see below.

To check whether any pair of stocks shares the same tail-index and scale parameter we concatenate any two series and apply the aforementioned test on the concatenated series. For the "Energy" and the "Consumer Staples" sector we summarize the results in the bottom row of Figure 2.3. These graphs show that the "null hypothesis" of an equal tail-index and scale parameter is rejected for more pairs in the "Energy" sector than it is for those in the "Consumer Staples" sector. This suggests that lower tail-indices of stocks in the "Energy" sector are more spread out than those of the "Consumer Staples" sector. Also observe that while 3 stocks, say A, B and C, test in favor of the relations $\alpha_A = \alpha_B, \alpha_A \neq \alpha_C$, it often happens that another test on B and C is supportive of $\alpha_B = \alpha_C$. Again, this is due to the limited power of the test. Based on such results, one may guess that α_B lies between α_A and α_C. The test is unable to recognize the smaller differences between α_A, α_B on one hand and between α_B, α_C on the other hand.

To get an idea about the power of the test we run it on a sample concatenated from two independent iid samples of the same size $n = 1304$ as the S&P 500 series. Both pieces are t-distributed with distinct degrees of freedom. The results are shown in Figure 2.5: the power of the test is the smaller the larger the minimum tail-index in the concatenated pair.

A major problem of this test is the asymptotic distribution of the test statistic under the null hypothesis. The rate at which the finite-sample distribution tends to its limit is not known. To find out about this problem we compared the distributions of T_{1304} for t-distributed X_t with $\alpha = 3$ and $\alpha = 4$ degrees of freedom with the limit distribution of T_n. The estimated density functions are shown on the right of Figure 2.5. As seen in the graph, the asymptotic distribution assigns significantly more mass to the tail than the distributions of T_n do. For comparison, we list a few quantiles of these distributions in Table 2.1, showing major differences between the asymptotic and finite-sample distributions.

Figure 2.6 points at another shortcoming of the test: we show T_n for an arbitrarily chosen random permutation of the concatenated data from two different stocks. In this case the null hypothesis that two series have the same tail-index is rejected much more

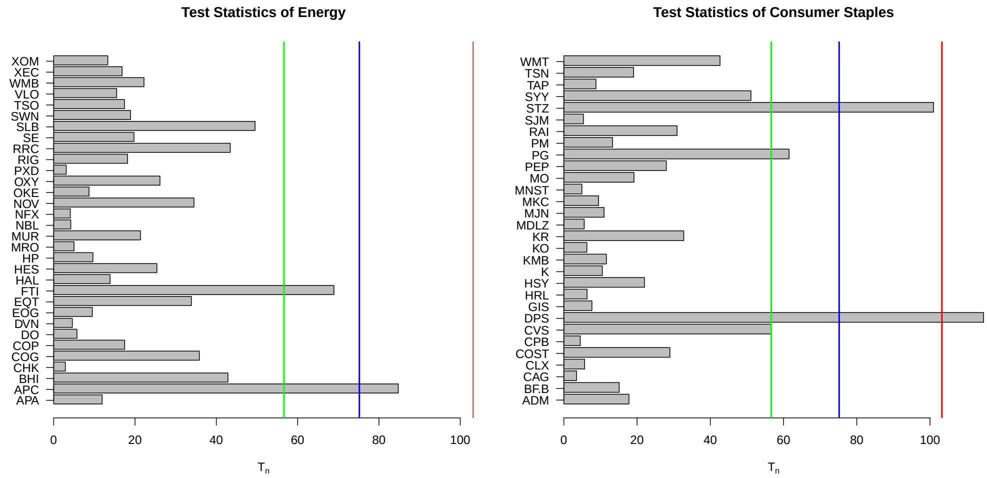


Figure 2.4: Test statistic T_n from (2.7) for the stocks in the "Energy" and "Consumer Staples" sectors of S&P 500. The green, blue and red lines correspond to the 85%, 90% and 95% quantiles of the limit distribution of T_n. They are derived by simulations from the limit distribution.

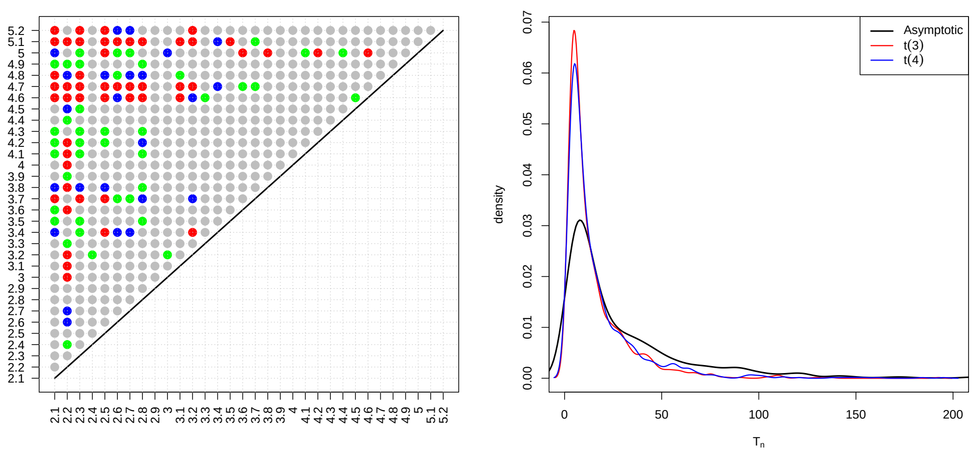


Figure 2.5: *Left*: Test of concatenated t-samples with different degrees of freedom α. Numbers on the axes are the degrees of freedom in the subsamples. For an interpretation of the colored bullets, see the caption for the bottom row of Figure 2.3. The graph shows the limited power of the test. In particular, if both degrees of freedom are relatively large it loses the capability of distinguishing between the distributions. *Right*: Comparison of the asymptotic distribution of the test statistic T_n in (2.7) under the null hypothesis and the distribution of T_n for $n = 1304$ iid t-distributed X_t with 3 and 4 degrees of freedom.

	Quantiles			
Distribution of X_t	80%	85%	90%	95%
Asymptotic	47.48	59.61	78.90	113.12
t(3)	23.30	28.27	35.04	46.75
t(4)	25.76	31.13	38.81	56.32

Table 2.1: Quantiles of the test statistic T_n for $n = 1304$ t-distributed samples with $\alpha = 3$ and $\alpha = 4$ degrees of freedom as well as the corresponding quantiles for the limiting distribution of T_n. In particular, there are huge differences between the three distributions for the higher quantiles.

often, as a comparison with the bottom graphs of Figure 2.3 shows. If the data in the concatenated series were iid a random permutation would not change the distribution of T_n. Thus the value of the test statistic T_n strongly depends on the dependence structure of the underlying data and therefore a test based on T_n may be misleading.

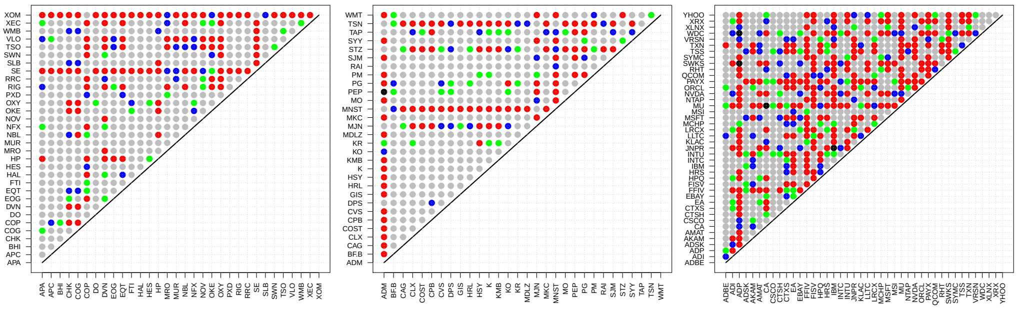


Figure 2.6: Test for changing tail-index or scale parameter of losses using Hoga's test based on concatenated series of pairs of stocks. A random permutation is applied to the observations of each series. The green, blue and red points correspond to pairs of stock in a sector when the test statistic T_n exceeds the 85%-, 90%-, 95%-quantile of the limit distribution. Grey points stand for pairs for which the test statistic is below the asymptotic 85%-quantile. Black points represent pairs for which the computation of T_n fails for given precision requirements and time limits.

2.3 Some theoretical arguments for equality of tail-indices

2.3.1 Multivariate GARCH models whose components have equal tail-indices

Among the models for returns the generalized autoregressive conditionally heteroscedastic (GARCH) model is certainly most popular because it is parsimonious, captures various of the stylized facts of real return data and can also be modified in various directions to capture specific behavior of time series such a asymmetry, skewness, long memory; see for example Andersen et al. [4], Part 1, for a collection of results on GARCH-type models. The original *univariate* GARCH model of Bollerslev [24] is a stochastic volatility model of the type $X_t = \sigma_t Z_t$, where (Z_t) is an iid mean-zero unit-variance sequence. In the simple case of a GARCH the *squared volatility* satisfies the *stochastic recurrence equation*

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2\,, \quad t \in \mathbb{Z}\,. \tag{2.8}$$

Here $\alpha_0 > 0$, α_1, β_1 are non-negative constants. For suitable choices of α_1, β_1 the equation (2.8) can be solved and the solution (σ_t^2) constitutes a strictly stationary sequence, implying that (X_t) is strictly stationary itself. A remarkable property of the process (σ_t) is that it has a power-law tail of the form

$$\mathbb{P}(\sigma_t > x) \sim c\, x^{-\alpha}\,, \qquad x \to \infty\,, \tag{2.9}$$

for some positive $c > 0$ and a positive tail-index α which is the unique solution of the equation $\mathbb{E}[(\alpha_1 Z_1^2 + \beta_1)^{\alpha/2}] = 1$ provided that the solution exists and some mild assumptions on the distribution of Z_t hold. This result follows by an application of the Kesten-Goldie theorem; see Kesten [84], Goldie [64], cf. Buraczewski et al. [29] for a recent textbook treatment. The latter result ensures power-law tails for the strictly stationary solution (Y_t) to the stochastic recurrence equation

$$Y_t = A_t\, Y_{t-1} + B_t\,, \qquad t \in \mathbb{Z}\,, \tag{2.10}$$

for an iid sequence of pairs (A_t, B_t), $t \in \mathbb{Z}$, with non-negative components satisfying $\mathbb{E}[A_1^{\alpha/2}] = 1$. In the model (2.8) we can choose $Y_t = \sigma_t^2$, $A_t = \alpha_1\, Z_{t-1}^2 + \beta_1$ and $B_t = \alpha_0$ to achieve (2.9). In turn, by an application of Breiman's lemma (see [29], p. 275) it follows that

$$\mathbb{P}(\pm X_t > x) \sim \mathbb{E}[(Z_t)_\pm^\alpha]\, \mathbb{P}(\sigma_t > x)\,, \qquad x \to \infty\,,$$

implying power-laws for the right and left tails of X_t caused by the power-law tail of σ_t.

A GARCH process of the order (p, q) can be embedded in a multivariate equation of the type (2.10), where $(\mathbf{A}_t)$ are iid random matrices and $\mathbf{B} = \mathbf{B}_t$ is a constant vector. Again, the Kesten theory [84] applies, implying that the marginal and finite-dimensional distributions of the GARCH process are regularly varying with a positive index α. We refrain from explaining the notion of multivariate regular variation which is needed in this context. For further details, see Buraczewski et al. [29] where the Kesten theorem and regular variation of GARCH processes are explained in detail.

There exist various extensions of the univariate GARCH model to the multivariate case. For the sake of argument, we stick here to the *constant conditional correlation*

(CCC) model of Bollerslev [25] and Jeantheau [80], and we only consider a special bivariate case. It is the model

$$\mathbf{X}_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} \end{pmatrix} \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} = \Sigma_t \, \mathbf{Z}_t \,, \qquad t \in \mathbb{Z} \,.$$

Thus both return components $X_{i,t}$ have the form of a univariate stochastic volatility model $X_{i,t} = \sigma_{i,t} Z_{i,t}$ with non-negative volatility $\sigma_{i,t}$ and an iid bivariate noise sequence $(\mathbf{Z}_t)$ with zero mean and unit variance components. We also have the specification

$$\begin{aligned} \mathbf{Y}_t = \begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{2,t}^2 \end{pmatrix} &= \begin{pmatrix} \alpha_{01} \\ \alpha_{02} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} X_{1,t-1}^2 \\ X_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{01} \\ \alpha_{02} \end{pmatrix} + \begin{pmatrix} \alpha_{11} Z_{1,t-1}^2 + \beta_{11} & \alpha_{12} Z_{2,t-1}^2 + \beta_{12} \\ \alpha_{21} Z_{1,t-1}^2 + \beta_{21} & \alpha_{22} Z_{2,t-1}^2 + \beta_{22} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \end{pmatrix} , \end{aligned} \quad (2.11)$$

for positive α_{0i} and suitable non-negative α_{ij}, β_{ij}, $i, j = 1, 2$. Writing

$$\mathbf{B}_t = \begin{pmatrix} \alpha_{01} \\ \alpha_{02} \end{pmatrix} \quad \text{and} \quad \mathbf{A}_t = \begin{pmatrix} \alpha_{11} Z_{1,t-1}^2 + \beta_{11} & \alpha_{12} Z_{2,t-1}^2 + \beta_{12} \\ \alpha_{21} Z_{1,t-1}^2 + \beta_{21} & \alpha_{22} Z_{2,t-1}^2 + \beta_{22} \end{pmatrix} ,$$

we see that we are again in the framework of a stochastic recurrence equation but this time for vector-valued $\mathbf{B}_t$ and matrix-valued $\mathbf{A}_t$:

$$\mathbf{Y}_t = \mathbf{A}_t \, \mathbf{Y}_{t-1} + \mathbf{B}_t \,, \qquad t \in \mathbb{Z} \,. \quad (2.12)$$

Kesten [84] also provided the corresponding theory for stationarity and tails in this case. Stărică [107] dealt with the corresponding problems for CCC-GARCH processes, making use of the theory in Kesten [84], Bougerol and Picard [26] and its specification to the tails of GARCH models in Basrak et al. [14]. Stărică [107] assumed the Kesten conditions for the matrices $\mathbf{A}_t$. These conditions ensure that the product matrices $\mathbf{A}_1 \cdots \mathbf{A}_n$ have positive entries for sufficiently large n. Then Kesten's theory implies that all components of the vector $\mathbf{X}_t$ have power-law tails with the same index α and also that the finite-dimensional distributions of the process $(\mathbf{X}_t)$ are regularly varying with index α.

Various GARCH modifications are derived by considering linear combinations of CCC-GARCH models. The property of multivariate regular variation of multivariate GARCH ensures that, after linear transformations, the new process in all components has again power-law tails with the same index as the original GARCH process; see Basrak et al. [14]. Models which are constructed in this way are the Orthogonal GARCH model of Alexander and Chibumba [2], its generalization GO-GARCH by van der Weide [113], the Full Factor GARCH model of Vrontos et al. [116] and the Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen [87]. These models are characterized by their treatment of each series as a linear combination of factors, and each of the factors is modeled as a GARCH process; see Silvennoinen and Teräsvirtä [104].

Not all choices of α- and β-parameters in the model (2.11) allow for an application of the Kesten theory. For example, assume that only the diagonal elements α_{ii} and β_{ii} are positive. Then $\mathbf{A}_t$ is diagonal and, hence, the condition that $\mathbf{A}_1 \cdots \mathbf{A}_n$ have positive entries for sufficiently large n cannot be satisfied. In the latter situation, both $(X_{1,t})$ and $(X_{2,t})$ are univariate GARCH processes. Assuming the conditions of the univariate Kesten-Goldie theorem for each component process, $(X_{1,t})$ and $(X_{2,t})$ have power-law tails with indices κ_1 and κ_2, respectively, given by the solutions to the equations $\mathbb{E}[(\alpha_{ii} Z_{i,t}^2 + \beta_{ii})^{\kappa_i/2}] = 1$, $i = 1, 2$. In this model, one can introduce dependence between the two component series $(X_{1,t})$ and $(X_{2,t})$ by assuming dependence between the noise variables $Z_{1,t}$ and $Z_{2,t}$. Another situation when the Kesten theory fails appears when $\mathbf{A}_t$ is an upper or lower triangle matrix: then the products $\mathbf{A}_1 \cdots \mathbf{A}_n$ are always of the same triangular type. Similar remarks apply when one considers a CCC model in general dimension. Of course, one may argue that the latter models are not natural: they are degenerate since they do not allow for a linear relationship between all squared volatilities on a given day.

2.3.2 A utility based argument for equal tail-indices

In this section we give an argument based on economic theory that suggests equality of tail-indices for equity return series. We follow an approach by Routledge and Zin [99] who introduced the notion of Generalized Disappointment Aversion (GDA). We consider the risky payoff C of an investor and assume that it has a continuous distribution on $(0,\infty)$ with distribution function F_C. Let u be a utility function assumed to be increasing and concave on $(0,\infty)$. Following Routledge and Zin [99], the utility of an agent with GDA preferences is given by

$$\widetilde{u} \quad = \quad \mathbb{E}[u(C)] - b\int_0^{\delta v} \big[u(\delta\, v) - u(x)\big] F_C(dx)\,,$$

where δ and v are positive constants, and $b \geq 0$. Here v can be thought of as the *certainty payoff* equivalent to the risky payoff C; δ tunes the *threshold of disappointment* in proportion to v; b determines the extra weight given to the expected return of C when C is below the disappointment threshold δv. If $b = 0$, preferences are the classical expected utility. If $\delta = 1$ and $b > 0$ preferences follow Gul's [66] disappointment aversion which were generalized by Routledge and Zin [99].

An agent guided by the utility function u will seek to maximize the functional $\widetilde{u}$. Routledge and Zin assumed a power-law utility function

$$u(x) = -\frac{1}{\xi}\, x^{-\xi}\,, \quad \xi > 0\,. \tag{2.13}$$

For the sake of argument, we assume that an investor initially has one unit of wealth. He invests $1-\phi \in (0,1)$ units in a risk-free bond with interest rate $r > 0$ and ϕ units in a risky asset with return X over one time unit, i.e.,

$$C(X) = (1-\phi)\,\mathrm{e}^{\,r} + \phi\,\mathrm{e}^{\,X}\,. \tag{2.14}$$

Then we have

$$\widetilde{u}(F_X,\phi) \quad = \quad \mathbb{E}[u(C)] + b\,\mathbb{E}\big[u(C)\mathbf{1}_{\{(\}}C \leq \delta v)\big] - b\,u(\delta\, v) F_X(q)\,,$$

where F_X is the distribution function of X and

$$q = \log\Big(\mathrm{e}^{\,r} + \frac{\delta\, v - \mathrm{e}^{\,r}}{\phi}\Big)\,.$$

Note that $C \leq \delta v$ if and only if $X \leq q$.

Naturally, if an agent invests in a risky asset instead of a riskless bond, he expects to obtain a higher (on average) return from the risky asset than he is guaranteed from the riskless bond. In our notation, this means $\delta\, v > \mathrm{e}^{\,r}$ or $q > r$. For given b, δ, v, the functional $\widetilde{u}$ depends only on ϕ and F_X, $\widetilde{u} = \widetilde{u}(F_X, \phi)$. We assume that

$$\widetilde{u}_{\max} = \widetilde{u}_{\max}(F_X) = \max_{0\le\phi\le 1} \widetilde{u}(F_X, \phi)$$

exists and that the maximum is achieved at a unique $\hat{\phi} \in (0, 1)$.

2.3.3 Pareto-distributed returns

Since we are interested in the influence of heavy-tailed losses on the preferences of an investor we assume the following toy model. We consider the case when X has a two-sided Pareto distribution given by

$$F_X(x) = \begin{cases} p\left(\frac{K}{K-x}\right)^{\alpha} & x \le 0\,, \\ 1-(1-p)\left(\frac{K'}{K'+x}\right)^{\beta} & x > 0\,, \end{cases} \tag{2.15}$$

where $\alpha, \beta > 0$, $K, K' > 0$, $0 < p < 1$. We also write f_X for the density function of F_X.

We have

$$\begin{aligned} \widetilde{u}(F_X, \phi) \quad = \quad & \alpha\, K^{\alpha}\, p \int_{-\infty}^{0} u\big((1-\phi)\mathrm{e}^{\,r} + \phi \mathrm{e}^{\,x}\big)\, \frac{1+b}{(K-x)^{\alpha+1}}\, dx \\ & +\beta\,(K')^{\beta}\,(1-p) \int_{0}^{\infty} u\big((1-\phi)\mathrm{e}^{\,r} + \phi\, \mathrm{e}^{\,x}\big)\, \frac{1+b\mathbf{1}_{\{(\}}x<q)}{(K'+x)^{\beta+1}}\, dx \\ & -b\, u(\delta v)\, F_X(q)\,. \end{aligned} \tag{2.16}$$

We observe the following property whose proof is given in Appendix 2.C.

Lemma 2.1. *Assume the two-sided Pareto model* (2.15)*, that there is no functional relationship between α, K and β, K' and the utility function u is increasing and differentiable. Then $\frac{\partial \widetilde{u}_{\max}}{\partial \alpha} > 0$ and $\frac{\partial \widetilde{u}_{\max}}{\partial K} < 0$.*

We conclude that $\widetilde{u}_{\max}$ increases with α and decreases with K. Therefore there is a curve of equal preference on the (α, K)-plane. Moving along this curve in the direction of increasing α, one expects the values of K to increase too, i.e., the estimated values of α and K should appear positively dependent. Figure 2.8 illustrates this scenario for $\xi = 1/2, 4$ for the power-utility function (2.13). In fact, this positive dependence is indeed observed for some real return data, e.g. the “Energy”, “Consumer Staples” and “Information Technology” sectors of the S&P 500 index; see Figure 2.2.

A particularly interesting case occurs when F_X is symmetric, i.e., when $\alpha = \beta$, $K = K'$ and $p = 0.5$. Then (2.15) turns into

$$\begin{aligned} \widetilde{u}(F_X, \phi) \quad = \quad & \frac{\alpha}{2} K^{\alpha}(1+b) \int_0^{\infty} \frac{u(C(x))\left[1-\frac{b}{1+b}\mathbf{1}_{\{x\ge q\}}\right] + u(C(-x))}{(K+x)^{\alpha+1}} dx \\ & -bu(\delta v) F_X(q)\,. \end{aligned} \tag{2.17}$$

This situation is not covered by Lemma 2.1: for the proof of the latter result we used Lemma 2.2 whose assumptions are not satisfied in the present situation. Indeed, the integrand

$$U_{\mathrm{all}}(x) = u(C(x))\left[1 - \frac{b}{1+b}\mathbf{1}_{\{(\}}x \ge q)\right] + u(C(-x))$$

is not monotone.[2]

We resort to numerical methods to gain some understanding of how $\tilde{u}(F_X, \phi)$ changes with α. The value of $\hat{\phi}$ can be calculated by numerical integration and optimization with respect to ϕ for given values of K, K', α, β. This is shown in Figure 2.7 for the power utility function (2.13) both for fixed K', β and for $K = K'$, $\alpha = \beta$. The corresponding values $\tilde{u}_{\max}(\alpha, K)$ are shown in Figure 2.8. If K' and β are fixed, both $\widetilde{u}_{\max}(\alpha, K)$ and $\hat{\phi}$ increase with α and decrease with K. This is in agreement with Lemma 2.1. In contrast, when $K = K'$ and $\alpha = \beta$ $\widetilde{u}_{\max}(\alpha, K)$ decreases with α but is rather insensitive with respect to K. On the other hand, $\hat{\phi}$ is not monotone with respect to α or K. For each fixed K, it peaks at an α-value somewhere below 1. For realistic values $\alpha \in (2, 4)$, $\hat{\phi}$ is a small value below 5%. Since $\widetilde{u}_{\max}(\alpha, K)$ decreases with α, investors who seek to maximize $\widetilde{u}_{\max}(\alpha, K)$ will prefer the smallest α in the market, resulting in similar values of α for different equities.

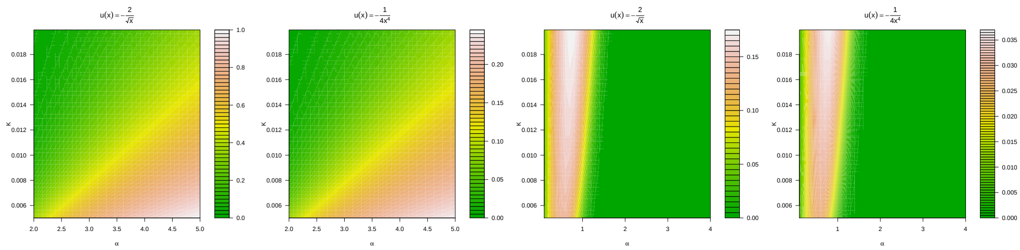


Figure 2.7: *The 1st and 2nd* graphs show $\hat{\phi}$, the optimal equity allocation as a function of α and K in the two-sided Pareto model (2.15) for fixed $K' = 0.012$, $\beta = 1.4$. *The 3rd and 4th* graphs show $\hat{\phi}$ as a function of α and K with $\beta = \alpha$ and $K' = K$. We choose the utility function u from (2.13) for $\xi = 1/2$ and $\xi = 4$, $b = 0.01$ in all cases.

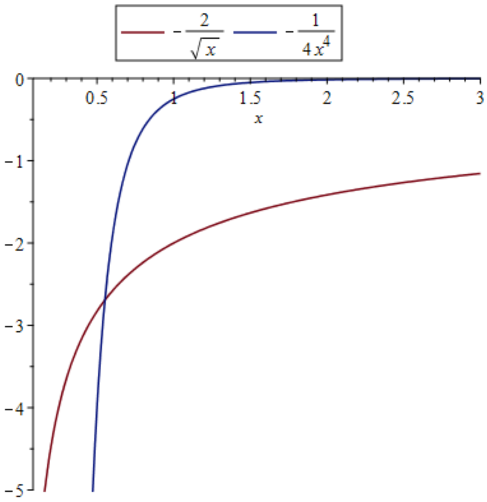
[2]To see this we may plug (2.13) in U_{all} and re-write it as

$$U_{\text{all}}(x) = -\frac{\phi^{-\xi}}{\xi}\Bigg\{ \underbrace{\left(1 - \frac{b}{1+b}\mathbf{1}_{\{(\}}x \geq q)\right)\Big[a + e^x\Big]^{-\xi} + \Big[a + e^{-x}\Big]^{-\xi}}_{U(x)} \Bigg\},$$

where $a = (1-\phi)e^r/\phi$. Direct computation gives

$$U'(x) \quad = \quad -\frac{(a+e^x)^{-\xi-1}\xi e^x}{1 + b\mathbf{1}_{\{(\}}x \geq q)} + (a + e^{-x})^{-\xi-1}\xi e^{-x}, \qquad x \neq q$$

The function $U(x)$, hence U_{all}, is not monotone because $U'(x) > 0$ for all large x while $U(x)$ decreases in a small neighborhood of q.



The economitrical differences between the two utility functions are illustrated in the graph to the left. We see that $u(x)$ grows slower and saturates later for small ξ than for large ξ. The latter case represents an agent who is more tolerant about low consumption and seeks wealth more aggressively. In other words, he is less risk-averse than one with a larger ξ. Such an agent will therefore invest more heavily in equity, as shown in Figure 2.7. Note that, while $\hat{\phi}$ changes nearly in the same way in both graphs, the scales in the graphcs are very different.

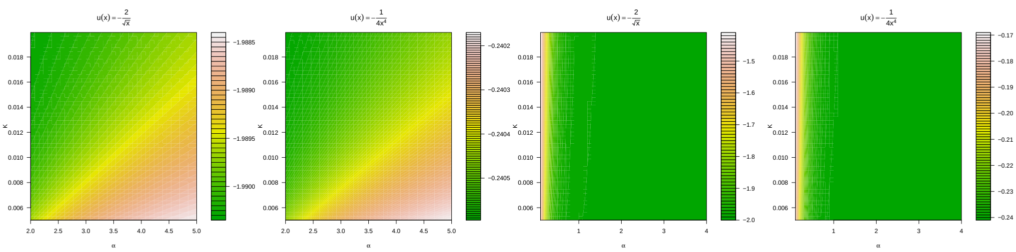


Figure 2.8: *The 1st and 2nd* graphs show $\tilde{u}_{\max}(\alpha, K)$, as a function of α and K in the two-sided Pareto model (2.15) with $K' = 0.012$, $\beta = 1.4$. Clearly, $\tilde{u}_{\max}(\alpha, K)$ increases with α and decreases with K when K' and β are fixed. *The 3rd and 4th* graphs show $\tilde{u}_{\max}(\alpha, K)$ as a function of α and K with $\beta = \alpha$ and $K' = K$. We choose the utility function u from (2.13) for $\xi = 1/2$ and $\xi = 4$. $b = 0.01$ in all cases.

If the parameter b is very small the GDA preference is closely approximated by the mean-utility preference, corresponding to $b = 0$. In this case, we show in the proof of Lemma 2.5 that the function $U_{\text{all}}(x)$ may increase or decrease depending on particular conditions on the values of ξ and $(1 - \phi)e^r/\phi$:

1. If $\max\{a, 1\} < \xi$, $U_{\text{all}}(\cdot)$ is monotone decreasing.
2. If $a < \xi < 1$ and $(a + y_-)/(ay_- + 1) < y_-^{(1-\xi)/(1+\xi)}$, $U_{\text{all}}(\cdot)$ is monotone decreasing.
3. If $\xi < a < 1$, $U_{\text{all}}(\cdot)$ is monotone increasing.
4. If $1 < \xi < a$ and $(a + y_+)/(ay_+ + 1) > y_+^{(1-\xi)/(1+\xi)}$, $U_{\text{all}}(\cdot)$ is monotone increasing.
5. In other cases, $U_{\text{all}}(\cdot)$ is not monotone.

where

$$\begin{aligned} a &= \frac{(1-\phi)e^r}{\phi} \\ y_\pm &= \frac{a^2 - \xi \pm \sqrt{(a^2-1)(a^2-\xi^2)}}{a(\xi-1)} \end{aligned}$$

Moreover, it is also easily checked that, when $x \in (0, K(e^{1/\alpha}-1)]$, the density function in the integral of (2.17) increases with α; when $x \in (K(e^{1/\alpha}-1), \infty)$ it decreases with α. Following the arguments for Lemma 2.1, and applying Lemma 2.2, it can be seen that $\widetilde{u}_{\max}(\alpha, K)$ increases/decreases with α when $U_{\mathrm{all}}(x)$ decreases/increases.

2.4 Conclusion

We have established that, in the case of an equity return series with two-sided, functionally independent Pareto tails, investor preference functionals are monotone increasing/decreasing with the tail index/scale parameters. Thus in a market dominated by such equities, the investors would pursue the largest tail index in the market, leading to a shared common tail index for all equities.

The empirical results presented in section 2.2 suggest this may well be the case for the "Consumer Staples" sector of S&P 500, given the Hill estimates of tail indices shown in figure 2.1 and the largely positive results of tests for equal tail indices shown in figure 2.3.

On the other hand, we have also seen that, when the left and the right tails have the same indices, investor preference over the equity has more sophisticated variations in the parameters' space including the tail parameters of the equity, the interest rate, the investor's risk apetite as captured by his utility function, and his threshold of disappointment.

We also acknowledge that our model of the market and the investor is a simple one, not accoounting for the dependence between equities, nor the categorization of investors and their interactions. These are potential topics of future work.

2.A A monotonicity lemma

Lemma 2.2. *Assume distribution function $F(x,\theta)$ parameterized by $\theta \in \Theta \subseteq \mathbb{R}$ has support $(a,b) \subseteq \mathbb{R}$, and in addition $F(x,\theta)$ has density function $f(x,\theta)$ that is differentiable with respect to θ for all $\theta \in \Theta$. Let $X \sim F$ and assume function $h(\cdot)$ is defined on (a,b) and is monotone throughout this interval. Moreover, we assume $h(x)$ and $f(x,\theta)$ satisfy*

$$\int_a^b \left| \frac{\partial f(x,\theta)}{\partial \theta} \right| dx < \infty \text{ and } \int_a^b \left| h(x) \frac{\partial f(x,\theta)}{\partial \theta} \right| dx < \infty \tag{2.18}$$

Then the following holds true:

1. *If $h(\cdot)$ is decreasing and $\exists x_0 \in (a,b)$ such that $\frac{\partial f}{\partial \theta}(x,\theta) > 0$ for $x \in (a, x_0)$ while $\frac{\partial f}{\partial \theta}(x,\theta) < 0$ for $x \in (x_0, b)$, then*

$$\frac{\partial \mathbb{E} h(X)}{\partial \theta} > 0$$

2. *If $h(\cdot)$ is increasing and $\exists x_0 \in (a,b)$ such that $\frac{\partial f}{\partial \theta}(x,\theta) < 0$ for $x \in (a, x_0)$ while $\frac{\partial f}{\partial \theta}(x,\theta) > 0$ for $x \in (x_0, b)$, then*

$$\frac{\partial \mathbb{E}h(X)}{\partial \theta} > 0$$

Remark 2.3. *Two other cases follow trivially from lemma 2.2:*

1. *If $h(\cdot)$ is increasing and $\frac{\partial f}{\partial \theta}$ satisfies the same conditions of the 1st case of lemma 2.2, $\frac{\partial \mathbb{E}h(X)}{\partial \theta} < 0$. This immediately follows from applying 1st case of lemma 2.2 to $-h(\cdot)$.*
2. *By the same argument, if $h(\cdot)$ is decreasing and $\frac{\partial f}{\partial \theta}$ satisfies the same conditions of the 2nd case of lemma 2.2, $\frac{\partial \mathbb{E}h(X)}{\partial \theta} < 0$.*

Proof. Firstly, by dominated convergence theorem, conditions (2.18) imply, for all $S \subseteq (a,b)$,

$$\begin{aligned} \frac{\partial}{\partial \theta}\int_S f(x,\theta)dx &= \int_S \frac{\partial}{\partial \theta} f(x,\theta)dx \\ \frac{\partial}{\partial \theta}\int_S h(x)f(x,\theta)dx &= \int_S h(x)\frac{\partial}{\partial \theta} f(x,\theta)dx \end{aligned}$$

Thus we have

$$\begin{aligned} \frac{\partial \mathbb{E}h(X)}{\partial \theta} &= \frac{\partial}{\partial \theta}\int_a^b h(x)f(x,\theta)dx \\ &= \int_a^b h(x)\frac{\partial f}{\partial \theta}(x,\theta)dx \\ &= \underbrace{\int_a^{x_0} h(x)\frac{\partial f}{\partial \theta}(x,\theta)dx}_{I_1} + \underbrace{\int_{x_0}^b h(x)\frac{\partial f}{\partial \theta}(x,\theta)dx}_{I_2} \end{aligned}$$

x_0 being located in the interior of (a,b) and $h(\cdot)$ being monotone imply $h(x_0) < \infty$.

1. When $h(x)$ is decreasing on (a,b) and $\frac{\partial f}{\partial \theta}(x,\theta) > 0$ on (a, x_0)

$$I_1 > h(x_0)\int_a^{x_0} \frac{\partial f}{\partial \theta}(x,\theta)dx$$

Similarly, because $\frac{\partial f}{\partial \theta}(x,\theta) < 0$ for $x \in (x_0, b)$ and $h(x)$ is decreasing, we have

$$I_2 = \int_{x_0}^b -h(x_0)\left|\frac{\partial f}{\partial \theta}(x,\theta)\right| dx > -h(x_0)\int_{x_0}^b \left|\frac{\partial f}{\partial \theta}(x,\theta)\right| dx$$

Finally we have

$$\frac{\partial \mathbb{E}h(X)}{\partial \theta} > h(x_0)\int_a^b \frac{\partial f}{\partial \theta}(x,\theta)dx = h(x_0)\frac{\partial}{\partial \theta}\int_a^b f(x,\theta)dx = 0$$

2. If $h(\cdot)$ is increasing and $\exists x_0 \in (a,b)$ such that $\frac{\partial f}{\partial \theta}(x_0,\theta) < 0$ on (a,x_0) while $\frac{\partial f}{\partial \theta}(x_0,\theta) > 0$ on (x_0,b), by similar arguments, one can show

$$\frac{\partial \mathbb{E}h(X)}{\partial \theta} > 0$$

□

2.B When equity returns follow Student's t-distribution

It is a common practice to use Student's t-distribution to model the stationary distribution of equity returns. So it is of interest to find out what implications this distribution has when it is combined with the PDA preference. Formally we assume

$$f(x;\alpha) = c(\alpha)\left(1+\frac{x^2}{\alpha}\right)^{-(\alpha+1)/2}$$

where $\alpha > 1$ and

$$c(\alpha) = \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\alpha/2)\sqrt{\alpha\pi}}$$

In the same way as for (2.17), we can write $\widetilde{u}(F,\phi)$ as

$$\begin{aligned}\widetilde{u}(F,\phi) &= (1+b)\int_0^\infty \underbrace{\left\{u(C(x))\left[1-\frac{b}{1+b}\mathbf{1}_{\{x\geq q\}}\right]+u(C(-x))\right\}}_{U_{\text{all}}} f(x,\alpha)dx \\ &\quad -bu(\delta v)F_X(q) \end{aligned} \tag{2.19}$$

where $C(\cdot)$ is defined in (2.14). As shown in lemma 2.5, when $b=0$ and $u(\cdot)$ takes the power-form of (2.13), U_{all} is monotone depending on the values of ξ and $(1-\phi)e^r/\phi$. As given in lemma 2.4, there is a point $x_0 > 0$ such that $\frac{\partial f}{\partial \alpha}(x_0,\alpha) = 0$ and $\forall x \in (0,x_0)$, $\frac{\partial f}{\partial \alpha}(x,\alpha) > 0$ and $\forall x \in (x_0,\infty)$, $\frac{\partial f}{\partial \alpha}(x,\alpha) < 0$. Thus it remains to verify $\int_0^\infty |\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$ and $\int_0^\infty |U_{\text{all}}(x)\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$ if we are to apply lemma 2.2.

As computed in the proof of lemma 2.4, $\frac{\partial f}{\partial \alpha}(x,\alpha)$ is given by (2.20). It is also shown there $\frac{d}{d\alpha}c(\alpha) > 0$. Thus for $\int_0^\infty |\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$ it suffices to show

$$\int_0^\infty \frac{x^2}{(x^2+\alpha)(1+x^2/\alpha)^{\alpha/2+1/2}}dx < \infty$$

We may write

$$\begin{aligned}&\int_0^\infty \frac{x^2}{(x^2+\alpha)(1+x^2/\alpha)^{\alpha/2+1/2}}dx \\ =& \left(\int_0^1+\int_1^\infty\right)\frac{x^2}{(x^2+\alpha)(1+x^2/\alpha)^{\alpha/2+1/2}}dx \\ =& I_1+I_2\end{aligned}$$

Clearly

$$I_1 < \int_0^1 \frac{1}{\alpha}dx < \infty$$

while

$$\begin{aligned} I_2 &= \int_1^\infty \frac{1}{1+\alpha/x^2} \frac{1}{(1/x^2+1/\alpha)^{(\alpha+1)/2}} \frac{dx}{x^{\alpha+1}} \\ &< \int_1^\infty \alpha^{(\alpha+1)/2} \frac{dx}{x^{\alpha+1}} < \infty \end{aligned}$$

So we conclude $\int_0^\infty |\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$. To see $\int_0^\infty |U_{\text{all}}(x)\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$, we note

$$\begin{aligned} \xi|U_{\text{all}}(x)| &< [(1-\phi)e^r + e^x]^{-\xi} + [(1-\phi)e^r + e^{-x}]^{-\xi} \\ &< 1 + (1-\phi)^{-\xi} e^{-r\xi} \end{aligned}$$

Since $\int_0^\infty |\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$, it follows from the above inequality $\int_0^\infty |U_{\text{all}}(x)\frac{\partial f}{\partial \alpha}(x,\alpha)|dx < \infty$. Thus by lemma 2.2, $\widetilde{u}_{\max}$ is monotone increasing/decreasing with α when $U_{\text{all}}(\cdot)$ is monotone decreasing/increasing. Accordingly, an investor guided by the utility function will seek the largest/smallest α observed in the market.

If however $b > 0$, Lemma 2.2 is not applicable anymore. Nonetheless, numerical analysis lends some insight. As shown in figure 2.9, $\hat{\phi}$ is monotone increasing for all 4 values of b, while $\tilde{u}_{\max}(\alpha)$ is increasing with α when b is relatively large, but decreasing with α when b is small. We note that a sizable value of b indicates a conservative, risk-averse investor.

Lemma 2.4. *Let f denotes the density function of the Student's t-distribution, i.e.*

$$f(x;\alpha) = c(\alpha)\left(1+\frac{x^2}{\alpha}\right)^{-(\alpha+1)/2}$$

where $\alpha > 1$ and

$$c(\alpha) = \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\alpha/2)\sqrt{\alpha\pi}}$$

Then there exists $x_0 > 0$ such that $\frac{\partial f}{\partial \alpha}(x_0,\alpha) = 0$ and $\forall x \in (0,x_0)$, $\frac{\partial f}{\partial \alpha}(x,\alpha) > 0$ and $\forall x \in (x_0,\infty)$, $\frac{\partial f}{\partial \alpha}(x,\alpha) < 0$.

Proof. Straightforward computation gives

$$\begin{aligned} \frac{\partial f(x,\alpha)}{\partial \alpha} &= \frac{c(\alpha)x^2(\alpha+1) + (2\alpha x^2 + 2\alpha^2)c'(\alpha) - \alpha c(\alpha)(x^2+\alpha)\log(1+x^2/\alpha)}{2\alpha(x^2+\alpha)(1+x^2/\alpha)^{\alpha/2+1/2}} \\ &:= \frac{P(x^2,\alpha)}{2\alpha(x^2+\alpha)(1+x^2/\alpha)^{\alpha/2+1/2}} \end{aligned} \tag{2.20}$$

While the denominator of the right side of $\frac{\partial f(x,\alpha)}{\partial \alpha}$ is always positive, its numerator $P(x^2,\alpha)$ has a single root:

$$x_0^2 = \alpha \exp\left\{W\left[-\left(1+\frac{1}{\alpha}\right)e^{-1-2c'(\alpha)/c(\alpha)-1/\alpha}\right] + 1 + \frac{1}{\alpha} + \frac{2c'(\alpha)}{c(\alpha)}\right\} - \alpha \tag{2.21}$$

where $W(\cdot)$ is the principle branch of the Lambert W function. and $c'(\cdot)$ is the derivative of $c(\cdot)$. To check the right side of (2.21) for positivity, we first note $c'(\alpha) > 0$:

$$c'(\alpha) = \frac{\pi\Gamma(\alpha/2+1/2)\left\{\alpha\left[\Psi(\alpha/2+1/2) - \Psi(\alpha/2)\right] - 1\right\}}{2\Gamma(\alpha/2)(\pi\alpha)^{3/2}}$$

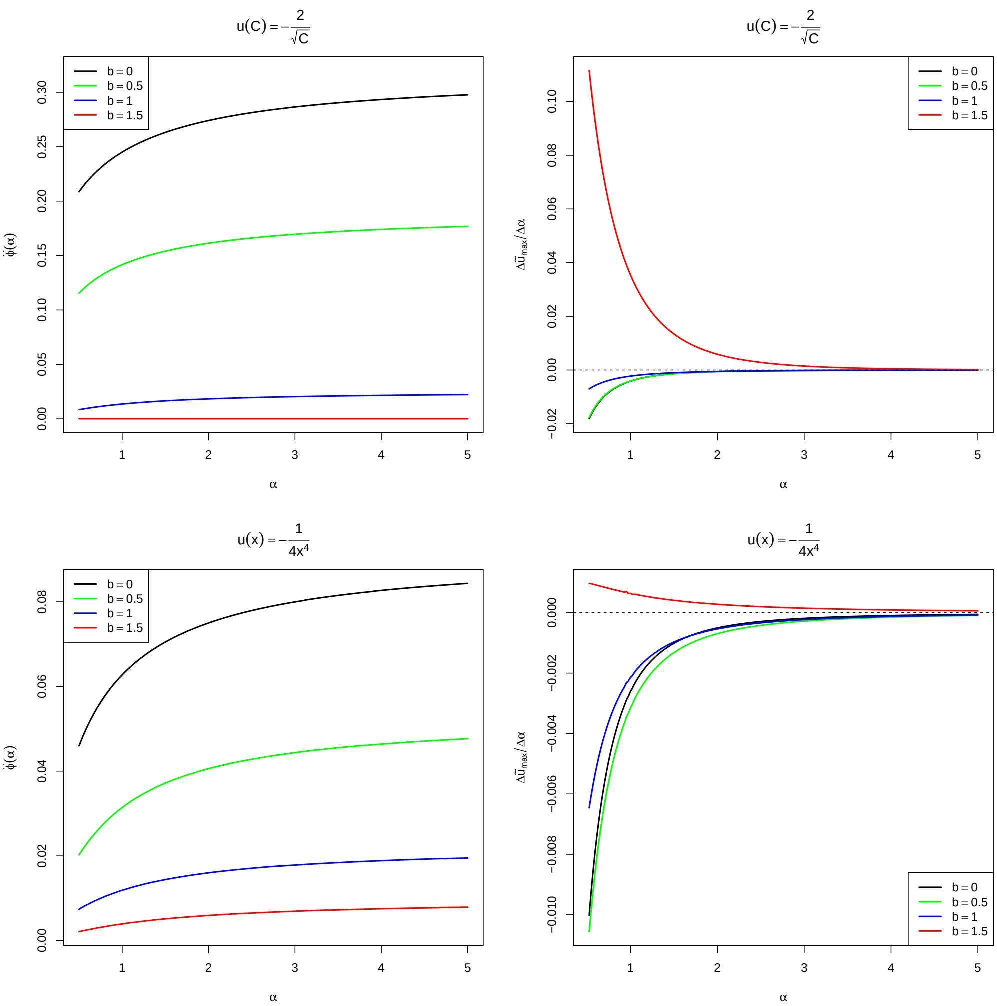
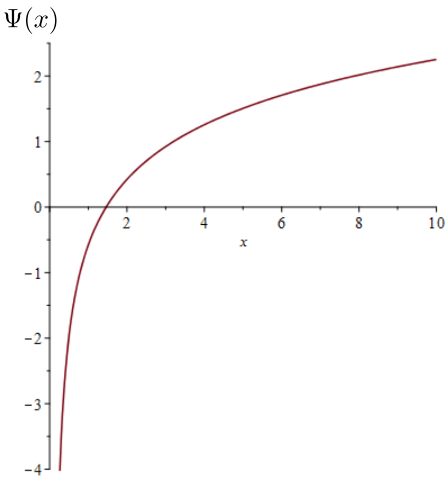


Figure 2.9: $\hat{\phi}$ (left) and $\frac{\partial \tilde{u}_{\max}}{\partial \alpha}$ (right). *top:* $\xi = 1/2$. *bottom:* $\xi = 4$.

where $\Psi(\cdot)$ is the digamma function:

$$\Psi(x) = \frac{d \log[\Gamma(x)]}{dx}$$



As shown in the figure to the left, $\Psi(x)$ is increasing for $x > 0$. This immediately follows from the series representation

$$\Psi(x+1) = -\gamma + \sum_{n=1}^{\infty} \frac{x}{n(n+x)} \qquad x \neq -1, -2, -3, \ldots$$

which gives

$$\frac{\partial \Psi(x+1)}{\partial x} = \sum_{n=1}^{\infty} \frac{1}{(n+x)^2} > 0$$

See Abramowitz and Stegun [1], p.259, formula 6.3.16. Therefore $\Psi(\alpha/2+1/2) - \Psi(\alpha/2) > 0$. So we have

$$\begin{aligned} & \alpha\left[\Psi(\alpha/2+1/2) - \Psi(\alpha/2)\right] - 1 \\ \geq\ & 1 \times \left[\Psi(1/2+1/2) - \Psi(1/2)\right] - 1 \\ =\ & \log(4) - \log(e) \\ >\ & 0 \end{aligned}$$

Thus $c'(\alpha) > 0$. Furthermore, we recall $W(\cdot)$ is increasing on its principle branch. So

$$\begin{aligned} & W\left[-\left(1+\frac{1}{\alpha}\right)e^{-1-2c'(\alpha)/c(\alpha)-1/\alpha}\right] + 1 + \frac{1}{\alpha} + \frac{2c'(\alpha)}{c(\alpha)} \\ >\ & W\left[-\left(1+\frac{1}{\alpha}+\frac{2c'(\alpha)}{c(\alpha)}\right)e^{-1-2c'(\alpha)/c(\alpha)-1/\alpha}\right] + 1 + \frac{1}{\alpha} + \frac{2c'(\alpha)}{c(\alpha)} \\ =\ & W(-ye^{-y}) + y \end{aligned}$$

where

$$y = 1 + \frac{1}{\alpha} + \frac{2c'(\alpha)}{c(\alpha)} > 1$$

Now notice

$$\log(ye^{-y}) = \log(y) - y$$

is a decreasing function for $y > 1$. Thus $-ye^{-y}$ is an increasing function. Hence we have

$$W(-ye^{-y}) + y \quad > \quad W(-e^{-1}) + 1 = 0$$

Now it is clear

$$\alpha \exp\left\{W\left[-\left(1+\frac{1}{\alpha}\right)e^{-1-2c'(\alpha)/c(\alpha)-1/\alpha}\right] + 1 + \frac{1}{\alpha} + \frac{2c'(\alpha)}{c(\alpha)}\right\} - \alpha > 0$$

Now that we have established that $\frac{\partial f}{\partial \alpha}(x,\alpha) = 0$ has a single positive root, it remains to determine the sign of $\frac{\partial f}{\partial \alpha}(x,\alpha)$ on the two sides of the root. For this purpose we observe

$$P(0,\alpha) = 2\alpha^2 c'(\alpha) > 0 \qquad (2.22)$$

So we want to investigate $\frac{\partial P}{\partial x}(x,\alpha)$:

$$\frac{\partial P}{\partial x}(x,\alpha) \quad = \quad 2\alpha c'(\alpha) + c(\alpha) - \alpha c(\alpha)\log\left(1+\frac{x}{\alpha}\right) \tag{2.23}$$

Clearly, $\frac{\partial P}{\partial x}(0,\alpha) > 0$. Hence from (2.22) and (2.23) it is clear

$$\text{sign}\left[\frac{\partial f}{\partial \alpha}(x,\alpha)\right] = \begin{cases} 1 & 0 < x < x_0 \\ -1 & x > x_0 \end{cases}$$

where x_0 is the positive root of (2.21). □

2.C Proof of Lemma 2.1

Proof. Let

$$\hat{\phi} := \underset{0<\phi\leq 1}{\text{argmax}}\, \widetilde{u}(F_X,\phi) \tag{2.24}$$

We have

$$\tilde{u}_{\max}(F_X) = \widetilde{u}(F_X,\hat{\phi})$$

It follows

$$\frac{d\tilde{u}_{\max}(F_X)}{d\alpha} \quad = \quad \left.\frac{\partial\widetilde{u}(\alpha,\phi)}{\partial\alpha}\right|_{\phi=\hat{\phi}} + \left.\frac{\partial\widetilde{u}(\alpha,\phi)}{\partial\phi}\right|_{\phi=\hat{\phi}}\frac{\partial\hat{\phi}}{\partial\alpha} \tag{2.25}$$

The definition (2.24) implies for all α

$$\left.\frac{\partial\widetilde{u}(\alpha,\phi)}{\partial\phi}\right|_{\phi=\hat{\phi}} = 0 \tag{2.26}$$

So the second term of (2.25) vanishes. It remains to show the first term is positive. From (2.16), it follows

$$\frac{\partial\widetilde{u}(\alpha,\phi)}{\partial\alpha} = \frac{\partial}{\partial\alpha}\mathbb{E}[u((1-\phi)e^r+\phi e^x)\mathbf{1}_{\{X<0\}}]$$

The function $u((1-\phi)e^r+\phi e^x)$ is obviously increasing with x. It follows

$$\frac{\partial f_X(x;\alpha,K)}{\partial\alpha} = \frac{\partial}{\partial\alpha}\frac{\alpha K^\alpha}{(K-x)^{\alpha+1}} = -\frac{K^\alpha}{(K-x)^{\alpha+1}}\left[\alpha\log\left(1-\frac{x}{K}\right)-1\right]$$

It is easily checked

$$\frac{\partial}{\partial\alpha}\frac{\alpha K^\alpha}{(K-x)^{\alpha+1}}\begin{cases} >0 & \text{when } x < K(1-e^{1/\alpha}) < 0 \\ <0 & \text{when } K(1-e^{1/\alpha}) < x < 0 \end{cases}$$

This is the second case of lemma 2.2. So we have $\frac{\partial\widetilde{u}(\alpha,\phi)}{\partial\alpha} > 0$. As for $\frac{\partial\tilde{u}_{\max}(F_X)}{\partial K}$, by the same argument, it suffices to show $\frac{\partial\widetilde{u}(K,\phi)}{\partial K} < 0$. We have

$$\frac{\partial\widetilde{u}(K,\phi)}{\partial K} = \frac{\partial}{\partial K}\mathbb{E}\left[u((1-\phi)e^r+\phi e^x)\mathbf{1}_{\{X<0\}}\right]$$

and

$$\frac{\partial f_X(x;\alpha,K)}{\partial K} = \frac{\partial}{\partial K}\frac{\alpha K^\alpha}{(K-x)^{\alpha+1}} = -\alpha K^{\alpha-1}\frac{\alpha x+K}{(K-x)^{\alpha+2}}$$

Clearly,

$$\frac{\partial}{\partial K}\frac{\alpha K^\alpha}{(K-x)^{\alpha+1}}\begin{cases} >0 & \text{when } x<-K/\alpha<0 \\ <0 & \text{when } -K/\alpha<x<0\end{cases}$$

So by the 1st case of Remark 2.3 we conclude $\frac{\partial \widetilde{u}(K,\phi)}{\partial K}<0$. □

2.D Lemma 2.5

Lemma 2.5. *Let $u(\cdot)$ and $C(\cdot)$ be defined as in (2.13) and (2.14) respectively. Define*

$$\begin{aligned} U_{all} &= u(C(x))+u(C(-x)) \quad x\geq 0 \\ a &= \frac{(1-\phi)e^r}{\phi} \\ y_\pm &= \frac{a^2-\xi\pm\sqrt{(a^2-1)(a^2-\xi^2)}}{a(\xi-1)} \end{aligned}$$

The following holds true:

1. *If $\max\{a,1\}<\xi$, $U_{all}(\cdot)$ is monotone decreasing.*
2. *If $a<\xi<1$ and $(a+y_-)/(ay_-+1)<y_-^{(1-\xi)/(1+\xi)}$, $U_{all}(\cdot)$ is monotone decreasing.*
3. *If $\xi<a<1$, $U_{all}(\cdot)$ is monotone increasing.*
4. *If $1<\xi<a$ and $(a+y_+)/(ay_++1)>y_+^{(1-\xi)/(1+\xi)}$, $U_{all}(\cdot)$ is monotone increasing.*
5. *In other case, $U_{all}(\cdot)$ is not monotone.*

Proof. It is convenient to re-write U_{all} as

$$U_{\text{all}} = -\frac{\phi^{-\xi}}{\xi}\left\{\underbrace{\left[\frac{(1-\phi)}{\phi}e^r+e^x\right]^{-\xi}+\left[\frac{(1-\phi)}{\phi}e^r+e^{-x}\right]^{-\xi}}_{U(x)}\right\}$$

Thus the monotonicity of $U_{\text{all}}(x)$ is the opposite of $U(x)$. For convenience of writing, let $a=(1-\phi)e^r/\phi$. First of all, we find the conditions on which $U(x)$ is monotone. Direct differentiation yields

$$\frac{\partial U(x)}{\partial x} = -(a+e^x)^{-\xi-1}\xi e^x+(a+e^{-x})^{-\xi-1}\xi e^{-x}$$

$\frac{\partial U(x)}{\partial x}=0$ is equivalent to

$$\begin{aligned} \frac{a+y}{ay+1} &= y^{\frac{1-\xi}{1+\xi}} \\ \underbrace{\log(a+y)-\log(ay+1)}_{f(y)} &= \frac{1-\xi}{1+\xi}\log(y) \end{aligned}$$

where we have defined $y = e^x$, and $f(y)$, $g(y)$ as above. Observe

$$\frac{\partial (f-g)(y)}{\partial y} = \frac{a(\xi-1)y^2 + 2(\xi - a^2)y + a(\xi-1)}{(1+\xi)(a+y)(ay+1)y} \tag{2.27}$$

$\frac{\partial(f-g)(y)}{\partial y} = 0$ has two roots when $\xi \neq 1$

$$y_{\pm} = \frac{a^2 - \xi \pm \sqrt{(a^2-1)(a^2-\xi^2)}}{a(\xi-1)}$$

1. If $\min\{1,\xi\} < a < \max\{1,\xi\}$, $y_{\pm}$ are not real, $\frac{\partial(f-g)(y)}{\partial y} = 0$ has no solution on $(1,\infty)$; $(f-g)(y)$ is monotone on $(1,\infty)$.

   a) If in addition $\xi > 1$, i.e. $1 < a < \xi$, $(f-g)(y)$ is monotone increasing on $(1,\infty)$, because the coefficient of the y^2 term of the numerator of (2.27), i.e. $a(\xi-1)$ is positive. We note $f(1) = g(1) = 0$; thus on $(1,\infty)$, there is no solution to $f(y) = g(y)$. It can be concluded that $U(\cdot)$ is monotone on $(0,\infty)$. Furthermore

   $$\left.\frac{\partial U(x)}{\partial x}\right|_{x=0} = 0$$

   For a small $\epsilon > 0$, the sign of $\frac{\partial U(x)}{\partial x}$ on $(0,\epsilon)$ is thus the same as $\left.\frac{\partial^2 U(x)}{\partial x^2}\right|_{x=0}$:

   $$\left.\frac{\partial^2 U(x)}{\partial x^2}\right|_{x=0} = 2(a+1)^{-\xi-2}(\xi - a) > 0$$

   Thus $\frac{\partial U(x)}{\partial x} > 0$ for $x \in (0,\infty)$; U_{all} is monotone decreasing.

   b) If instead $\xi < 1$, i.e. $\xi < a < 1$, by a similar argument as in the previous case, U_{all} is monotone increasing on $(0,\infty)$.

2. If $a < \min\{1,\xi\}$ and $\xi > 1$, i.e. $a < 1 < \xi$, it is clear

   $$y_- = \frac{a^2 - \xi - \sqrt{(a^2-1)(a^2-\xi^2)}}{a(\xi-1)} < 0$$

   It remains to compare y_+ with 1 to determine whether $\frac{\partial(f-g)(y)}{\partial y} = 0$ has a solution on $(1,\infty)$. Assume $y_+ > 1$. Then

   $$\begin{aligned} a^2 - \xi + \sqrt{(a^2-1)(a^2-\xi^2)} &> a(\xi-1) \\ 2a(\xi-1)(a+1)(a-\xi) &> 0 \end{aligned} \tag{2.28}$$

   This contradicts the assumption $a < \xi$ and $\xi > 1$. Hence $y_+ < 1$. So we conclude $(f-g)(y)$ is monotone increasing on $(1,\infty)$. Following the same analysis as in the case (1.a), one can see U_{all} is monotone decreasing.

3. If $a < \min\{1,\xi\}$ and $\xi < 1$, i.e. $a < \xi < 1$, it is clear $y_- > 0$ and $y_+ < y_-$. Moreover, $y_- > 1$ is equivalent to

   $$\sqrt{(a^2-1)(a^2-\xi^2)} > a^2 - \xi a + a - \xi = (a+1)(a-\xi)$$

The last inequality is obviously true in this case. So $y_- > 1$. $\frac{\partial(f-g)(y)}{\partial y}$ has a maximum at $y_- \in (1,\infty)$. We know $(f-g)(y_-)$ is a maximum because (2.27) shows that, for $y > y_-$, $\frac{\partial(f-g)(y)}{\partial y} < 0$.

If $(f-g)(y_-) < 0$, $(f-g)(y) = 0$ has no solution on $(1,\infty)$. Hence $U(\cdot)$ is monotone on both $(0,\infty)$. The same analysis as in the case (1.a) shows U_{all} is monotone decreasing on $(0,\infty)$.

If $(f-g)(y_-) > 0$, $(f-g)(y) = 0$ must have a solution on (y_-,∞). So $U(\cdot)$ is not monotone.

4. $a > \max\{1,\xi\}$ and $\xi > 1$, i.e. $1 < \xi < a$. By the same argument that leads to (2.28), we see $y_+ > 1$. From (2.27) it is clear $(f-g)(y_+)$ is a minimum. If $(f-g)(y_+) > 0$, there is no solution to $(f-g)(y) = 0$ on $(1,\infty)$; $U(x)$ is monotone. By the same analysis as in the case (1.a), we know $U(x)$ is monotone decreasing and U_{all} is monotone increasing.

   if $(f-g)(y_+) < 0$, there must be a solution to $(f-g)(y) = 0$ on (y_+,∞). U_{all} is not monotone.

□

Chapter 3

Rare event simulation for GARCH(p,q) processes

Jeffrey Collamore, Anand Vidyashankar & Xiaolei Xie
technical report

Abstract

We propose an efficient importance sampling estimator for the rare event probability $\mathbb{P}(|V| > u)$ where V is a vector following the stationary distribution of a GARCH(p, q) process. Recall $V_t = A_t V_{t-1} + B$ is the matrix recurrence equation of a GARCH(p, q) process. We emanate the process $\{V_t\}$ from a set $\mathcal{C} = \{V : |V| \leq M\}$ for some predefined positive constant M, and then we introduce a dual change of measure for the originally iid matrices $\{A_t\}$: for $t < T_u = \min\{i > 0 : |V_i| > u\}$, we exponentially tilt the distribution of $A_t|X_{t-1}$, where $X_0 = V_0$ and $X_t = \prod_{i=1}^{t} X_0 / |\prod_{i=1}^{t} X_0|$, so that $|A_t X_{t-1}|^{\xi}$ is more likely to take on large values and hence $|V_t|$ is more likely to exceed u. Once the exeedance has happened, we change the distribution of $\{A_t\}$ back to the original and continue the process until V_t returns to the set $\mathcal{C}$. Along each simulated path emenating from $\mathcal{C}$ and ending in $\mathcal{C}$ we compute $N_u = \sum_{i=1}^{K} \mathbf{1}_{\{|V_i|>u\}}$, where $K = \min\{i > 0 : |V_i| \leq M\}$.

The pursuit estimate of $P(|V| > u)$ is then a weighted average of N_u computed along each path. The weight depends on the path.

3.1 Introduction

Since the seminal papers by Bollerslev [24] and Taylor [111] (cf. also Andersen et al [4]), the GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*) model has been widely used in finance and economics, and has inspired numerous variants such as GJR-GARCH of Glosten et al [63], *Asymmetric* GARCH of Engle and Ng [60] and the *Quadratic* GARCH of Sentana [100], among others. The basic GARCH model of Bollerslev [24] and Taylor [111] defines the conditional variance via the stochastic recurrence equation

$$\begin{aligned} R_t &= \sigma_t Z_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i R_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{3.1}$$

where $\{R_t\}_{t\in\mathbb{Z}}$ is the return series in question; $\{Z_t\}_{t\in\mathbb{Z}}$ is an iid sequence of random variables with zero mean and unit variance; the distribution function of Z_t is assumed to have a density. $\omega, \alpha_i, i \in \{1, \ldots, p\}$ and $\beta_j, j \in \{1, \ldots, q\}$ are constant parameters. A process defined by (3.1) is called a GARCH(p, q) process. When $p = q = 1$,

$$\sigma_t^2 = \omega + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2$$

When $p > 1$ or $q > 1$, the GARCH(p, q) process is given by a matrix recurrence equation (cf. Davis and Mikosch [41]). Define $d = p + q - 1$ and we can write the recurrence equation as

$$V_t = A_t V_{t-1} + B_t \tag{3.2}$$

where V_t and B_t are d-dimensional vectors; A_t are $d \times d$ matrices. The sequences A_t and B_t are both iid. Of course, B_t are really constant vectors, but we postpone this specialization for now and generalize the equation (3.2) to the broader context of matrix recursions. There is already a rich literature on this subject. Kesten [84] showed that, when A_t and B_t were almost surely non-negative, had no row or column of only zeros, and there was a positive probability that B_t was strictly positive, the strictly stationary solution to the equation $V \overset{d}{=} AV + B$ had power-law tails for its marginal distributions, assuming the following conditions (M) and (A):

- Condition (M)
  1. The top Lyapunov exponent
  $$\gamma = \inf_{n \geq 1} \frac{1}{n} \mathbb{E} \log \|A_n \cdots A_1\|$$
  is negative.
  2. There exists $\xi > 0$ such that
  $$1 = \lambda(\xi) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E} \|A_n \cdots A_1\|^{\xi}$$
  3. $\mathbb{E}(\|A_1\|^{\xi} \log^{+} \|A_1\|) < \infty$
  4. $\mathbb{E}|B_1|^{\xi} < \infty$
- Condition (A) : The group generated by
  $$\{\log \rho(s) : s = A_n \cdots A_1 \text{ for some } n \geq 1\}$$
  is dense in $\mathbb{R}$, where $\rho(s)$ denotes the spectral radius of matrix s.

Upon these conditions, Kesten's theorem gives

$$u^{\xi} \mathbb{P}(u^{-1} V \in \cdot) \overset{v}{\to} \mu_{\xi}(\cdot) \tag{3.3}$$

where μ_{ξ} is a non-null Radon measure on $\mathbb{R}_{+}^{d} \setminus \{0\}$ with the property $\mu_{\xi}(aA) = a^{-\xi} \mu_{\xi}(A)$.

Recently, Collamore and Mentemeier [37] extended Kesten's result and gave an explicit expression for μ_{ξ}:

$$\lim_{u \to \infty} u^{\xi} \mathbb{E}\left[f(u^{-1}V)\right] = \frac{C}{\lambda'(\xi)} \int_{\mathbb{S}_{+}^{d-1} \times \mathbb{R}} e^{-\xi s} f(e^{s} x) \ell_{\xi}(dx) ds \tag{3.4}$$

where C is a constant (cf. eq.(2.15) of Collamore and Mentemeier [37]), $f(\cdot)$ is any bounded continuous function on $\mathbb{R}_{+}^{d} \setminus \{0\}$ and ℓ_{ξ} is a probability measure on $\mathbb{S}_{+}^{d-1}$. Its definition is also found in (3.16).

From (3.4) a representation for $\mu_{\xi}(\cdot)$ immediately follows

$$\mu_{\xi}(\cdot) = \frac{C}{\lambda'(\xi)} \mathcal{L}_{\xi}(\cdot)$$

Here $\mathcal{L}_\xi$ is a non-null Radon measure on $\mathbb{R}_+^d \setminus \{0\}$ that satisfies, for all bounded continuous function $f(\cdot)$ on $\mathbb{R}_+^d \setminus \{0\}$:

$$\int_{\mathbb{R}_+^d \setminus \{0\}} f(x)\mathcal{L}_\xi(dx) = \int_{\mathbb{S}_+^{d-1} \times R} e^{-\xi s} f(e^s x)\ell_\xi(dx)ds$$

A key ingredient of Collamore and Mentemeier's approach is Hennion's uniform convergence result on the product of iid random matrices [70]:

$$\limsup_{n\to\infty} \left\{ \frac{1}{n}\mathbf{1}_{\{n>T\}} \left|\log\langle y, A_n \cdots A_1 x\rangle - \gamma\right| : x, y \geq 0, |x| = 1, |y| = 1 \right\} = 0$$

where

$$T = \min\{n \geq 1, A_n \cdots A_1 > 0\}. \tag{3.5}$$

Conditions (I, II, III) of hypothesis 1 and (ii) of remark 3.1 imply T as defined in (3.5) is almost surely finite. Cf. Buraczewski et al [29], example 4.4.13, Kesten [84], eq.(2.56) and Hennion [70], lemma 3.1., T is almost surely finite. Here $x \geq 0$ means every component of x is non-negative.

In addition to non-negative matrices, two other classes of random matrices have been shown to lead to power-law tails via the recurrence relation (3.2). Alsmeyer and Mentemeier [3] considered invertible matrices whose distribution has a density. Let $M(d, \mathbb{R})$ denote the space of $d \times d$ matrices with real entries that are invertible with probability 1. They replaced Kesten's condition $\mathbb{E}(\|A\|^\xi \log^+ \|A\|) < \infty$ with a stronger counterpart $\mathbb{E}[\|A\|^\xi(\log^+ \|A\| + \log \|A^{-1}\|)] < \infty$, and lifted the condition (A). In addition, they assumed

1. The Markov chain X_n on $\mathbb{S}^{d-1}$, namely $X_n = A_n X_{n-1}/|A_n X_{n-1}|$, is irreducible, i.e. for any open set $U \subset \mathbb{S}^{d-1}$ and any $u \in \mathbb{S}^{d-1}$, $\exists n \geq 1$ such that $\mathbb{P}(X_n u \in U) > 0$.
2. There exist $N \geq 1$, $c, \epsilon > 0$ and an invertible matrix $\bar{A} \in M(d, \mathbb{R})$ such that for any set $C \subset M(d, \mathbb{R})$, it holds true $\mathbb{P}(A_N \cdots A_1 \in C) \geq c|B_\epsilon(\bar{A}) \cap C|$, where $|\cdot|$ denotes the Lebesgue measure.

These assumptions are termed conditions (id). Furthermore, they assumed that there was no point in $\mathbb{R}^d$ such that the recurrence equation (3.2) was stuck at this point with probability 1: $\mathbb{P}(AX + B = X) < 1$ for all $X \in \mathbb{R}^d$ and all $A \in M(d, \mathbb{R})$. With these assumptions, they showed

$$\lim_{u\to\infty} u^\xi \mathbb{P}(\langle x, V\rangle > u) = e_\xi(x)$$

where $x \in \mathbb{S}^{d-1}$ and $e_\xi(\cdot)$ is a continuous function $\mathbb{S}^{d-1} \to \mathbb{R}_+$.

The second of the (id) conditions, which is satisfied when the distribution of A has a Lebesgue density, can actually be lifted if stronger moment conditions are imposed on A and B, and in addition, a proximity condition is satisfied by the support of A. This is the result of Guivarc'h and Le Page, et al [65]. Let G_A denote the semi-group generated by $\{\Pi_n : \Pi_n = A_n \cdots A_1, A_i \in M(d, \mathbb{R})\}$. The authors assumed

1. There is no finite union W of proper subspaces of $\mathbb{R}^d$ that satisfies $\forall a \in G_A, aW = W$.
2. G_A contains a proximal element, i.e. an element a whose largest singular value is an algebraically simple eigenvalue of a.

These two assumptions are termed (ip) conditions. Replacing the (id) conditions of Alsmeyer and Mentemeier with (ip) and the moment conditions of the former with

$$\mathbb{E}\|A\|^{\xi+\delta} < \infty, \quad \mathbb{E}(\|A\|^{\xi}\|A^{-1}\|^{\delta}) < \infty, \quad \mathbb{E}(|B|^{\xi+\delta} < \infty), \qquad \text{for some } \delta > 0$$

Guivarc'h and Le Page et al.proved the same vague convergence result of (3.3).

In the special case of GARCH(p, q), Bollerslev [24] showed that the equation $X \stackrel{d}{=} AX + B$ has a unique, strictly stationary solution with finite variance if and only if

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \tag{3.6}$$

In the rest of this paper, we always assume condition (3.6) is satisfied. For convenience of narration, let π denote this unique stationary probability measure and let $V \sim \pi$, $Z \sim \mu_Z$. More generally we write μ_U for the probability measure of U, no matter what type of object U may be.

Buraczewski et al.[29] (proposition 4.3.1) derived the support of π assuming the condition of Bollerslev 3.6. We omit the formula here and refer to it as χ hereafter.

In addition to (3.6), we assume:

Hypothesis 1. *All the following conditions hold:*

(I) $\exists s > 0$ *such that* $1 < \mathbb{E}(\alpha_1 Z^2 + \beta_1)^s < \infty$

(II) If $p, q \geq 2$, *there exists an non-empty open set* $S \subset supp\, \mu_Z$.

(III) $\alpha_p > 0$ *and* $\beta_q > 0$.

Clearly, these conditions are satisfied when Z has normal or t distributions.

Remark 3.1. *From hypothesis 1, a few implications immediately follow*

(i) (3.1) *implies*

$$\sigma_t^2 \geq \omega \left(1 - \sum_{j=1}^{q} \beta_j\right)^{-1} = \sigma_{min}^2 > 0$$

Then it follows $\chi \subseteq [\sigma_{min}^2, \infty)^q \times [0, \infty)^{p-1}$, *so* $V_n \in \chi$ *for all* $n \geq 0$. *Since the random variable* Z_{n-1}^2 *is assumed to have a continuously differentiable distribution function,* $\mathbb{P}(Z_{n-1}^2 = x) = 0$ *for all* $x \in supp\, \mu_{Z^2}$. *Furthermore,* Z_{n-1}^2 *uniquely determines the matrix* A_n, *so it follows* $\mathbb{P}(Av + B = v) = 0$ *for all* $v \in \chi$.

(ii) (3.6) *implies the top Lyapunov exponent of* A_n

$$\gamma = \inf_{n \geq 1} \frac{1}{n} \mathbb{E}\left(\log \|A_n \cdots A_1\|\right) \tag{3.7}$$

is negative. Cf. Buraczewski et al [29], prop. 4.1.12.

(iii) That $\mathbf{0} \notin \chi$ *and* A_n *has a Lebesgue density implies the stationary distribution* π *is absolutely continuous with respect to Lebesgue measure. This immediately follows from lemma 4.2.2 of Buraczewski et al.[29].*

(iv) Condition (III) ensures that, with probability 1, every row and column of the matrix A_n has at least one postive component.

The implications (i), (ii), (iii) are in fact the conditions of proposition 4.2.1 of Buraczewski et al [29], from which we conclude V_n is an aperiodic, positive Harris chain that is in addition π-irreducible on χ. Moreover, from (I) it follows $0 < \exists\xi < s$ such that

$$\lambda(\xi) = \lim_{n\to\infty} (\mathbb{E}\|A_n \cdots A_1\|^{\xi})^{1/n} = 1 \tag{3.8}$$

and

$$\mathbb{E}\|A\|^{\xi} < \infty \tag{3.9}$$

The existence of ξ together with Conditions (ii, iv) and (I, II) allow the application of Kesten's theorem (cf. Buraczewski et al [29], example 4.4.13).

Although the probability $\mathbb{P}(\langle \tilde{x}, y\rangle > u)$ has been given asymptotically by Kesten's theorem, one often wishes to know this probability more precisely, due to the importance of risk management. Now that more detailed analytic description of the probability is unknown, one has to resort to numerical methods. But the occurrence of $\langle \tilde{x}, V\rangle > u$ for a large u is a rare event; a naive Monte-Carlo approach will be very inefficient. Cf. Asmussen and Glynn [6]. One way to increase the efficiency of Monte-Carlo methods is importance sampling.

The idea of importance sampling with exponential shift dates back to Siegmund [103], who devised an algorithm for estimating the excursion probability of 1D random walk with iid increments. Following his work, various importance sampling algorithms have been proposed for rare event simulation in a variety of problems.

Let $W_n = \sum_{i=1}^{n} X_n$ be a random walk. Blanchet and Glynn [22] proposed a state-dependent importance sampling algorithm to estimate the tail of $\max\{W_n, n \geq 1\}$ and showed that their estimator had bounded relative error (cf. Asmussen and Glynn [6]). In the case of light tailed increments, their estimator recovers that of Siegmund.

In 2010, Blanchet and Liu [23] presented an importance sampling algorithm for the first passage time of a multidimensional random walk with heavy-tailed increments.

However, to our best knowledge, no importance sampling estimator has been proposed in the literature for the computation of $\mathbb{P}(\langle \tilde{x}, V\rangle > u)$ or for the more general problem when the defining recurrence equation of V_n i.e. (3.2) is more general than that of GARCH(p, q). We present a solution in this paper.

When $p = q = 1$, V_n reduces to a scalar. An importance sampling estimator was proposed and shown to be efficient in the sense of bounded relative error by Collamore et al.[36]. We consider our work as a multivariate extension to theirs.

3.2 Statement of results

Our solution involves associating a *Markov Additive* process (X_n, S_n) to the Markov chain V_n:

$$X_t = \frac{A_t A_{t-1} \cdots A_1 \tilde{V}_0}{|A_t A_{t-1} \cdots A_1 \tilde{V}_0|}, \quad X_0 = \tilde{V}_0 \tag{3.10}$$

$$S_t = \log|A_t \cdots A_1 \tilde{V}_0| \tag{3.11}$$

$$l_t = S_t - S_{t-1} = \log|A_t X_{t-1}| \tag{3.12}$$

where $\tilde{v} = v/|v|$ for a vector v. From the GARCH(p, q) recurrence relation

$$\begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \vdots \\ \sigma_{t-q+2}^2 \\ \sigma_{t-q+1}^2 \\ R_{t-1}^2 \\ R_{t-2}^2 \\ \vdots \\ R_{t-p+2}^2 \\ R_{t-p+1}^2 \end{pmatrix} = \begin{pmatrix} \alpha_1 Z_{t-1}^2 + \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ Z_{t-1}^2 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{t-1}^2 \\ \sigma_{t-2}^2 \\ \vdots \\ \sigma_{t-q+1}^2 \\ \sigma_{t-q}^2 \\ R_{t-2}^2 \\ R_{t-3}^2 \\ \vdots \\ R_{t-p+1}^2 \\ R_{t-p}^2 \end{pmatrix} + \begin{pmatrix} \omega \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{3.13}$$

it is obvious that, by define mapping

$$g : (x, y, l) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \times \mathbb{R} \to \mathbb{R}_+ \ni \frac{\langle \mathbf{e}_{q+1}, e^l y \rangle}{\langle \mathbf{e}_1, x \rangle}$$

one has the relation $Z_{n-1}^2 = g(X_{n-1}, X_n, l_n)$. Let

$$\mathcal{F}_n = \mathcal{B}(X_0, X_1, \ldots, X_n, l_1, l_2, \ldots, l_n)$$

where $\mathcal{B}(\cdot)$ denotes the σ-field generated by $\cdot$. It is clear $\mathcal{B}(V_n) \subseteq \mathcal{F}_n$. Let P denote the transition kernel of (X_n, S_n). We have

$$P(x, dy \times dl) = \mathbb{P}(X_n \in dy, l_n \in dl | X_{n-1} = x) = \mathbb{P}(Z_{n-1}^2 \in g(x, dy, dl))$$

Note $g(\mathbb{S}^{d-1}, \mathbb{S}^{d-1}, \mathbb{R}) = \operatorname{Img} Z^2$, where $\operatorname{Img} Z^2$ denotes the image of Z^2. Choose a set $\mathcal{S} \subset \mathbb{S}^{d-1}$ such that

$$\inf_{w \in g(\mathcal{S}, \mathbb{S}^{d-1}, \mathbb{R})} f_{Z^2}(w) > 0$$

We have

$$P(x, dy \times dl) \geq \mathbf{1}_{\mathcal{S}}(x) \inf_{w \in g(\mathcal{S}, dy, dl)} f_{Z^2}(w) |g(\mathcal{S}, dy, dl)|$$

where $f_{Z^2}(\cdot)$ is the density function of Z^2 with respect to the Lebesgue measure and $|\cdot|$ denotes the Lebesgue measure. It is easy to see

$$\int_{\mathbb{S}^{d-1} \times \mathbb{R}} \inf_{w \in g(\mathcal{S}, dy, dl)} f_{Z^2}(w) |g(\mathcal{S}, dy, dl)| < \infty$$

Clearly

$$\begin{aligned} & \int_{\mathbb{S}^{d-1} \times \mathbb{R}} \inf_{w \in g(\mathcal{S}, dy, dl)} f_{Z^2}(w) |g(\mathcal{S}, dy, dl)| \leq \int_{\mathbb{S}^{d-1} \times \mathbb{R}} \int_{g(\mathcal{S}, dy, dl)} f_{Z^2}(w) dw \\ = & \int_{g(\mathcal{S}, \mathbb{S}^{d-1}, \mathbb{R})} f_{Z^2}(w) dw < \int_{\operatorname{Img} Z^2} f_{Z^2}(w) dw = 1 \end{aligned}$$

Let

$$\delta = \int_{\mathbb{S}^{d-1} \times \mathbb{R}} \inf_{w \in g(\mathcal{S}, dy, dl)} f_{Z^2}(w) |g(\mathcal{S}, dy, dl)| < 1$$

and

$$\nu(dy \times dl) = \frac{1}{\delta} \inf_{w \in g(\mathcal{S}, dy, dl)} f_{Z^2}(w) |g(\mathcal{S}, dy, dl)|$$

Now that $\nu(\mathbb{S}^{d-1} \times \mathbb{R}) = 1$, $\nu(\cdot)$ is a probability measure. We have the minorization condition

$$P(x, dy \times dl) \geq \delta \mathbf{1}_{\mathcal{S}}(x) \nu(dy \times dl) \tag{3.14}$$

By Ney and Nummelin [93], lemma 3.1, (3.14) implies the MA-process (X_n, S_n) has a regenerative structure:

(1) There exist random variables $0 < \tau_0 < \tau_1 < \dots$, $i = 0, 1, 2, \dots$ such that $\tau_{i+1} - \tau_i$, are iid.

(2) The blocks

$$(X_{\tau_i}, X_{\tau_i+1}, \dots, X_{\tau_{i+1}-1}, l_{\tau_i}, l_{\tau_i+1}, \dots, l_{\tau_{i+1}-1}) \quad i = 0, 1, 2, \dots$$

are independent of each other.

(3)
$$\mathbb{P}[(X_{\tau_i}, l_{\tau_i}) \in S \times \Gamma | \mathcal{F}_{\tau_i - 1}] = \nu(S \times \Gamma)$$

Furthermore, (3.14) means $P(x, dy \times dl)$ can be decomposed as

$$P(x, dy \times dl) = P'(dx, dy \times dl) + \delta \mathbf{1}_{\mathcal{S}}(x) \nu(dy \times dl)$$

Thus the MA-process regenerates only when it is in $\mathcal{S}$ and in this case it regenerates with probability δ. That is

$$\mathbb{P}[(X_n, S_n) \text{ regenerates } | X_{n-1} = x] = \delta \mathbf{1}_{\mathcal{S}}(x)$$

There is yet another useful property of the iid matrices A_n. Define mapping

$$A \cdot x : (A, x) \in \operatorname{Img}(A) \times \mathbb{S}^{d-1} \to \mathbb{S}^{d-1} \ni \frac{Ax}{|Ax|}$$

and operator $\mathscr{P}^\theta$ for $\theta \in \mathbb{R}$, $f : \mathbb{S}^{d-1} \to \mathbb{R}_+$:

$$\mathscr{P}^\theta f(x) = \mathbb{E}\left[|Ax|^\theta f(A \cdot x)\right] \tag{3.15}$$

By Lemma 2.2 of Collamore and Mentemeier [37], (3.9) means $\lambda(\xi) = 1$ is the spectral radius of $\mathscr{P}^\xi$ and there is a unique, strictly positive eigenfunction $r_\xi(\cdot)$ associated with $\lambda(\xi)$ i.e. $\mathscr{P}^\xi r_\xi(x) = \lambda(\xi) r_\xi(x)$. Moreover, r_ξ is $\max\{\xi, 1\}$-Hölder continuous, implying r_ξ is bounded from above and below by positive constants. From now on, we use the notations

$$\bar{r}_\xi = \sup_{x \in \mathbb{S}^{d-1}} r_\xi(x), \quad \underline{r}_\xi = \inf_{x \in \mathbb{S}^{d-1}} r_\xi(x)$$

There is also an eigenmeasure ℓ_ξ on $\mathcal{B}(\mathbb{S}^{d-1})$ associated with the operator $\mathscr{P}^\xi$ that corresponds to the eigenvalue $\lambda(\xi) = 1$ and eigenfunction r_ξ:

$$\mathbb{E}\left[|Ax|^\xi \ell(A \cdot dx)\right] = \ell_\xi \mathscr{P}^\xi(dx) = \lambda(\xi) \ell_\xi(dx) \tag{3.16}$$

The eigenfunction r_ξ and eigenmeasure ℓ_ξ are called right eigenfunction and left eigenmeasure, respectively. They satisfy the identity $\ell_\xi r_\xi = \int_{\mathbb{S}^{d-1}} r_\xi(x) \ell_\xi(dx) = 1$. Cf. Collamore and Mentemeier [37], Lemma 2.2.

Naively one would estimate $\mathbb{P}(|V| > u)$ as $n^{-1}\sum_{i=1}^{n} \mathbf{1}_{\{|V_i|>u\}}$, applying the law of large numbers. The difficulty with this naive method is that, when u is large, $|V_i| > u$ happens very rarely, resulting in a large variance of the estimator. To tackle this problem, we use importance sampling and exponentially shift the transition kernel of the MA process (X_i, S_i), i.e. the conditional probability $P(x, dy \times dl)$, until $|V_t| > u$. Let

$$P^\theta(x, dy \times dl) = \frac{e^{\theta l}}{\lambda(\theta)} \frac{r_\theta(y)}{r_\theta(x)} P(x, dy \times dl)$$

Since the matrix A_t depends only on Z_{t-1}^2, shifting the transition kernel of (X_t, S_t) is equivalent to shifting the conditional distribution of Z_{t-1}^2. It follows from the above equation

$$\frac{\mathbb{P}^\theta(Z_{t-1}^2 \in dw | X_{t-1} = x)}{\mathbb{P}(Z_{t-1}^2 \in dw | X_{t-1} = x)} = \frac{|A(w)x|^\theta}{\lambda(\theta)} \frac{r_\theta(A(w)\cdot x)}{r_\theta(x)} \tag{3.17}$$

where $\mathbb{P}^\theta(\cdot|\cdot)$ denotes the shifted conditional probability measure.

Now we are ready to introduce our importance sampling estimator. Define M and $\{K_i\}_{i=0,1,\ldots}$ as in lemma 3.5. We start the process V_t from within $\mathcal{C} = \{v \in \chi : |v| < M\}$ and let $V_0 \sim \eta$, where the probability measure η is defined as

$$\eta(S) = \pi(S)/\pi(\mathcal{C}) \quad \forall S \in \mathcal{B}(\mathcal{C})$$

Let

$$\begin{aligned} R_n &:= \sup\{i \geq 0 : K_i \leq n\} \\ T_u &= \inf\{n \geq 1 : |V_n| > u\} \\ N_u &:= \sum_{i=0}^{K_1-1} \mathbf{1}_{\{V_i > u\}} \\ \mathcal{E}_u &= \pi(\mathcal{C}) N_u \mathbf{1}_{\{T_u < K_1\}} e^{-\xi S_{T_u}} \frac{r_\xi(X_0)}{r_\xi(X_{T_u})} \end{aligned}$$

$\mathcal{E}_u$ is our estimator. We have

Theorem 3.2. *The estimator $\mathcal{E}_u$ is unbiased, i.e.*

$$\mathbb{P}(|V| > u) = \mathbb{E}_\eta^{\mathcal{D}} \mathcal{E}_u \tag{3.18}$$

The superscript $\mathcal{D}$, short for "dual", is to remind us that the expectation is taken with respect to the shifted kernel P_ξ before the threshold is exceeded, and with respect to the original kernel P thereafter. The subscript η means that V_0 is drawn from the distribution η.

While unbiased, the estimator $\mathcal{E}_u$ is also efficient, i.e. its relative error is bounded. Cf. Asmussen and Glynn [6]. This constitutes the next theorem:

Theorem 3.3. *Let M and $K_i, i = 0, 1, 2, \ldots$ be defined as in Lemma 3.5 and $0 < b < 1$ be the constant shown to exist by lemma 3.5. Assume $b^{1-\xi/s} < e^\gamma$, where s is the positive constant satisfying condition (I) of hypothesis 1 and γ is the top Lyapunov exponent defined by (3.7). Then the estimator $\mathcal{E}_u$ has bounded relative error, i.e.*

$$\limsup_{u \to \infty} \frac{var(\mathcal{E}_u)}{[\mathbb{P}(|V| > u)]^2} < \infty$$

In §3.3 we show that, with certain shifted kernels, the MA-process drifts towards a set of bounded $|V_t|$. This is a crucial fact for the consistency and efficiency of $\mathcal{E}_u$. Then in §3.4 we prove theorem 3.2 and in §3.5 we prove theorem 3.3.

3.3 V_n drifts towards a small set

3.3.1 A drift condition

Lemma 3.4. *Let $0 < \theta < \xi$. Then there exist $0 < b_\theta < 1, M_\theta > 0$ such that*

$$\mathbb{E}\left[|V_n|^\theta r_\theta(\tilde{V}_n)\mathbf{1}_{\{|V_{n-1}|>M_\theta\}}\middle|\mathcal{F}_{n-1}\right] \le b_\theta |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1}) \tag{3.19}$$

Proof. $\lambda(\cdot)$ is a convex continuous function (cf. Buraczewski et al [29], §4.4.3), so $\lambda(0) = 1 = \lambda(\xi)$ implies $\lambda(\theta) < 1$. By Buraczewski et al.[28] Proposition 3.1, an eigenfunction $r_\theta(\cdot)$ and an eigenmeasure $\ell_\theta(\cdot)$ exist for the operator $\mathscr{P}^\theta$. In particular, the right eigenfunction can be represented as

$$r_\theta(x) = c(\theta)\int_{\mathbb{S}^{d-1}} \langle x, y\rangle^\theta \ell^*_\theta(dy)$$

Thus we have

$$\begin{aligned} & \mathbb{E}\left[|V_n|^\theta r_\theta(\tilde{V}_n)\mathbf{1}_{\{|V_{n-1}|>M_\theta\}}|\mathcal{F}_{n-1}\right] \\ = \;& \mathbf{1}_{\{|V_{n-1}|>M_\theta\}}\mathbb{E}\left[c(\theta)\int_{\mathbb{S}^{d-1}}\langle V_n, y\rangle^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] \\ = \;& \mathbf{1}_{\{|V_{n-1}|>M_\theta\}}\mathbb{E}\left[c(\theta)\int_{\mathbb{S}^{d-1}}(\langle A_nV_{n-1}, y\rangle + \langle B, y\rangle)^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] \end{aligned} \tag{3.20}$$

case 1. *If $\theta \le 1$, by subadditivity we have*

$$\begin{aligned} & \mathbb{E}\left[c(\theta)\int_{\mathbb{S}^{d-1}}(\langle A_nV_{n-1}, y\rangle + \langle B, y\rangle)^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] \\ \le \;& \mathbb{E}\left[c(\theta)\int_{\mathbb{S}^{d-1}}\langle A_nV_{n-1}, y\rangle^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] + \mathbb{E}\left[c(\theta)\int_{\mathbb{S}^{d-1}}\langle B, y\rangle^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] \\ = \;& \mathbb{E}\left[|V_{n-1}|^\theta |A_n\tilde{V}_{n-1}|^\theta c(\theta)\int_{\mathbb{S}^{d-1}}\langle A_n\cdot \tilde{V}_{n-1}, y\rangle^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] + |B|^\theta r_\theta(\tilde{B}) \\ = \;& |V_{n-1}|^\theta \mathbb{E}\left[|A_n\tilde{V}_{n-1}|^\theta r_\theta(A_n\cdot\tilde{V}_{n-1})\middle|\mathcal{F}_{n-1}\right] + |B|^\theta r_\theta(\tilde{B}) \\ = \;& |V_{n-1}|^\theta \lambda(\theta) r_\theta(\tilde{V}_{n-1}) + |B|^\theta r_\theta(\tilde{B}) \\ = \;& |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})\lambda(\theta)\left[1 + \frac{|B|^\theta r_\theta(\tilde{B})}{\lambda(\theta)|V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})}\right] \end{aligned}$$

Then we have

$$\begin{aligned} & \mathbb{E}\left[|V_n|^\theta r_\theta(\tilde{V}_n)\mathbf{1}_{\{|V_{n-1}|>M_\theta\}}|\mathcal{F}_{n-1}\right] \\ \le \;& |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})\lambda(\theta)\left[1 + \frac{|B|^\theta r_\theta(\tilde{B})}{\lambda(\theta)|V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})}\right]\mathbf{1}_{\{|V_{n-1}|>M_\theta\}} \\ \le \;& |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})\lambda(\theta)\left[1 + \frac{|B|^\theta r_\theta(\tilde{B})}{\lambda(\theta)M_\theta \underline{r}_\theta}\right] \end{aligned}$$

Since $\lambda(\theta) < 1$, there exists

$$M_\theta = \frac{|B|^\theta r_\theta(\tilde{B})}{[1-\lambda(\theta)]\underline{r}_\theta} + \epsilon \tag{3.21}$$

for an $\epsilon > 0$ such that

$$b_\theta = \lambda(\theta)\left[1 + \frac{|B|^\theta r_\theta(\tilde{B})}{\lambda(\theta) M_\theta \underline{r}_\theta}\right] < 1 \tag{3.22}$$

Thus (3.19) holds.

case 2. *If $\theta > 1$, applying Minkowski's inequality to the RHS of (3.20) gives*

$$\begin{aligned}
& \mathbb{E}\left[c(\theta)\int_{\mathbb{S}^{d-1}} (\langle A_n V_{n-1}, y\rangle + \langle B, y\rangle)^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right] \\
\leq & \left\{\left[\mathbb{E}\left(c(\theta)\int_{\mathbb{S}^{d-1}} \langle A_n V_{n-1}, y\rangle^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right)\right]^{1/\theta}\right. \\
& \left. + \left[\mathbb{E}\left(c(\theta)\int_{\mathbb{S}^{d-1}} \langle B, y\rangle^\theta \ell^*_\theta(dy)|\mathcal{F}_{n-1}\right)\right]^{1/\theta}\right\}^\theta \\
\leq & \left\{|V_{n-1}|\lambda(\theta)^{1/\theta} r_\theta(\tilde{V}_{n-1})^{1/\theta} + |B| r_\theta(\tilde{B})^{1/\theta}\right\}^\theta \\
\leq & |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})\left[\lambda(\theta)^{1/\theta} + \frac{|B| r_\theta(\tilde{B})^{1/\theta}}{|V_{n-1}| r_\theta(\tilde{V}_{n-1})}\right]^\theta
\end{aligned}$$

Thus, as in the previous case, we have

$$\begin{aligned}
& \mathbb{E}\left[|V_n|^\theta r_\theta(\tilde{V}_n)\mathbf{1}_{\{|V_{n-1}|>M_\theta\}}|\mathcal{F}_{n-1}\right] \\
\leq & |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})\left[\lambda(\theta)^{1/\theta} + \frac{|B| r_\theta(\tilde{B})^{1/\theta}}{|V_{n-1}| r_\theta(\tilde{V}_{n-1})}\right]^\theta \mathbf{1}_{\{|V_{n-1}|>M_\theta\}} \\
\leq & |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})\left[\lambda(\theta)^{1/\theta} + \frac{|B| r_\theta(\tilde{B})^{1/\theta}}{M_\theta \underline{r}_\theta}\right]^\theta
\end{aligned}$$

Choose

$$\begin{aligned}
M_\theta &= \frac{|B| r_\theta(\tilde{B})^{1/\theta}}{(1-\lambda(\theta)^{1/\theta})\underline{r}_\theta} + \epsilon \text{ for some } \epsilon > 0 \\
b_\theta &= \left[\lambda(\theta)^{1/\theta} + \frac{|B| r_\theta(\tilde{B})^{1/\theta}}{M_\theta \underline{r}_\theta}\right]^\theta < 1
\end{aligned} \tag{3.23}$$

Then we have

$$\mathbb{E}\left[|V_n|^\theta r_\theta(\tilde{V}_n)\mathbf{1}_{\{|V_{n-1}|>M_\theta\}}|\mathcal{F}_{n-1}\right] \leq b_\theta |V_{n-1}|^\theta r_\theta(\tilde{V}_{n-1})$$

We have proved the lemma. □

The conclusion of lemma 3.4 allows us to bound the return time of V_n to the set $\mathcal{C} = \{v \in \chi, |v| < M\}$, where max is a postive constant. This is the next lemma.

3.3.2 A bound on the return time to $\mathcal{C}$

Lemma 3.5. *Let* Θ *be a proper subset of* $(0, \xi)$*, i.e.* $\inf \Theta > 0$ *and* $\sup \Theta < \xi$*. Define*

$$M = \sup_{\theta \in \Theta} M_\theta \tag{3.24}$$

$$K_0 = 0, \quad K_i = \inf\{n > K_{i-1} : |V_n| \leq M\}, \quad i \geq 1 \tag{3.25}$$

Then there exist $0 < b < 1$ *and* $\rho > 0$ *such that for* $n \geq 1$*,*

$$\mathbb{P}(K_{j+1} - K_j > n) \leq b^n \rho \tag{3.26}$$

Proof. Iterating (3.19) yields, for $j \geq 0, n > 1$ and $\theta \in \Theta$,

$$\mathbb{E}\left[|V_{K_j+n}|^\theta r_\theta(\tilde{V}_{K_j+n}) \prod_{i=1}^{n-1} \mathbf{1}_{\{|V_{K_j+i}| > M_\theta\}} \middle| \mathcal{F}_{K_j+1} \right] \leq b_\theta^{n-1} |V_{K_j+1}|^\theta r_\theta(\tilde{V}_{K_j+1})$$

Because $\{K_{j+1} - K_j > n-1\} \subseteq \bigcap_{i=K_j+1}^{K_j+n-1} \{|V_i| > M_\theta\}$,

$$\begin{aligned} \mathbb{E}\left[|V_{K_j+n}|^\theta \underline{r}_\theta \mathbf{1}_{\{K_{j+1}-K_j>n-1\}} \middle| \mathcal{F}_{K_j+1} \right] &\leq b_\theta^{n-1} |V_{K_j+1}|^\theta \bar{r}_\theta, \\ \mathbb{E}\left[|V_{K_j+n}|^\theta \underline{r}_\theta \mathbf{1}_{\{K_{j+1}-K_j>n-1\}} \right] &\leq b_\theta^{n-1} \mathbb{E}|V_{K_j+1}|^\theta \bar{r}_\theta \end{aligned} \tag{3.27}$$

If $\theta < 1$, by subadditivity we have

$$\mathbb{E}|V_{K_j+1}|^\theta \leq \mathbb{E}\left(|A_{K_j+1} V_{K_j}|^\theta + |B|^\theta \right) = \mathbb{E}\left(|A_{K+1} \tilde{V}_K|^\theta |V_K| \right) \leq M^\theta \mathbb{E}\|A_{K+1}\|^\theta$$

Since $0 < \theta < \xi$ and $\mathbb{E}\|A_{K+1}\|^0 = 1$, $\mathbb{E}\|A_{K+1}\|^\xi < \infty$, it follows by continuity $\mathbb{E}\|A_{k+1}\|^\theta < \infty$. So

$$\mathbb{E}|V_{K_j+1}|^\theta \leq M^\theta \mathbb{E}\|A\|^\theta + |B|^\theta$$

If $\theta \geq 1$, by Minkowski inequality we have

$$\begin{aligned} (\mathbb{E}|V_{K+1}|^\theta)^{1/\theta} &\leq (\mathbb{E}|A_{K+1} V_K|^\theta)^{1/\theta} + |B| \leq M (\mathbb{E}|A_{K+1} \tilde{V}_K|^\theta)^{1/\theta} + |B| \\ &\leq M (\mathbb{E}\|A\|^\theta)^{1/\theta} + |B|, \\ \mathbb{E}|V_{K+1}|^\theta &\leq \left[M (\mathbb{E}\|A\|^\theta)^{1/\theta} + |B| \right]^\theta < \infty \end{aligned}$$

Then it follows from (3.27) for $n \geq 1$,

$$\underline{V}^\theta \underline{r}_\theta \mathbb{P}(K_{j+1} - K_j > n-1) \leq \mathbb{E}\left[|V_{K_j+n}|^\theta \underline{r}_\theta \mathbf{1}_{\{K_{j+1}-K_j>n-1\}} \right] \leq b_\theta^{n-1} \mathbb{E}|V_{K_j+1}|^\theta \bar{r}_\theta$$

That is

$$\mathbb{P}(K_{j+1} - K_j > n) < b_\theta^n \rho_\theta \tag{3.28}$$

where

$$\rho_\theta = \frac{\bar{r}_\theta}{\underline{r}_\theta \underline{V}^\theta} \times \begin{cases} M^\theta \mathbb{E}\|A\|^\theta + |B|^\theta & \theta < 1 \\ \left[M (\mathbb{E}\|A\|^\theta)^{1/\theta} + |B| \right]^\theta & \theta \geq 1 \end{cases}$$

Since the inequality (3.28) holds for all $\theta \in \Theta$, we have for $n \geq 1$,

$$\mathbb{P}(K_{j+1} - K_j > n) \leq \inf_{\theta \in \Theta} b_\theta^n \rho_\theta \leq \left(\inf_{\theta \in \Theta} b_\theta \right)^n \sup_{\theta \in \Theta} \rho_\theta$$

Thus (3.26) holds with

$$b = \inf_{\theta \in \Theta} b_\theta, \quad \rho = \sup_{\theta \in \Theta} \rho_\theta \tag{3.29}$$

□

3.4 The estimator is unbiased

In this section we prove theorem 3.2.

Proof. By the *strong law of large numbers* for Markov chains,

$$\frac{1}{n}\sum_{i=0}^{n}\mathbf{1}_{\{|V_i|>u\}} \overset{a.s.}{\to} \mathbb{P}(|V|>u)$$

Define $R_n = \sup\{i \geq 0 : K_i \leq n\}$. Then one can write

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{|V_i|>u\}} = \frac{1}{n}\left[\sum_{i=0}^{K_{R_n}-1}\mathbf{1}_{\{|V_i|>u\}} + \sum_{i=K_{R_n}}^{n}\mathbf{1}_{\{|V_i|>u\}}\right] \tag{3.30}$$

For the 2nd term on the right side, we show in the following

$$\frac{1}{n}\sum_{i=K_{R_n}}^{n}\mathbf{1}_{\{|V_i|>u\}} \overset{a.s.}{\to} 0 \tag{3.31}$$

This is, by definition, for all $\epsilon > 0$

$$\mathbb{P}\left[\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left\{\frac{1}{n}\sum_{i=K_{R_n}}^{n}\mathbf{1}_{\{|V_i|>u\}} \leq \epsilon\right\}\right] = 1$$

By Borel-Cantelli lemma, it suffices to show

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\frac{1}{n}\sum_{i=K_{R_n}}^{n}\mathbf{1}_{\{|V_i|>u\}} > \epsilon\right] < \infty$$

Clearly

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}\left[\frac{1}{n}\sum_{i=K_{R_n}}^{n}\mathbf{1}_{\{|V_i|>u\}} > \epsilon\right] &\leq \sum_{n=1}^{\infty}\mathbb{P}\left[\frac{1}{n}\sum_{i=K_{R_n}}^{K_{R_n+1}-1}\mathbf{1}_{\{|V_i|>u\}} > \epsilon\right] \\ &\leq \sum_{n=1}^{\infty}\mathbb{P}\left(K_{R_n+1} - K_{R_n} > \lfloor \epsilon n \rfloor\right)\end{aligned}$$

It suffices to show

$$\sum_{n=\lceil 1/\epsilon \rceil}^{\infty}\mathbb{P}\left(K_{R_n+1} - K_{R_n} > \lfloor \epsilon n \rfloor\right) < \infty$$

By Lemma 3.5, $\mathbb{P}(K_{j+1} - K_j > k) < b^n \rho$ for $k \geq 1$. Thus

$$\sum_{n=\lceil 1/\epsilon \rceil}^{\infty}\mathbb{P}\left(K_{R_n+1} - K_{R_n} > \lfloor \epsilon n \rfloor\right) < \sum_{n=\lceil 1/\epsilon \rceil}^{\infty} b^{\lfloor \epsilon n \rfloor}\rho < \infty$$

This shows (3.31) holds. For the 1st term on the right side of (3.30), we have

$$\frac{1}{n}\sum_{i=0}^{K_{R_n}-1}\mathbf{1}_{\{|V_i|>u\}} = \frac{R_n}{n}\frac{1}{R_n}\sum_{i=1}^{R_n}\sum_{j=K_{i-1}}^{K_i-1}\mathbf{1}_{\{|V_i|>u\}}$$

It can be shown $(V_{K_i}, \sum_{j=K_{i-1}}^{K_i-1}\mathbf{1}_{\{|V_i|>u\}})$ is a positive Harris chain. Moreover

$$\mathbb{E}\left(\sum_{j=K_{i-1}}^{K_i-1}\mathbf{1}_{\{|V_j|>u\}}\right) \le \mathbb{E}(K_i - K_{i-1}) < \sum_{n=1}^{\infty} n b_0^n \rho_0 < \infty$$

Therefore, by the law of large numbers for Markov chains,

$$\frac{R_n}{n}\frac{1}{R_n}\sum_{i=1}^{R_n}\sum_{j=K_{i-1}}^{K_i-1}\mathbf{1}_{\{|V_i|>u\}} \quad \stackrel{a.s.}{\to} \quad \pi(\mathcal{C})\mathbb{E}_\eta N_u$$

Hence $\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{|V_i|>u\}} \stackrel{a.s.}{\to} \pi(\mathcal{C})\mathbb{E}_\eta N_u$. On the other hand, by the very definition of N_u, $\mathbb{E}_\eta N_u = \mathbb{E}_\eta\left(N_u \mathbf{1}_{\{T_u<K_1\}}\right)$. We have

$$\begin{aligned}
&\mathbb{E}_\eta\left(N_u\mathbf{1}_{\{T_u<K_1\}}\right)\\
=&\int_{(\mathbb{S}^{d-1}\times\mathbb{R})^{K_1-1}} N_u\mathbf{1}_{\{T_u<\tau\}}\prod_{i=1}^{T_u} e^{-\xi l_i}\frac{r_\xi(x_{i-1})}{r_\xi(x_i)}P_\xi(x_{i-1}, dx_i\times dl_i)\prod_{i=T_u+1}^{K_1-1}P(x_{i-1}, dx_i\times dl_i)\\
=&\mathbb{E}_\eta^{\mathcal{D}}\left[N_u\mathbf{1}_{\{T_u<K_1\}}e^{-\xi S_{T_u}}\frac{r_\xi(X_0)}{r_\xi(X_{T_u})}\right]
\end{aligned}$$

Thus we have proved the theorem. □

3.5 The estimator has bounded relative error

In this section we prove that the estimator $\mathcal{E}_u$ is efficient, i.e. theorem 3.3.

Proof. The assertion is implied by, for all $X_0 \in \mathcal{C}$,

$$\limsup_{u\to\infty}\frac{\mathbb{E}_{X_0}^{\mathcal{D}}\mathcal{E}_u^2}{[\mathbb{P}(|V|>u)]^2} < \infty$$

For notational simplicity, we omit the subscript X_0 and write $\mathbb{E}^{\mathcal{D}}$ for $\mathbb{E}_{X_0}^{\mathcal{D}}$ in the rest of the proof. By Kesten's theorem [84], $\mathbb{P}(|V|>u) \sim Cu^{-\xi}$. Hence, to prove the assertion, one needs to check

$$\limsup_{u\to\infty} u^{2\xi}\mathbb{E}^{\mathcal{D}}\mathcal{E}_u^2 < \infty$$

That is,

$$\limsup_{u\to\infty}\mathbb{E}^{\mathcal{D}}\left[u^{2\xi}N_u^2\mathbf{1}_{\{T_u<K_1\}}e^{-2\xi S_{T_u}}\frac{r_\xi^2(X_0)}{r_\xi^2(X_{T_u})}\right] < \infty$$

We note $V_t = \sum_{n=0}^{t} A_t\cdots A_{n+1}B$ and $|V_{T_u}| > u$. Moreover r_ξ is bounded from above and below by positive constants. So it suffices to show

$$\limsup_{u\to\infty}\mathbb{E}^{\mathcal{D}}\left(\left|\sum_{n=0}^{T_u}\frac{N_u^{1/\xi}A_{T_u}\cdots A_{n+1}B}{|A_{T_u}\cdots A_1X_0|}\mathbf{1}_{\{T_u<K_1\}}\right|^{2\xi}\right) \quad < \quad \infty \tag{3.32}$$

In the rest of the proof, we write $c, c_1, c_2, \dots$ for constants whose values have no importance and depend on the context. Moreover, we use the notation

$$\Pi_{i,j} = \begin{cases} A_i A_{i-1} \cdots A_j & i \geq j \\ 1 & i < j \end{cases}$$

If $2\xi > 1$, by Minkowski's inequality it suffices to show

$$\limsup_{u\to\infty} \sum_{n=0}^{\infty} \left(\mathbb{E}^{\mathcal{D}} \left| N_u^{1/\xi} \frac{\Pi_{T_u,n+1} B}{|\Pi_{T_u,1} X_0|} \mathbf{1}_{\{n \leq T_u < K_1\}} \right|^{2\xi} \right) < \infty$$

The sum on the left side is bounded by

$$\begin{aligned} & c \sum_{n=0}^{\infty} \mathbb{E}^{\mathcal{D}} \left(\frac{|\Pi_{T_u,n+1} X_n|^{2\xi} N_u^2 \mathbf{1}_{\{n \leq T_u < K_1\}}}{|\Pi_{T_u,n+1} X_n|^{2\xi} |\Pi_{n,1} X_0|^{2\xi}} \right) \\ \leq\ & c \sum_{n=0}^{\infty} \mathbb{E} \left(\frac{|\Pi_{T_u,n+1} X_n|^{\xi}}{|\Pi_{n,1} X_0|^{\xi}} N_u^2 \mathbf{1}_{\{n \leq T_u < K_1\}} \right) \end{aligned}$$

If $2\xi < 1$, due to subadditivity, the sum on the left side of (3.32) is bounded by

$$\begin{aligned} & \sum_{n=0}^{\infty} \mathbb{E}^{\mathcal{D}} \left(\frac{|\Pi_{T_u,n+1} B|^{2\xi}}{|\Pi_{T_u,1} X_0|^{2\xi}} N_u^2 \mathbf{1}_{\{n \leq T_u < K_1\}} \right) \\ \leq\ & c \sum_{n=0}^{\infty} \mathbb{E}^{\mathcal{D}} \left(\frac{|\Pi_{T_u,n+1} X_n|^{2\xi}}{|\Pi_{T_u,1} X_0|^{2\xi}} N_u^2 \mathbf{1}_{\{n \leq T_u < K_1\}} \right) \\ \leq\ & c \sum_{n=0}^{\infty} \mathbb{E} \left(\frac{|\Pi_{T_u,n+1} X_n|^{\xi}}{|\Pi_{n,1} X_0|^{\xi}} N_u^2 \mathbf{1}_{\{n \leq T_u < K_1\}} \right) \end{aligned}$$

This is the same sum as in the previous case except for the multiplicative constant. So in either case we need to show

$$\limsup_{u\to\infty} \sum_{n=0}^{\infty} \mathbb{E} \left(\frac{|\Pi_{T_u,n+1} X_n|^{\xi}}{|\Pi_{n,1} X_0|^{\xi}} N_u^2 \mathbf{1}_{\{n \leq T_u < K_1\}} \right) < \infty \tag{3.33}$$

We write the sum of (3.33) as

$$\sum_{n=0}^{\infty} \sum_{m=n}^{\infty} \mathbb{E} \left[\frac{|\Pi_{m,n+1} X_n|^{\xi}}{|\Pi_{n,1} X_0|^{\xi}} N_u^2 \mathbf{1}_{\{m < K_1\}} \mathbf{1}_{\{T_u = m\}} \right]$$

Since $\|\Pi_{n+m,n+1}\| \to 0$ a.s. as $m \to \infty$ (cf. Buraczewski et al.[29], theorem 4.1.3), it is useful to consider the family of sets $S_1(\epsilon)$ for each $\epsilon > 0$:

$$S_1(\epsilon) = \{\exists N_1 \geq 1, \text{ such that } \forall m \geq N_1, \|\Pi_{n+m,n+1}\| < \epsilon\} \tag{3.34}$$

Note

$$\mathbb{P}(S_1(\epsilon)) = 1$$

Thus we have

$$\sum_{n=0}^{\infty} \sum_{m=n}^{\infty} \mathbb{E} \left[\frac{|\Pi_{m,n+1} X_n|^{\xi}}{|\Pi_{n,1} X_0|^{\xi}} N_u^2 \mathbf{1}_{\{m < K_1\}} \mathbf{1}_{\{T_u = m\}} \mathbf{1}_{\{S_1(\epsilon)\}} \right] = \mathscr{D} \tag{3.35}$$

Let's temporarily specialize the general norm to 1-norm. In this case

$$\begin{aligned} |\Pi_{n,1}X_0|_1 &= \sum_{i=1}^{d}\sum_{j=1}^{d}\Pi_{n,1}(i,j)X_0(j) \\ &= d\sum_k X_0(k)\sum_{i=1}^{d}\frac{1}{d}\sum_{j=1}^{d}\frac{\Pi_{n,1}(i,j)X_0(j)}{\sum_k X_0(k)} \end{aligned}$$

where $\Pi_{n,1}(i,j)$ refers to the (i,j)-th component of matrix $\Pi_{n,1}$ and $X_0(j)$ to the j-th component of X_0. By theorem 2 of Hennion [70], for every $\epsilon > 0$:

$$\mathbb{P}\left(\exists N_2 > T \text{ such that } \sup_{n\geq N_2}\left|\frac{1}{n}\log\left[\frac{|\Pi_{n,1}X_0|_1}{d|X_0|_1}\right] - \gamma\right| < \epsilon\right) = 1$$

which implies

$$\mathbb{P}\left(\exists N_2 > T \text{ such that } \forall n \geq N_2,\ |\Pi_{n,1}X_0|_1 > d|X_0|_1 e^{(\gamma-\epsilon)n}\right) = 1$$

By equivalence of vector norms on R^d, $|\Pi_{n,1}X_0| \geq c_1|\Pi_{n,1}X_0|_1$ and $|X_0|_1 \geq c_2|X_0| = c_2$ for some constants $c_1, c_2 > 0$. Thus we may define sets $S_2(\epsilon)$:

$$S_2(\epsilon) = \left\{\exists N_2 > T \text{ such that } \forall n \geq N_2,\ |\Pi_{n,1}X_0| > c\cdot d\cdot e^{(\gamma-\epsilon)n}\right\}$$

where $c > 0$ is a constant. With $S_2(\epsilon)$ defined as such, we have $\mathbb{P}(S_2(\epsilon)) = 1$. Now that we have $\mathbb{P}(S_1(\epsilon)) = 1$ and $\mathbb{P}(S_2(\epsilon)) = 1$, we may restrict the expectation in (3.35) to the set $S_1(\epsilon)\cap S_2(\epsilon)$, i.e.

$$\begin{aligned} \mathscr{D} &= \sum_{n=0}^{\infty}\sum_{m=n}^{\infty}\mathbb{E}\left[\frac{|\Pi_{m,n+1}X_n|^\xi}{|\Pi_{n,1}X_0|^\xi}N_u^2\mathbf{1}_{\{m<K_1\}}\mathbf{1}_{\{T_u=m\}}\mathbf{1}_{\{S_1(\epsilon)\cap S_2(\epsilon)\}}\right] \\ &= \left(\sum_{n=0}^{N_2-1}\sum_{m=n}^{n+N_1-1} + \sum_{n=0}^{N_2-1}\sum_{m=n+N_1}^{\infty} + \sum_{n=N_2}^{\infty}\sum_{m=n}^{n+N_1-1} + \sum_{n=N_2}^{\infty}\sum_{m=n+N_1}^{\infty}\right) \\ &\quad \mathbb{E}\left[\frac{|\Pi_{m,n+1}X_n|^\xi}{|\Pi_{n,1}X_0|^\xi}N_u^2\mathbf{1}_{\{m<K_1\}}\mathbf{1}_{\{T_u=m\}}\mathbf{1}_{\{S_1(\epsilon)\cap S_2(\epsilon)\}}\right] \\ &= \mathscr{D}_1 + \mathscr{D}_2 + \mathscr{D}_3 + \mathscr{D}_4 \end{aligned}$$

To show $\mathscr{D} < \infty$, it suffices to show $\mathscr{D}_i < \infty$ for each $i = 1,2,3,4$. $\mathscr{D}_1$ sums only finitely many terms, so it suffices to show for each fixed n and m,

$$\mathbb{E}\left[\frac{|\Pi_{m,n+1}X_n|^\xi}{|\Pi_{n,1}X_0|^\xi}N_u^2\mathbf{1}_{\{m<K_1\}}\mathbf{1}_{\{T_u=m\}}\mathbf{1}_{\{S_1(\epsilon)\cap S_2(\epsilon)\}}\right] < \infty \tag{3.36}$$

Firstly, we observe $|\Pi_{n,1}X_0|$ is bounded from below by a positive constant for any $n < \infty$

$$|\Pi_{n,1}X_0|_2 > d^{-1/2}\min_l X_0(l)\|\Pi_{n,1}\|_2 > 0$$

The first inequality is due to Kesten [84]. Thus (3.36) is implied by

$$\mathbb{E}\left[\|\Pi_{m,n+1}\|^\xi N_u^2\mathbf{1}_{\{m<K_1\}}\right] < \infty$$

By Hölder's inequality, the left side of the above inequality is bounded by

$$\begin{aligned} & (\mathbb{E}\|\Pi_{m,n+1}\|^{p\xi})1/p[\mathbb{E}(N_u^{2q}\mathbf{1}_{\{m<K_1\}})]^{1/q} \\ \leq \quad & (\mathbb{E}\|A\|^{p\xi})^{(m-n)/p}[\mathbb{E}(K_1^{2q}\mathbf{1}_{\{m<K_1\}})]^{1/q} \end{aligned}$$

where $p, q > 1$ and $1/p + 1/q = 1$. Because $\mathbb{E}\|A\|^s < \infty$ for some $s > \xi$ as assumed in condition (I), p can be chosen sufficiently close to 1 such that $\mathbb{E}\|A\|^{p\xi} < \infty$. Meanwhile

$$\mathbb{E}(K_1^{2q}\mathbf{1}_{\{m<K_1\}}) \leq \sum_{i=m+1}^{\infty} i^{2q}\mathbb{P}(K_1 > i-1) \leq \sum_{i=m+1}^{\infty} i^{2q}\rho_0 b_0^{i-1} < \infty$$

where we have used lemma 3.5 to reach the last line.

As for $\mathscr{D}_2$, we note $\|\Pi_{m,n+1}\|\mathbf{1}_{\{S_1(\epsilon)\}} < \epsilon$ for $m \geq n+N_1$, and for $n < N_2$, $|\Pi_{n,1}X_0| > 0$. So for $\mathscr{D}_2 < \infty$, it suffices to show

$$\sum_{n=0}^{N_2-1}\sum_{m=n+N_1}^{\infty}\mathbb{E}(N_u^2\mathbf{1}_{\{K_1>m\}}) < \infty$$

The left side is bounded by

$$\begin{aligned} & \sum_{n=0}^{N_2-1}\sum_{m=n+N_1}^{\infty}\sum_{i=m+1} i^2\mathbb{P}(K_1 > i-1) \\ = \quad & \sum_{n=0}^{N_2-1}\sum_{m=n+N_1}^{\infty}(c_2m^2 + c_1m + c_0)b_0^m < \infty \end{aligned}$$

where c_0, c_1, c_2 are constants.

For $\mathscr{D}_3$, we have for $n \geq N_2$, $|\Pi_{n,1}X_0|\mathbf{1}_{\{S_1(\epsilon)\}} > c \cdot de^{(\gamma-\epsilon)n}$. Thus

$$\begin{aligned} \mathscr{D}_3 \quad & < \quad \sum_{n=N_2}^{\infty}\sum_{m=n}^{n+N_1-1}\mathbb{E}\left[\frac{e^{(\epsilon-\gamma)n}}{cd}\|\Pi_{m,n+1}\|^{\xi}N_u^2\mathbf{1}_{\{K_1>m\}}\right] \\ & \leq \quad \frac{1}{cd}\sum_{n=N_2}^{\infty}e^{(\epsilon-\gamma)n}\sum_{m=n}^{n+N_1-1}\sum_{i=m+1}^{\infty}i^2\mathbb{E}\left[\|\Pi_{m,n+1}\|^{\xi}\mathbf{1}_{\{K_1=i\}}\right] \\ & \leq \quad \frac{1}{cd}\sum_{n=N_2}^{\infty}e^{(\epsilon-\gamma)n}\sum_{m=n}^{n+N_1-1}\sum_{i=m+1}^{\infty}i^2[\mathbb{E}\|\Pi_{m,n+1}\|^{p\xi}]^{1/p}[\mathbb{P}(K_1 > i-1)]^{1/q} \\ & \leq \quad \frac{1}{cd}\sum_{n=N_2}^{\infty}e^{(\epsilon-\gamma)n}[\mathbb{E}\|A\|^{p\xi}]^{-n/p}\sum_{m=n}^{n+N_1-1}[\mathbb{E}\|A\|^{p\xi}]^{m/p}\sum_{i=m+1}^{\infty}i^2\rho_o^{1/q}b_0^{(i-1)/q} \end{aligned}$$

The last sum over i evaluates to $(c_2m^2+c_1m+c_0)b_0^{m/q}$. So, by Cauchy-Schwarz inequality we have

$$\begin{aligned} \mathscr{D}_3 \quad \leq \quad & c_3\rho_o^{1/q}\sum_{n=N_2}^{\infty}e^{(\epsilon-\gamma)n}[\mathbb{E}\|A\|^{p\xi}]^{-n/p} \times \\ & \left(\sum_{m=n}^{n+N_1-1}[\mathbb{E}\|A\|^{p\xi}]^{2m/p}b_0^{2m/q}\right)^{1/2}\left(\sum_{m=n}^{n+N_1-1}(c_2m^2 + c_1m + c_0)^2\right)^{1/2} \end{aligned}$$

where c_3 is a positive constant. Since positive multiplicative constants do not affect the finiteness, we shall no longer keep track of their values but recycle the symbols $c, c_0, c_1, c_2, \ldots$ to denote different constants in different contexts. In this notation, the second sum over m is bounded by $c_4(n + N_1 - 1)$. We have

$$\mathscr{D}_3 \leq c \sum_{n=N_2}^{\infty} e^{(\epsilon-\gamma)n} [\mathbb{E}\|A\|^{p\xi}]^{-n/p} (n + N_1 - 1) \left(\sum_{m=n}^{n+N_1-1} [\mathbb{E}\|A\|^{p\xi}]^{2m/p} b_0^{2m/q} \right)^{1/2}$$

To show $\mathscr{D}_3 < \infty$, it is sufficient to show

$$\sum_{n=N_2}^{\infty} n e^{(\epsilon-\gamma)n} b_0^{n/q} < \infty$$

Condition (I) gives $\mathbb{E}\|A\|^s < \infty$. So we may choose $p = s/\xi$, i.e. $q = (1 - \xi/s)^{-1}$. We have assumed $b_0^{1/q} e^{-\gamma} < 1$, so there exists $\epsilon > 0$ as small as to make $e^{(\epsilon-\gamma)n} b_0^{1/q} < 1$. Therefore the last inequality holds. We have shown $\mathscr{D}_3 < \infty$.

To see $\mathscr{D}_4 < \infty$, we observe

$$\frac{|\Pi_{m,n+1} X_n|^\xi}{|\Pi_{n,1} X_0|^\xi} \mathbf{1}_{\{S_1(\epsilon) \cap S_2(\epsilon)\}} < c e^{(\epsilon-\gamma)n}$$

using the previous convention about multiplicative constants. Thus, to show $\mathscr{D}_4 < \infty$, it suffices to show

$$\sum_{n=N_2}^{\infty} e^{(\epsilon-\gamma)n} \sum_{m=n+N_1}^{\infty} \mathbb{E}(N_u^2 \mathbf{1}_{\{m<K_1\}}) < \infty$$

The left side is bounded by

$$\begin{aligned} & c \sum_{n=N_2}^{\infty} e^{(\epsilon-\gamma)n} \sum_{m=n+N_1}^{\infty} \sum_{i=m+1}^{\infty} i^2 b_0^{i-1} \\ \leq \; & c \sum_{n=N_2}^{\infty} e^{(\epsilon-\gamma)n} \sum_{m=n+N_1}^{\infty} (c_2 m^2 + c_1 m + c_0) b_0^m \\ \leq \; & c \sum_{n=N_2}^{\infty} e^{(\epsilon-\gamma)n} b_0^{n+N_1} [c_2(n + N_1)^2 + c_1(n + N_1) + c_0] \end{aligned}$$

Since $b_0 < 1$, it is clear $b_0 < b_0^{1/q}$. As argued in the case of $\mathscr{D}_3$, $b_0^{1/q} e^{\epsilon-\gamma} < 1$. So the last sum is finite. Thus $\mathscr{D}_4 < \infty$ and

$$\mathscr{D} = \mathscr{D}_1 + \mathscr{D}_2 + \mathscr{D}_3 + \mathscr{D}_4 < \infty$$

The proof is complete. □

3.6 Estimation of tail index

3.6.1 The algorithm

The idea is to estimate $\log[\lambda(\alpha)]$ according to

$$\log[\lambda(\alpha)] = \lim_{n\to\infty} \frac{1}{n} \log \left(\mathbb{E}|\Pi_{n,1} X_0|^\alpha \right) \tag{3.37}$$

and then solve $\log[\lambda(\xi)] = 0$ for ξ. The difficulty with brute-force simulation and estimation is that, when n is large, the variance of $|\Pi_{n,1}X_0|^\alpha$ is very large too, resulting in uselessly inaccurate estimations. Vanneste [114] proposed an importance sampling algorithm using a resampling step: we divide the estimation of $\log[\lambda(\alpha)]$ into n steps and we simulate K realizations of $A_n, \ldots, A_1$ and $X_0, X_1, \ldots, X_n$. We denote the l-th realization with an superscript l. The estimator $\mathscr{E}_\alpha$ for $\log[\lambda(\alpha)]$ is as follows:

$$\mathscr{E}_\alpha = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{K}\sum_{l=1}^{K} |A_i^l X_{i-1}^{w_{l,i-1}}|^\alpha\right) \tag{3.38}$$

where A_i^l, X_I^l denote the l-th realization of A_i and X_i respectively, and the random variable $w_{l,i-1}$ has conditional distribution

$$\begin{aligned}\mathbb{P}(w_{l,i-1} = j | w_{1,i-2}, \ldots, w_{K,i-2}) &= \frac{a_{i-1}^j}{b_{i-1}}\\ a_{i-1}^j &= |A_{i-1}^j X_{i-2}^{w_{j,i-2}}|^\alpha\\ b_{i-1} &= \sum_{l=1}^{K} a_{i-1}^l\end{aligned}$$

Figure 3.1 shows a possible realization of the resampling procedure and algorithm 1 outlines an implementation of $\mathscr{E}_\alpha$. The norm $|\cdot|$ is not restricted to any particular form, but obviously must be consistent with one's choice of the norm when computing the right eigenfunction, which is detailed in §3.8.1.

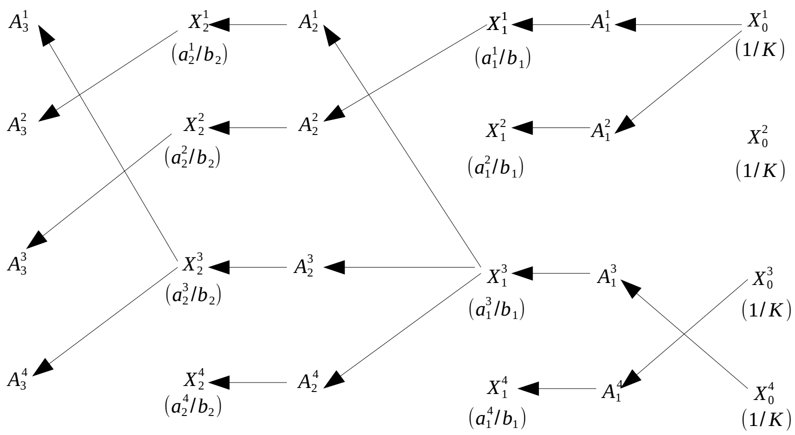


Figure 3.1: A possible realization of the re-sampling procedure. $n = 3$, $K = 4$. A number in a parenthesis indicates the probability of the unit vector above it being chosen to the next step.

Note for an GARCH(p, q) processes, the A_i matrices have dimension $d \times d$, where $d = p + q - 1$, and the X_i are d-dimensional unit vectors.

Algorithm 1 Algorithm for estimating $\log[\lambda(\alpha)] = \lim_{n\to\infty} \frac{1}{n} \log\left(\mathbb{E}|\Pi_{n,1}|^{\alpha}\right)$

```
procedure ℰ_α(n, K)
    Define K d-dimensional vectors X^1, ..., X^K
    Define K d-dimensional vectors Y^1, ..., Y^K
    Define K-dimensional vector a ← (1, 1, ..., 1)          ▷ Initialize the weights
    for i from 1 to K do                                     ▷ Generate initial unit vectors
        for k from 1 to d do
            Generate a U(0, 1) variable U
            X^i(k) ← U                                   ▷ X^i(k) is the k-th component of X^i
        end for
        X^i ← X^i/|X^i|
    end for
    Λ ← 0
    for j from 1 to n do
        Define K-dimensional vector Q
        Q(k) ← Σ_{i=1}^{k} a(i) for all k = 1, 2, ..., K
        Generate K d × d random matrices A^1, ..., A^K.
        for k from 1 to K do
            Generate a U(0, Q(K)) variable U.
            l ← min{1 ≤ i ≤ K : Q(i) > U}
            Y^k ← A^k X^l
            a(k) ← |Y^k|
            Y^k ← Y^k/a(k)
            a(k) ← a(k)^α
        end for
        For all k = 1, ..., K, X^k ← Y^k
        Λ ← Λ + (1/n) log(Q(K)/K)
    end for
    Λ ← Λ + (1/n) log[(1/K) Σ_{k=1}^{K} a(k)]
    return Λ
end procedure
```

3.7 Sampling from the shifted conditional distribution

While the Radon-Nikodym derivative of the shifted conditional probability measure with respect to the original measure is given by (3.17), the procedure of sampling from the shifted conditional distribution is not trivial. For this purpose, we first draw a sequence of $\chi^2(1)$ random variables, call them $Z_1^2, Z_2^2, \ldots, Z_K^2$. Then for each $1 \le i \le K$, we compute

$$W_i = \frac{|A(Z_i^2) \cdot X_{i-1}|^{\xi} r_{\xi}(A(Z_i^2) \cdot X_{i-1})}{r_{\xi}(X_{i-1})}$$

Finally we draw a sample Z'^2 from the sequence $Z_1^2, Z_2^2, \ldots, Z_K^2$, such that

$$\mathbb{P}(Z'^2 = Z_i^2) = \frac{W_i}{\sum_{j=1}^{K} W_j}$$

Then the matrix A_t from the shifted distribution conditional on X_{t-1} is $A(Z'^2)$.

3.8 An example: GARCH(2,1) processes

As an example of the algorithms described in the previous sections, we consider GARCH(2,1) processes. In this particular case we have

$$\sigma_t^2 = \omega + \alpha_1 R_{t-1}^2 + \alpha_2 R_{t-2}^2 + \beta_1 \sigma_{t-1}^2$$

Or in matrix forms

$$\begin{aligned} \begin{pmatrix} \sigma_t^2 \\ R_{t-1}^2 \end{pmatrix} &= \begin{pmatrix} \alpha_1 Z_{t-1}^2 + \beta_1 & \alpha_2 \\ Z_{t-1}^2 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{t-1}^2 \\ R_{t-2}^2 \end{pmatrix} + \begin{pmatrix} \omega \\ 0 \end{pmatrix} \\ V_t &= A_t V_{t-1} + B \end{aligned} \tag{3.39}$$

where R_t is the t-th observation of the sequence in question and Z_t are i.i.d $N(0,1)$ random variables. In the following we check the conditions of theorem 3.3 against this process.

As mentioned in remark 3.1, Bollerslev's assumption $\alpha_1 + \alpha_2 + \beta_1 < 1$ implies that the top Lyapunov exponent associated with the model (3.39) is negative. In the following we check the condition of theorem 3.3 with regard to this model.

Lemma 3.6. *Assume Z_t are i.i.d $N(0,1)$ random variables. Then $b^{1-\xi/s} < e^{\gamma}$, where b is defined in lemma 3.5, ξ is the tail index of the stationary distribution of the markov chain V_t as defined by (3.39) and γ is the top Lyapunov exponent of the matrices A_t in (3.39).*

Proof. First of all, we note s of condition (I), hypothesis 1 can be arbitrarily large since Z_t are i.i.d $N(0,1)$ random variables. Therefore $b^{1-\xi/s} < e^{\gamma}$ holds if $b < e^{\gamma}$. To show the latter inequality holds, it suffices to show there exists $\theta \in (0,\xi)$ such that $b_\theta < e^{\gamma}$, where b_θ is defined in lemma 3.4.

If $\theta < 1$, b_θ is given by (3.22):

$$b_\theta = \lambda(\theta)\left[1 + \frac{|B|^\theta r_\theta(\tilde{B})}{\lambda(\theta) M_\theta \underline{r}_\theta}\right]$$

Because M_θ can be chosen arbitrarily large, $b_\theta < e^{\gamma}$ holds if $\lambda(\theta) < e^{\gamma}$, or equivalently $\log(\lambda(\theta)) < \gamma$. Applying theorem 2 of Hennion [70] gives, for an arbitrary fixed $\epsilon > 0$, there exists $N > T$ such that for all $n > N$, $\frac{1}{n}\log\|\Pi_{n,1}\| < \gamma + \epsilon$. This implies

$$\frac{1}{n}\log\left(\mathbb{E}\|\Pi_{n,1}\|^\theta\right) < \theta(\gamma + \epsilon)$$

Thus

$$\log(\lambda(\theta)) = \lim_{n\to\infty}\frac{1}{n}\log\left(\mathbb{E}\|\Pi_{n,1}\|^\theta\right) < \theta(\gamma+\epsilon)$$

Thus, if θ is chosen such that $0 < \theta < \frac{\gamma}{\gamma+\epsilon} < 1$, we have $b_\theta < e^{\gamma}$. The proof is complete. □

3.8.1 Evaluation of the right eigenfunction

Recall that, for $0 < \theta < \xi$, the right eigenfunction $r_\theta(\cdot)$ corresponding to eigenvalue $\lambda(\theta)$ of the operator $\mathscr{P}^\theta$ defined in (3.15) satisfies

$$\mathbb{E}\left[|Ax|^\theta r_\theta(A\cdot x)\right] = \mathscr{P}^\theta r_\theta(x) = \lambda(\theta) r_\theta(x) \tag{3.40}$$

In the following we take $|\cdot|$ as the Euclidean norm. Then x can be written as $x = (\cos w, \sin w)^\top$, $w \in (0, \pi/2)$ and

$$Ax = \begin{pmatrix} \alpha_2 \sin w + (Z_t^2 \alpha_1 + \beta_1) \cos w \\ Z^2 \cos w \end{pmatrix}$$

Let $A \cdot x = (\cos\varphi, \sin\varphi)^\top$. Then

$$\begin{aligned} Z_t^2 &= \frac{\tan\varphi(\alpha_2 \sin w + \beta_1 \cos w)}{\cos w(1 - \alpha_1 \tan\varphi)} \\ \frac{dZ_t^2}{d\varphi} &= \frac{(\alpha_2 \sin w + \beta_1 \cos w)\sec^2\varphi}{\cos w(1 - \alpha_1 \tan\varphi)^2} \end{aligned} \tag{3.41}$$

Using (3.41) we can write

$$|Ax|^\theta = (\alpha_2^2 + \beta_1^2)^{\theta/2} \frac{\sin^\theta\left(w + \arctan\frac{\beta_1}{\alpha_2}\right)}{\cos^\theta\varphi(1 - \alpha_1 \tan\varphi)^\theta}$$

It also becomes clear from (3.41)

$$x \in \left\{(\cos w, \sin w)^\top : w \in \Omega = \left[0, \arctan\frac{1}{\alpha_1}\right)\right\} \quad t = 0, 1, 2, \ldots$$

Define

$$h_\theta(w) : w \in \Omega \to (\underline{r}_\theta, \bar{r}_\theta) \ni r_\theta((\cos w, \sin w)^\top)$$

Then (3.40) can be rewritten as

$$\begin{aligned} \lambda(\theta)h_\theta(w) &= \int_\Omega (\alpha_2^2 + \beta_1^2)^{\theta/2} \frac{\sin^\theta\left(w + \arctan\frac{\beta_1}{\alpha_2}\right)}{\cos^\theta\varphi(1 - \alpha_1 \tan\varphi)^\theta} f_{\chi^2}\left[\frac{\tan\varphi(\alpha_2 \sin w + \beta_1 \cos w)}{\cos w(1 - \alpha_1 \tan\varphi)}\right] \times \\ &\quad h_\theta(\varphi) \frac{(\alpha_2 \sin w + \beta_1 \cos w)\sec^2\varphi}{\cos w(1 - \alpha_1 \tan\varphi)^2} d\varphi \\ \lambda(\theta)h_\theta(w) &= \int_\Omega H_\theta(w, \varphi) h_\theta(\varphi) d\varphi \end{aligned} \tag{3.42}$$

where $f_{\chi^2}(\cdot)$ is the probability density function of the χ^2 distribution; the function $H_\theta(w, \varphi)$ has been defined for convenience. We may approximate the last integral with a sum and the function $h_\theta(\cdot)$ with a vector:

$$\lambda(\theta)h_\theta(i\Delta_n) \approx \Delta_n \sum_{j=0}^{n-1} H_\theta(i\Delta_n, j\Delta_n) h_\theta(j\Delta_n)$$

where $\Delta_n = \frac{\arctan(1/\alpha_1)}{n}$. Thus $\lambda(\theta)$ can be found as the spectral radius of matrix $\Delta_n H_\theta(i\Delta_n, j\Delta_n)$ and $\{h_\theta(i\Delta_n)\}_{i=1,2,\ldots,n}$ as the associated eigenvector (cf. Collamore and Mentemeier [37], lemma 2.2).

3.8.2 The algorithm

In this section we describe the implementation of the proposed estimator $\mathcal{E}_u$. In previous sections we have shown that $\mathcal{E}_u$ is efficient for GARCH$(2,1)$ and described how the right eigenfunctions can be approximately evaluated. It remains to choose the set $\mathcal{C} = \{v \in \chi, |V_n| < M\}$. As given in (3.24), $M = \sup_{\theta\in\Theta} M_\theta$ while M_θ is in turn given by (3.21) or (3.23). One viable option is to choose M_θ such that $b_\theta = \frac{1+\lambda(\theta)}{2} < 1$. For the range of θ, which is denoted Θ, we take $\Theta = [0.1, 0.1 + 0.99 \times (\xi - 0.1)]$, where ξ is the tail index.

Before coding the algorithm of estimating $\mathbb{P}(|V| > u)$, we first need a procedure for simulating a sample path under the dual transition kernel of the MA-process (X_t, S_t). For convenience, let $A(W)$ denotes the matrix A_t that appears in (3.39) with $Z_{t-1}^2 = W$. We describe the procedure in algorithm 2.

With a procedure established for simulating a sample path under the dual transition kernel, we are at a position to describe the procedure for estimating $\mathbb{P}(|V| > u)$. We present it as algorithm 3.

3.8.3 Simulaton and Results

In this section we present estimation results of the algorithm outlined in §3.8.2 when applied to a few real-world GARCH$(2,1)$ time series. Table 3.1 lists the estimated parameters of the S& P 500 index and the Dow Jones Industrial Average index. Judging by the *Akaike Information Criterior*, the two series are indeed better described by GARCH$(2,1)$ models than by the simpler GARCH$(1,1)$ models. In the following table 3.2 we tabulate

	ω	α_1	α_2	β_1	ξ	AIC
SP500 GARCH$(1,1)$	7.4×10^{-6}	0.15		0.72		-7.0270
SP500 GARCH$(2,1)$	9.2×10^{-6}	0.088	0.097	0.65		-7.0274
DJIA GARCH$(1,1)$	5.6×10^{-6}	0.16		0.73	3.76	-7.1720
DJIA GARCH$(2,1)$	7.3×10^{-6}	0.075	0.13	0.65	4.12	-7.1752

Table 3.1: GARCH$(1,1)$ and GARCH$(2,1)$ models of returns of Standard & Poor 500 index (SP500) and the Dow Jones Industrial Average index (DJIA). The price data are downloaded from Yahoo and cover the period 2012-01-01 to 2014-12-31. 753 observations are included. AIC stands for *Akaike Information Criterior.* The model parameters are estimated using the fGARCH package [31] of the R language.

the values of u, M, and $\mathbb{P}(|V| > u)$ for the SP500 and DJIA series described in table 3.1.

Algorithm 2 Algorithm for simulating a sample path under the dual kernel

```
procedure SIM(u, V_0)                      ▷ Simulate a path under the dual kernel
    V ← V_0
    X ← V_0/|V_0|
    B ← (ω, 0)^⊤
    (N_u, S, I, m) ← (0, 0, 0, 0)
    loop
        Generate U(0, 1) r.v. U
        if I = 0 then          ▷ u is not exceeded. Sample from the shifted distribution
            draw W|X from the shifted distribution as described in §3.7
        else                   ▷ u is exceeded. Sample from the original distribution
            draw W from χ^2(1) distribution
        end if
        V ← A(W)V + B
        if I = 0 then          ▷ Once u is exceeded, X, S don't need to be computed.
            S ← log |A(W)X|
            X ← A(W)X / |A(W)X|
        end if
        if I = 0 and |V| ≤ M then    ▷ The chain returns to C before exceeding u.
            (N_u, S) ← (0, 0)
            m ← m + 1
        else if I = 0 and |V| > M then
            if |V| > u then                      ▷ u is exceeded for the 1st time
                N_u ← N_u + 1
                I ← 1
            end if
        else if I = 1 and |V| ≤ M then
            break
        else if I = 1 and |V| > u then
            N_u ← N_u + 1
        end if
    end loop
    p ← N_u exp(−ξS) r_ξ(V_0/|V_0|)/r_ξ(X)
    q ← m + 1
    return (p, q)
end procedure
```

Algorithm 3 Algorithm for estimating $\mathbb{P}(|V| > u)$

procedure $\mathcal{E}_u(u, K)$
  Define $W_0 = (\frac{\omega}{1-\beta_1}, 0)^\top$ ▷ Create a sample of $\eta : \eta(S) = \pi(S)/\pi(\mathcal{C}), S \subseteq \mathcal{C}$
  Define $B = (\omega, 0)^\top$.
  for i from 1 to 1000 **do**
    Generate $\chi^2(1)$ random variable Z_i^2
    generate random matrix $A_i(Z_i^2)$.
    $W_i \leftarrow A_i W_{i-1} + B$
  end for
  for i from 1 to 800 **do** ▷ Discard the first 20% of the sample.
    $W_i \leftarrow W_{i+200}$
  end for
  Discard all W_i for which $|W_i| > M$.
  $N \leftarrow$ number of remaining W_i
  Define a vector E of length K.
  $n \leftarrow 0$
  for i from 1 to K **do** ▷ Simulate K sample paths according to the dual transition kernel
    Generate $U(0, N)$ random variable U
    $k \leftarrow \lceil U \rceil$
    $(p, q) \leftarrow \text{sim}(u, W_k)$ ▷ Simulate a sample path
    $E_i \leftarrow p$
    $n \leftarrow n + q$;
  end for
  $\pi(\mathcal{C}) \leftarrow N/1000$
  $\hat{\mathcal{E}_u} \leftarrow \frac{\pi(\mathcal{C})}{n} \sum_{i=1}^{K} E_i$ ▷ $\hat{\mathcal{E}_u}$ is our estimate of $\mathbb{P}(|V| > u)$
  return $\hat{\mathcal{E}_u}$
end procedure

u	$\mathbb{P}(\|V\| > u)$	u	$\mathbb{P}(\|V\| > u)$
6	2.16×10^{-2}	17	4.03×10^{-4}
7	$1.06 \times 10^{-}2$	18	3.25×10^{-4}
8	6.37×10^{-3}	19	2.41×10^{-4}
9	4.06×10^{-3}	20	2.08×10^{-4}
10	2.95×10^{-3}	50	4.78×10^{-6}
11	2.04×10^{-3}	51	4.43×10^{-6}
12	1.55×10^{-3}	52	3.65×10^{-6}
13	9.92×10^{-4}	53	3.96×10^{-6}
14	7.94×10^{-4}	53	3.77×10^{-6}
15	6.24×10^{-4}	54	3.42×10^{-6}
16	5.09×10^{-4}	55	3.05×10^{-6}

Table 3.2: Estimates of $\mathbb{P}(|V| > u)$ for the GARCH$(2, 1)$ model of DJIA. $M = 4.65$, $\xi = 4.12$.

Chapter 4

The eigenvalues of the sample covariance matrix of a multivariate heavy-tailed stochastic volatility model

Anja Janssen, Thomas Mikosch,
Mohsen Rezapour & Xiaolei Xie
Bernoulli, to appear

Abstract

We consider a multivariate heavy-tailed stochastic volatility model and analyze the large-sample behavior of its sample covariance matrix. We study the limiting behavior of its entries in the infinite-variance case and derive results for the ordered eigenvalues and corresponding eigenvectors. Essentially, we consider two different cases where the tail behavior either stems from the iid innovations of the process or from its volatility sequence. In both cases, we make use of a large deviations technique for regularly varying time series to derive multivariate α-stable limit distributions of the sample covariance matrix. While we show that in the case of heavy-tailed innovations the limiting behavior resembles that of completely independent observations, we also derive that in the case of a heavy-tailed volatility sequence the possible limiting behavior is more diverse, i.e. allowing for dependencies in the limiting distributions which are determined by the structure of the underlying volatility sequence.

4.1 Introduction

4.1.1 Background and Motivation

The study of sample covariance matrices is fundamental for the analysis of dependence in multivariate time series. Besides from providing estimators for variances and covariances of the observations (in case of their existence), the sample covariance matrices are a starting point for dimension reduction methods like principal component analysis. Accordingly, the special structure of sample covariance matrices and their largest eigenvalues has been intensively studied in random matrix theory, starting with iid Gaussian observations and more recently extending results to arbitrary distributions which satisfy some moment assumptions like in the four moment theorem of Tao and Vu [109].

However, with respect to the analysis of financial time series, such a moment assumption is often not suitable. Instead, in this work, we will analyze the large sample behavior of sample covariance matrices under the assumption that the marginal distributions of our observations are regularly varying with index $\alpha < 4$ which implies that fourth moments do not exist. In this case, we would expect the largest eigenvalues of the sample covariance matrix to inherit heavy-tailed behavior as well; see for example Ben Arous and Guionnet [18], Auffinger et al. [7], Soshnikov [105, 106], Davis et al. [38], Heiny and Mikosch [68] for the case of iid entries. Furthermore, in the context of financial time series we have to allow for dependencies both over time and between different components and indeed it is the very aim of the analysis to discover and test for these dependencies from the resulting sample covariance matrix as has for example been done in Plerou et al. [95] and Davis et al. [44, 45]. The detection of dependencies among assets also plays a crucial role in portfolio optimization based on multi-factor prizing models, where principal component analysis is one way to derive the main driving factors of a portfolio; cf. Campbell et al. [30] and recent work by Lam and Yao [86].

The literature on the asymptotic behavior of sample covariance matrices derived from dependent heavy-tailed data is, however, relatively sparse up till now. Starting with the analysis of the sample autocorrelation of univariate linear heavy-tailed time series in Davis and Resnick [46, 47], the theory has recently been extended to multivariate heavy-tailed time series with linear structure in Davis et al. [44, 45], cf. also the recent survey article by Davis et al. [38]. But most of the standard models for financial time series such as GARCH and stochastic volatility models are non-linear. In this paper we will therefore focus on a class of multivariate stochastic volatility models of the form

$$X_{it} = \sigma_{it}\, Z_{it}\,, \qquad t \in \mathbb{Z}\,, \quad 1 \le i \le p, \tag{4.1}$$

where (Z_{it}) is an iid random field independent of a strictly stationary ergodic field (σ_{it}) of non-negative random variables; see Section 4.2 for further details. Stochastic volatility models have been studied in detail in the financial time series literature; see for example Andersen et al. [4], Part II. They are among the simplest models allowing for conditional heteroscedasticity of a time series. In view of independence between the Z- and σ-fields dependence conditions on (X_{it}) are imposed only via the stochastic volatility (σ_{it}). Often it is assumed that $(\log \sigma_{it})$ has a linear structure, most often Gaussian.

In this paper we are interested in the case when the marginal and finite-dimensional distributions of (X_{it}) have power-law tails. Due to independence between (σ_{it}) and (Z_{it}) heavy tails of (X_{it}) can be due to the Z- or the σ-field. Here we will consider two cases: (1) the tails of Z dominate the right tail of σ and (2) the right tail of σ dominates the tail of Z. The third case when both σ and Z have heavy tails and are tail-equivalent will not be considered in this paper. Case (1) is typically more simple to handle; see Davis and Mikosch [40, 41, 43] for extreme value theory, point process convergence and central limit theory with infinite variance stable limits. Case (2) is more subtle as regards the tails of the finite-dimensional distributions. The literature on stochastic volatility models with a heavy-tailed volatility sequence is so far sparse but the interest in these models has been growing recently; see Mikosch and Rezapour [89], Kulik and Soulier [85] and Janßen and Drees [78]. In particular, it has been shown that these models offer a lot of flexibility with regard to the extremal dependence structure of the time series, ranging from asymptotic dependence of consecutive observations (cf. [89]) to asymptotic independence of varying degrees (cf. [85] and [78]).

4.1.2 Aims, main results and structure

After introducing the general model in Section 4.2 we first deal with the case of heavy-tailed innovations and a light-tailed volatility sequence in Section 4.3. The first step in our analysis is to describe the extremal structure of the corresponding process by deriving its so-called tail process; see Section 4.2.3 and Proposition 4.1. This allows one to apply an infinite variance stable central limit theorem from Mikosch and Wintenberger [90] (see Appendix 4.6) to derive the joint limiting behavior of the entries of the sample covariance matrix of this model. This leads to the main results in the first case: Theorems 4.3 and 4.6. They say, roughly speaking, that all values on the off-diagonals of the sample covariance matrix are negligible compared to the values on the diagonals. Furthermore, the values on the diagonal converge, under suitable normalization, to independent α-stable random variables, so the limiting behavior of this class of stochastic volatility models is quite similar to the case of iid heavy-tailed random variables. This fairly tractable structure allows us also to derive explicit results about the asymptotic behavior of the ordered eigenvalues and corresponding eigenvectors which can be found in Sections 4.3.3 and 4.3.4. In particular, we will see that in this model the eigenvectors are basically the unit canonical basis vectors which describe a very weak form of extremal dependence. With a view towards portfolio analysis, our assumptions imply that large movements of the market are mainly driven by one single asset, where each asset is equally likely to be this extreme driving force.

In the second case of a heavy-tailed volatility sequence combined with light-tailed innovations, which we analyze in Section 4.4, we see that the range of possible limiting behaviors of the entries of the sample covariance matrix is more diverse and depends on the specific structure of the underlying volatility process. We make the common assumption that our volatility process is log-linear, where we distinguish between two different cases for the corresponding innovation distribution of this process. Again, for both cases, we first derive the specific form of the corresponding tail process (see Proposition 4.14) which then allows us to derive the limiting behavior of the sample covariance matrix entries, leading to the main results in the second case: Theorems 4.16 and 4.20. We show that the sample covariance matrix can feature non-negligible off-diagonal components, therefore clearly distinguishing from the iid case, if we assume that the innovations of the log-linear volatility process are convolution equivalent. We discuss concrete examples for both model specifications and the corresponding implications for the asymptotic behavior of ordered eigenvalues and corresponding eigenvectors at the end of Section 4.4.

Section 4.5 contains a small simulation study which illustrates our results for both cases and also includes a real-life data example for comparison. From the foreign exchange rate data that we use, it is notable that the corresponding sample covariance matrix features a relatively large gap between the largest and the second largest eigenvalue and that the eigenvector corresponding to the largest eigenvalue is fairly spread out, i.e., all its components are of a similar order of magnitude. This implies that the model discussed in Section 4.3 may not be that suitable to catch the extremal dependence of this data, and that there is not one single component that is most affected by extreme movements but instead all assets are affected in a similar way. We perform simulations for three different specifications of models from Sections 4.3 and 4.4. They illustrate that the models analyzed in Section 4.4 are capable of exhibiting more diverse asymptotic behaviors of the sample covariance matrix and in particular non-localized dominant eigenvectors.

Some useful results for the (joint) tail and extremal behavior of random products are gathered in Appendix 4.7. These results may be of independent interest when studying the extremes of multivariate stochastic volatility models with possibly distinct tail indices. We mention in passing that there is great interest in non-linear models for log-returns of speculative prices when the number of assets p increases with the sample size n. We understand our analysis as a first step in this direction.

4.2 The model

We consider a stochastic volatility model

$$X_{it} = \sigma_{it} Z_{it}\,, \qquad i,t \in \mathbb{Z}\,, \tag{4.2}$$

where (Z_{it}) is an iid field independent of a strictly stationary ergodic field (σ_{it}) of non-negative random variables. We write Z, σ, X for generic elements of the Z-, σ- and X-fields σ and Z are independent. A special case appears when $\sigma > 0$ is a constant: then (X_{it}) constitutes an iid field.

For the stochastic volatility model as in (4.1) we construct the multivariate time series

$$\mathbf{X}_t = (X_{1t}, \dots, X_{pt})', \quad t \in \mathbb{Z}, \tag{4.3}$$

for a given dimension $p \geq 1$. For $n \geq 1$ we write $\mathbf{X}^n = \operatorname{vec}\big((\mathbf{X}_t)_{t=1,\dots,n}\big) \in \mathbb{R}^{p\times n}$ and consider the non-normalized sample covariance matrix

$$\mathbf{X}^n(\mathbf{X}^n)' = (S_{ij})_{i,j=1,\dots,p}\,, \qquad S_{ij} = \sum_{t=1}^{n} X_{it}X_{jt}\,, \qquad S_i = S_{ii}\,. \tag{4.4}$$

4.2.1 Case (1): Z dominates the tail

We assume that Z is regularly varying with index $\alpha > 0$, i.e.,

$$\mathbb{P}(Z > x) \sim p_+ \frac{L(x)}{x^\alpha} \qquad \text{and} \qquad \mathbb{P}(Z < -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \tag{4.5}$$

where p_+ and p_- are non-negative numbers with $p_+ + p_- = 1$ and L is a slowly varying function. If we assume $\mathbb{E}[\sigma^{\alpha+\delta}] < \infty$ for some $\delta > 0$ then, in view of a result by Breiman [27] (see also Lemma 4.24), it follows that

$$\mathbb{P}(X > x) \sim \mathbb{E}[\sigma^\alpha]\,\mathbb{P}(Z > x) \quad \text{and} \quad \mathbb{P}(X < -x) \sim \mathbb{E}[\sigma^\alpha]\,\mathbb{P}(Z < -x)\,, \qquad x \to \infty\,, \tag{4.6}$$

i.e., X is regularly varying with index α. Moreover, we know from a result by Embrechts and Goldie [57] that for independent copies Z_1 and Z_2 of Z, $Z_1 Z_2$ is again regularly varying with index α; cf. Lemma 4.24. Therefore, using again Breiman's result under the condition that $\mathbb{E}[(\sigma_{i0}\sigma_{j0})^{\alpha+\delta}\mathbf{1}(i \neq j) + \sigma_{i0}^{\alpha+\delta}] < \infty$ for some $\delta > 0$, we have

$$\mathbb{P}(\pm X_{it}\,X_{jt} > x) \sim \begin{cases} \mathbb{E}[(\sigma_{it}\,\sigma_{jt})^\alpha]\,\mathbb{P}(\pm Z_i\,Z_j > x) & i \neq j\,, \\ \mathbb{E}[\sigma^\alpha]\,\mathbb{P}(Z^2 > x) & i = j\,, \end{cases} \qquad x \to \infty\,. \tag{4.7}$$

4.2.2 Case (2): σ dominates the tail

We assume that $\sigma \geq 0$ is regularly varying with some index $\alpha > 0$: for some slowly varying function ℓ,

$$\mathbb{P}(\sigma > x) = x^{-\alpha}\,\ell(x)\,,$$

and $\mathbb{E}[|Z|^{\alpha+\delta}] < \infty$ for some $\delta > 0$. Now the Breiman result yields

$$\mathbb{P}(X > x) \sim \mathbb{E}[Z_+^\alpha]\,\mathbb{P}(\sigma > x) \qquad \text{and} \qquad \mathbb{P}(X < -x) \sim \mathbb{E}[Z_-^\alpha]\,\mathbb{P}(\sigma > x)\,, \qquad x \to \infty\,.$$

Since we are also interested in the tail behavior of the products $X_{it}X_{jt}$ we need to be more precise about the joint distribution of the sequences (σ_{it}). We assume

$$\sigma_{it} = \exp\Big(\sum_{k,l=-\infty}^{\infty} \psi_{kl}\,\eta_{i-k,t-l}\Big)\,, \qquad i,t\in\mathbb{Z}\,, \tag{4.8}$$

where (ψ_{kl}) is a field of non-negative numbers (at least one of them being positive) (without loss of generality) $\max_{kl}\psi_{kl} = 1$ and (η_{it}) is an iid random field a generic element η satisfies

$$\mathbb{P}\big(\mathrm{e}^{\,\eta} > x\big) = x^{-\alpha}\,L(x)\,, \tag{4.9}$$

for some $\alpha > 0$ and a slowly varying function L. We also assume $\sum_{k,l}\psi_{kl} < \infty$ to ensure absolute summability of $\log\sigma_{it}$. A distribution of η that fits into this scheme is for example the exponential distribution; cf. also Rootzén [98] for further examples and extreme value theory for linear processes of the form $\sum_{l=-\infty}^{\infty}\psi_l\,\eta_{t-l}$.

4.2.3 Regularly varying sequences

In Sections 4.3.1 and 4.4.1 we will elaborate on the joint tail behavior of the sequences (σ_{it}), (X_{it}), $(\sigma_{it}\sigma_{jt})$, and $(X_{it}X_{jt})$. We will show that, under suitable conditions, these sequences are regularly varying with positive indices.

The notion of a *univariate regularly varying sequence* was introduced by Davis and Hsing [39]. Its extension to the multivariate case does not represent difficulties; see Davis and Mikosch [42]. An $\mathbb{R}^d$-valued strictly stationary sequence $(\mathbf{Y}_t)$ is *regularly varying with index* $\gamma > 0$ if each of the vectors $(\mathbf{Y}_t)_{t=0,\dots,h}$, $h\ge 0$, is regularly varying with index γ, i.e., there exist non-null Radon s μ_h on $[-\infty,\infty]^{d(h+1)}\backslash\{\mathbf{0}\}$ which are homogeneous of order $-\gamma$ such that

$$\frac{\mathbb{P}(x^{-1}(\mathbf{Y}_t)_{t=0,\dots,h}\in\cdot)}{\mathbb{P}(\|\mathbf{Y}_0\|>x)} \stackrel{v}{\to} \mu_h(\cdot)\,. \tag{4.10}$$

Here $\stackrel{v}{\to}$ denotes vague convergence on the Borel σ-field of $[-\infty,\infty]^{d(h+1)}\backslash\{\mathbf{0}\}$ and $\|\cdot\|$ denotes any given norm; see Resnick's books [96, 97] as general references to multivariate regular variation.

Following Basrak and Segers [15], an $\mathbb{R}^d$-valued strictly stationary sequence $(\mathbf{Y}_t)$ is regularly varying with index $\gamma>0$ if and only if there exists a sequence of $\mathbb{R}^d$-valued random vectors $(\mathbf{\Theta}_h)$ independent of a Pareto(γ) random variable Y, i.e., $\mathbb{P}(Y>x) = x^{-\gamma}$, $x>1$, for any $k\ge 0$,

$$\mathbb{P}(x^{-1}(\mathbf{Y}_0,\dots,\mathbf{Y}_k)\in\cdot\mid\|\mathbf{Y}_0\|>x) \stackrel{w}{\to} \mathbb{P}\big(Y\,(\mathbf{\Theta}_0,\dots,\mathbf{\Theta}_k)\in\cdot\big)\,, \qquad x\to\infty\,. \tag{4.11}$$

We call $(\mathbf{\Theta}_h)$ the *spectral tail process* of $(\mathbf{Y}_t)$ and $(Y\mathbf{\Theta}_h)$ the *tail process*. We will use both defining properties (i.e., (4.10) and (4.11)) of a regularly varying sequence.

4.3 Case (1): Z dominates the tail

4.3.1 Regular variation of the stochastic volatility model and its product processes

Proposition 4.1. *We assume the stochastic volatility model* (4.2) *and that Z is regularly varying with index $\alpha>0$ in the sense of* (4.5).

1. *If* $\mathbb{E}[\sigma^{\alpha+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$ *the sequence* $(X_{it})_{t\in\mathbb{Z}}$ *is regularly varying with index* α *and the corresponding spectral tail process* $(\Theta_h^i)_{h\geq 1}$ *vanishes.*

2. *For any* $i \neq j$*, if* $\mathbb{E}[(\sigma_{i0}\sigma_{j0})^{\alpha+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$ *then the sequence* $(X_{it}X_{jt})$ *is regularly varying with index* α *and the corresponding spectral tail process* $(\Theta_h^{ij})_{h\geq 1}$ *vanishes.*

Remark 4.2. If $\mathbb{E}[(\sigma_{ik}\sigma_{jl})^{\alpha+\varepsilon_{ik,jl}}] < \infty$ for some $\varepsilon_{ik,jl} > 0$ and any $(i,k) \neq (j,l)$ it is also possible to show the joint regular variation of the processes $(X_{it}X_{jt})$, $i \neq j$, with index α. The description of the corresponding spectral tail process is slightly tedious. It is not needed for the purposes of this paper and therefore omitted.

Proof. Regular variation of the marginal distributions of (X_{it}) and $(X_{it}X_{jt})$ follows from Breiman's result; see (4.6) and (4.7). As regards the regular variation of the finite-dimensional distributions of (X_{it}), we have for $h \geq 1$,

$$\begin{aligned}\mathbb{P}(|X_{ih}| > x \mid |X_{i0}| > x) &= \frac{\mathbb{P}(\min(|X_{i0}|, |X_{ih}|) > x)}{\mathbb{P}(|X_{i0}| > x)} \\ &\leq \frac{\mathbb{P}(\max(\sigma_{i0}, \sigma_{ih}) \min(|Z_{i0}|, |Z_{ih}|) > x)}{\mathbb{P}(|X_{i0}| > x)} \to 0\,, \qquad x \to \infty\,.\end{aligned}$$

In the last step we used Markov's inequality together with the moment condition $\mathbb{E}[\sigma^{\alpha+\varepsilon}] < \infty$ and the fact that $\min(|Z_{i0}|, |Z_{ih}|)$ is regularly varying with index 2α. This means that $\Theta_h^i = 0$ for $h \geq 1$.

Similarly, for $i \neq j$, $h \geq 1$,

$$\mathbb{P}(|X_{ih}X_{jh}| > x \mid |X_{i0}X_{j0}| > x) \leq \frac{\mathbb{P}(\max(\sigma_{i0}\sigma_{j0}, \sigma_{ih}\sigma_{jh}) \min(|Z_{i0}Z_{j0}|, |Z_{ih}Z_{jh}|) > x)}{\mathbb{P}(|X_{i0}X_{j0}| > x)} \to 0\,.$$

In the last step we again used Markov's inequality, the fact that $Z_{i0}Z_{j0}$ is regularly varying with index α (see Embrechts and Goldie [57]; cf. Lemma 4.24(1) below), hence $\min(|Z_{i0}Z_{j0}|, |Z_{ih}Z_{jh}|)$ is regularly varying with index 2α, and the moment condition $\mathbb{E}[(\sigma_{i0}\sigma_{j0})^{\alpha+\varepsilon}] < \infty$. Hence $\Theta_h^{ij} = 0$ for $i \neq j$, $h \geq 1$. □

4.3.2 Infinite variance stable limit theory for the stochastic volatility model and its product processes

Theorem 4.3. *Consider the stochastic volatility model* (4.2) *and assume the following conditions:*

1. Z *is regularly varying with index* $\alpha \in (0,4) \setminus \{2\}$.

2. $\big((\sigma_{it})_{t=1,2,\ldots}\big)_{i=1,\ldots,p}$ *is strongly mixing with rate function* (α_h) *for some* $\delta > 0$,

$$\sum_{h=0}^{\infty} \alpha_h^{\delta/(2+\delta)} < \infty\,. \tag{4.12}$$

3. *The moment condition*

$$\mathbb{E}[\sigma^{2\max(2+\delta,\alpha+\epsilon)}] < \infty \tag{4.13}$$

holds for the same $\delta > 0$ *as in* (4.12) *and some* $\epsilon > 0$.

Then

$$a_n^{-2}\big(S_1 - c_n, \dots, S_p - c_n\big) \stackrel{\rm d}{\to} (\xi_{1,\alpha/2}, \dots, \xi_{p,\alpha/2})\,, \tag{4.14}$$

where $(\xi_{i,\alpha/2})$ *are iid* $\alpha/2$*-stable random variables which are totally skewed to the right,*

$$c_n = \begin{cases} 0 & \alpha \in (0,2)\,, \\ n\,\mathbb{E}[X^2] & \alpha \in (2,4)\,, \end{cases} \tag{4.15}$$

and (a_n) *satisfies* $n\,\mathbb{P}(|X| > a_n) \to 1$ *as* $n \to \infty$.

Remark 4.4. From classical limit theory (see Feller [61], Petrov [94]) we know that (4.14) holds for an iid random field (X_{it}) with regularly varying X with index $\alpha \in (0,4)$. In the case $\alpha = 2$ one needs the special centering $c_n = n\,\mathbb{E}[X^2\mathbf{1}(|X| \le a_n)]$ which often leads to some additional technical difficulties. For this reason we typically exclude this case in the sequel.

Remark 4.5. It follows from standard theory that α-mixing of (σ_{it}) with rate function (α_h) implies α-mixing of (X_{it}) with rate function $(4\alpha_h)$; see Davis and Mikosch [43].

Proof. Recall the definition of $(\mathbf{X}_t)$ from (4.3). We will verify the conditions of Theorem 4.22 for $\mathbf{X}_t^2 = (X_{it}^2)_{i=1,\dots,p}$, $t = 0,1,2,\dots$.

(1) We start by verifying the regular variation condition for $(\mathbf{X}_t)$; see (4.11). We will determine the sequence $(\mathbf{\Theta}_h)$ corresponding to $(\mathbf{X}_t)$. We have for $t \ge 1$, with the max-norm $\|\cdot\|$,

$$\begin{aligned}
\mathbb{P}\big(\|\mathbf{X}_t\| > x \mid \|\mathbf{X}_0\| > x\big) &\le \frac{\mathbb{P}\big(\|\mathbf{X}_t\| > x\,, \cup_{i=1}^p \{|X_{i0}| > x\}\big)}{\mathbb{P}(\|\mathbf{X}_0\| > x)} \\
&\le \sum_{i=1}^p \frac{\mathbb{P}\big(\|\mathbf{X}_t\| > x\,, |X_{i0}| > x\big)}{\mathbb{P}(\|\mathbf{X}_0\| > x)} \\
&\le \sum_{i=1}^p \sum_{j=1}^p \frac{\mathbb{P}\big(|X_{jt}| > x\,, |X_{i0}| > x\big)}{\mathbb{P}(|X| > x)} \\
&\le \sum_{i=1}^p \sum_{j=1}^p \frac{\mathbb{P}\big(\max(\sigma_{jt}, \sigma_{i0}) \min(|Z_{jt}|, |Z_{i0}|) > x\big)}{\mathbb{P}(\sigma|Z| > x)}\,.
\end{aligned}$$

We observe that by Breiman's result and in view of the moment condition (4.13), for $t \ge 1$ and some positive constant c,

$$\frac{\mathbb{P}\big(\max(\sigma_{jt}, \sigma_{i0}) \min(|Z_{jt}|, |Z_{i0}|) > x\big)}{\mathbb{P}(\sigma|Z| > x)} \sim c\,\frac{\mathbb{P}(\min(|Z_{jt}|, |Z_{i0}|) > x)}{\mathbb{P}(|Z| > x)}\,,$$

and the right-hand side converges to zero as $x \to \infty$. We conclude that $\mathbf{\Theta}_h = \mathbf{0}$ for $h \ge 1$. We also have for $i \ne j$,

$$\frac{\mathbb{P}(|X_{i0}| > x\,, |X_{j0}| > x)}{\mathbb{P}(|X| > x)} \le \frac{\mathbb{P}\big(\max(\sigma_{i0}, \sigma_{j0}) \min(|Z_{i0}|, |Z_{j0}|) > x\big)}{\mathbb{P}(\sigma|Z| > x)} \to 0\,, \qquad x \to \infty\,.$$

Then, in a similar way, one can show

$$\mathbb{P}(\mathbf{X}_0/\|\mathbf{X}_0\| \in \cdot \mid \|\mathbf{X}_0\| > x) \stackrel{w}{\to} \mathbb{P}(\mathbf{\Theta}_0 \in \cdot) = \frac{1}{p}\sum_{i=1}^p \big(p_+ \varepsilon_{\mathbf{e}_i}(\cdot) + p_- \varepsilon_{-\mathbf{e}_i}(\cdot)\big)\,. \tag{4.16}$$

where $\mathbf{e}_i$ are the canonical basis vectors in $\mathbb{R}^p$, $\varepsilon_{\mathbf{x}}$ is Dirac measure at $\mathbf{x}$ and $p_\pm$ are the tail balance factors in (4.5).

We conclude that the spectral tail process $(\mathbf{\Theta}_h^{(2)})$ of $(\mathbf{X}_t^2)$ is given by $\mathbf{\Theta}_h^{(2)} = \mathbf{0}$ for $h \ge 1$ and from (4.16) we also have

$$\mathbb{P}(\mathbf{\Theta}_0^{(2)} \in \cdot) = \frac{1}{p}\sum_{i=1}^{p} \varepsilon_{\mathbf{e}_i}(\cdot)\,. \tag{4.17}$$

In particular, the condition $\sum_{i=1}^\infty \mathbb{E}[\|\mathbf{\Theta}_i^{(2)}\|] < \infty$ in Theorem 4.22(4) is trivially satisfied.

(2) Next we want to prove the mixing condition (4.46) for the sequence $(\mathbf{X}_t^2)$. We start by observing that there are integer sequences (l_n) and (m_n) $k_n\,\alpha_{l_n} \to 0$, $l_n = o(m_n)$ and $m_n = o(n)$. Then we also have for any $\gamma > 0$,

$$k_n\,\mathbb{P}\big(\sum_{t=1}^{l_n} \mathbf{X}_t^2 \mathbf{1}(\|\mathbf{X}_t\| > \varepsilon a_n) > \gamma a_n^2\big) \le k_n\, l_n\, \mathbb{P}(\|\mathbf{X}_t\| > \varepsilon a_n) \le c\, l_n/m_n = o(1)\,. \tag{4.18}$$

Relation (4.46) turns into

$$\mathbb{E}\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=1}^{n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)} - \big(\mathbb{E}\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=1}^{m_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)}\big)^{k_n} \to 0\,, \qquad \mathbf{s}\in\mathbb{R}^p\,.$$

In view of (4.18) it is not difficult to see that we can replace the sum in the former characteristic function by the sum over the index set $J_n = \{1,\ldots,m_n-l_n, m_n+1,\ldots,2m_n-l_n,\ldots,\} \subset \{1,\ldots,n\}$ and in the latter characteristic function by the sum over the index set $\{1,\ldots,m_n-l_n\}$. Without loss of generality we may assume that n/m_n is an integer. Thus it remains to show that the following difference converges to zero for every $\mathbf{s}\in\mathbb{R}^p$:

$$\begin{aligned}
&\Big|\mathbb{E}\big[\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t\in J_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)}\big] - \Big(\mathbb{E}\big[\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=1}^{m_n-l_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)}\big]\Big)^{k_n}\Big|\\
&= \Big|\sum_{v=1}^{k_n}\mathbb{E}\Big[\prod_{j=1}^{v-1}\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=(j-1)m_n+1}^{jm_n-l_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)}\\
&\qquad\times\big(\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=(v-1)m_n+1}^{vm_n-l_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)} - \mathbb{E}\big[\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=(v-1)m_n+1}^{vm_n-l_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)}\big]\big)\Big]\\
&\qquad\times\prod_{j=v+1}^{k_n}\mathbb{E}\big[\mathrm{e}^{\,i\mathbf{s}' a_n^{-2}\sum_{t=(j-1)m_n+1}^{jm_n-l_n}\mathbf{X}_t^2\mathbf{1}(\|\mathbf{X}_t\|>\varepsilon a_n)}\big]\Big|\,.
\end{aligned}$$

In view of a standard inequality for covariances of strongly mixing sequences of bounded random variables (see Doukhan [49], p. 3) the right-hand side is bounded by $c\,k_n\alpha_{l_n}$ which converges to zero by construction. Here and in what follows, c stands for any positive constant whose value is not of interest. Its value may change from line to line. This finishes the proof of the mixing condition.

(3) Next we check the anti-clustering condition (4.47) for $(\mathbf{X}_t)$ with normalization (a_n), implying the corresponding condition for $(\mathbf{X}_t^2)$ with normalization (a_n^2). By similar meth-

ods as for part (1) of the proof, assuming that $\|\cdot\|$ is the max-norm, we have

$$
\begin{aligned}
&\mathbb{P}\big(\max_{t=l,\ldots,m_n}\|\mathbf{X}_t\| > \gamma a_n \mid \|\mathbf{X}_0\| > \gamma a_n\big)\\
&\le \sum_{t=l}^{m_n}\mathbb{P}\big(\|\mathbf{X}_t\| > \gamma a_n \mid \|\mathbf{X}_0\| > \gamma a_n\big)\\
&\le c\sum_{t=l}^{m_n}\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\mathbb{P}\big(|X_{it}| > \gamma a_n\,,|X_{j0}| > \gamma a_n\big)}{\mathbb{P}(|Z| > \gamma a_n)}\\
&\le c\sum_{t=l}^{m_n}\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\mathbb{P}\big(\max(\sigma_{it},\sigma_{j0})\min(|Z_{it}|,|Z_{j0}|) > \gamma a_n\big)}{\mathbb{P}(|Z| > \gamma a_n)}\\
&\le c\sum_{t=l}^{m_n}\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\mathbb{P}\big(\sigma_{it}\min(|Z_{it}|,|Z_{j0}|) > \gamma a_n\big)}{\mathbb{P}(|Z| > \gamma a_n)}\,.
\end{aligned}
$$

By stationarity the probabilities on the right-hand side do not depend on $t \ge l$. Therefore and by Breiman's result, the right-hand side is bounded by

$$
c\,m_n\frac{\mathbb{P}\big(\min(|Z_{it}|,|Z_{j0}|) > \gamma a_n\big)}{\mathbb{P}(|Z| > \gamma a_n)} = O((m_n/n)[n\,\mathbb{P}(|Z| > a_n)]) = o(1)\,.
$$

This proves (4.47) for $(\mathbf{X}_t)$.

(4) Next we check the vanishing small values condition (4.48) for the partial sums of $(\mathbf{X}_t^2)$ and $\alpha \in (2,4)$. It is not difficult to see that it suffices to prove the corresponding result for the component processes:

$$
\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\Big(\Big|\sum_{t=1}^{n}\big(X_{it}^2\mathbf{1}(|X_{it}|\le \varepsilon a_n) - \mathbb{E}[X_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)]\big)\Big| > \gamma a_n^2\Big) = 0\,, \quad (4.19)
$$
$$
\gamma > 0\,,\; i = 1,\ldots,p\,.
$$

We have

$$
\begin{aligned}
&a_n^{-2}\sum_{t=1}^{n}\sigma_{it}^2\mathbb{E}\big[Z_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)\mid\sigma_{it}] - a_n^{-2}\,n\,\mathbb{E}[X_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)]\\
&= a_n^{-2}\sum_{t=1}^{n}(\sigma_{it}^2 - \mathbb{E}[\sigma_{it}^2])\,\mathbb{E}[Z^2] - a_n^{-2}\sum_{t=1}^{n}\big(\sigma_{it}^2\mathbb{E}[Z_{it}^2\mathbf{1}(|X_{it}|>\varepsilon a_n)\mid\sigma_{it}] - \mathbb{E}[X_{it}^2\mathbf{1}(|X_{it}|>\varepsilon a_n)]\big)\\
&= I_1 + I_2\,.
\end{aligned}
$$

The sequence (σ_{it}^2) satisfies the central limit theorem with normalization $\sqrt{n}$. This follows from Ibragimov's central limit theorem for strongly mixing sequence whose rate function (α_h) satisfies (4.12) and has moment $\mathbb{E}[\sigma^{2(2+\delta))}] < \infty$ (see (4.13)); cf. Doukhan [49], p. 45. We know that $\sqrt{n}/a_n^2 \to 0$ for $\alpha \in (2,4)$. Therefore $I_1 \stackrel{\mathbb{P}}{\to} 0$. We also have

$$
\begin{aligned}
\mathbb{E}[I_2^2] &\le \frac{n}{a_n^4}\mathbb{E}\big[\sigma^4(\mathbb{E}[Z^2\mathbf{1}(|X|>\varepsilon a_n)\mid\sigma])^2\big]\\
&\quad +2\,\frac{n}{a_n^4}\sum_{h=1}^{n}|\mathrm{cov}(\sigma_{i0}^2\mathbb{E}\big[Z_{i0}^2\mathbf{1}(|X_{i0}^2|>\varepsilon a_n)\mid\sigma_{i0}],\sigma_{ih}^2\mathbb{E}\big[Z_{ih}^2\mathbf{1}(|X_{ih}^2|>\varepsilon a_n)\mid\sigma_{ih}])|\\
&= I_3 + I_4\,.
\end{aligned}
$$

In view of the moment conditions on σ and since $\mathbb{E}[Z^2]<\infty$, $I_3\le c(n/a_n^4)\to 0$. In view of Doukhan [49], Theorem 3 on p. 9, we have

$$I_4 \quad\le\quad c\,\frac{n}{a_n^4}\sum_{h=1}^n \alpha_h^{\delta/(2+\delta)}(\mathbb{E}|\sigma|^{2(2+\delta)})^{2/(2+\delta)}\to 0\,.$$

Thus it suffices for (4.19) to prove

$$\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\Big(\Big|\sum_{t=1}^n\big(\sigma_{it}^2\mathbb{E}[Z_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)\mid\sigma_{it}]-X_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)\big)\Big|>\gamma\,a_n^2\Big)=0\,,\qquad \gamma>0\,.$$

The summands are independent and centered, conditional on the σ-field generated by $(\sigma_{it})_{t=1,\dots,n}$. An application of Čebyshev's inequality conditional on this σ-field and Karamata's theorem yield, as $n\to\infty$,

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{P}\Big(\Big|\sum_{t=1}^n\big(\sigma_{it}^2\mathbb{E}[Z_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)\mid\sigma_{it}]-X_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)\big)\Big|>\gamma\,a_n^2\big|(\sigma_{is})\Big)\Big]\\
\le\quad& c\,a_n^{-4}\mathbb{E}\Big[\sum_{t=1}^n \mathrm{var}(X_{it}^2\mathbf{1}(|X_{it}|\le\varepsilon a_n)\mid\sigma_{it})\mid(\sigma_{is})\Big]\\
\le\quad& c\,n\,\varepsilon^4\,\mathbb{E}[|X/(\varepsilon a_n)|^4\mathbf{1}(|X|\le\varepsilon a_n)]\to c\,\varepsilon^{4-\alpha}\,.
\end{aligned}$$

The right-hand side converges to zero as $\varepsilon\downarrow 0$.

This proves that all assumptions of Theorem 4.22 are satisfied. Therefore the random variables on the left-hand side of (4.14) converge to an α-stable random vector with log characteristic function

$$\begin{aligned}
&\int_0^\infty \mathbb{E}\big[\mathrm{e}^{\,i\,y\,\mathbf{t}'\sum_{j=0}^\infty\mathbf{\Theta}_j^{(2)}}-\mathrm{e}^{\,i\,y\,\mathbf{t}'\sum_{j=1}^\infty\mathbf{\Theta}_j^{(2)}}-i\,y\,\mathbf{t}'\mathbf{1}_{(1,2)}(\alpha/2)\big]\,d(-y^{\alpha/2})\\
=\quad&\sum_{j=1}^p\frac{1}{p}\int_0^\infty\mathbb{E}\big[\mathrm{e}^{\,i\,y\,t_j}-i\,y\,t_j\mathbf{1}_{(1,2)}(\alpha/2)\big]\,d(-y^{\alpha/2})\,,\qquad \mathbf{t}=(t_1,\dots,t_p)'\in\mathbb{R}^p,
\end{aligned}$$

where we used (4.17) and that $\mathbf{\Theta}_h^{(2)}=\mathbf{0}$ for $h\ge 1$. One easily checks that all summands in this expression are homogeneous functions in t_j of degree $\alpha/2$. Therefore, the limiting random vector in (4.14) has the same distribution as the sum $\sum_{j=1}^p\mathbf{e}_j\xi_{j,\alpha/2}$ for iid $\xi_{j,\alpha/2}$ which are $\alpha/2$-stable and totally skewed to the right (because all the summands in S_j are non-negative). □

4.3.3 Eigenvalues of the sample covariance matrix

We have the following approximations:

Theorem 4.6. *Assume that one of the following conditions holds:*

1. *(X_{it}) is an iid field of regularly varying random variables with index $\alpha\in(0,4)$. If $\mathbb{E}[|X|]<\infty$ we also assume $\mathbb{E}[X]=0$.*
2. *(X_{it}) is a stochastic volatility model (4.2) satisfying the regular variation, mixing and moment conditions of Theorem 4.3. If $\mathbb{E}[|Z|]<\infty$ we also assume $\mathbb{E}[Z]=0$.*

Then, with $\mathbf{X}^n$ *as in* (4.4),

$$a_n^{-2}\|\mathbf{X}^n(\mathbf{X}^n)' - \operatorname{diag}(\mathbf{X}^n(\mathbf{X}^n)')\|_2 \xrightarrow{\mathbb{P}} 0\,,$$

where $\|\cdot\|_2$ *is the spectral norm and* (a_n) *is a sequence* $n\,\mathbb{P}(|X|>a_n)\to 1$.

Proof. Part (1). Recall that for a $p\times p$ matrix $\mathbf{A}$ we have $\|\mathbf{A}\|_2\le\|\mathbf{A}\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. Hence

$$\begin{aligned} a_n^{-4}\|\mathbf{X}^n(\mathbf{X}^n)' - \operatorname{diag}(\mathbf{X}^n(\mathbf{X}^n)')\|_2^2 &\le a_n^{-4}\|\mathbf{X}^n(\mathbf{X}^n)' - \operatorname{diag}(\mathbf{X}^n(\mathbf{X}^n)')\|_F^2 \\ &= \sum_{1\le i\ne j\le p}\big(a_n^{-2}S_{ij}\big)^2\,. \end{aligned} \tag{4.20}$$

In view of the assumptions, $(X_{it}\,X_{jt})_{t=1,2,\ldots}$, $i\ne j$, is an iid sequence of regularly varying random variables with index α which is also centered if $\mathbb{E}[|X|]<\infty$. We consider two different cases.

The case $\alpha\in(0,2)$. According to classical limit theory (see Feller [61], Petrov [94]) we have for $i\ne j$, $b_n^{-1}S_{ij}\xrightarrow{\mathrm{d}}\xi_\alpha$, (see (4.4) for the definition of S_{ij}) where ξ_α is an α-stable random variable and (b_n) is chosen $n\,\mathbb{P}(|X_1X_2|>b_n)\to 1$ for independent copies X_1,X_2 of X. Since (b_n) and (a_n^2) are regularly varying with indices $1/\alpha$ and $2/\alpha$, respectively, the right-hand side in (4.20) converges to zero in probability.

The case $\alpha\in[2,4)$. In this case the distribution of X_1X_2 is in the domain of attraction of the normal law. Since X_1X_2 has mean zero we can apply classical limit theory (see Feller [61], Petrov [94]) to conclude that $b_n^{-1}S_{ij}\xrightarrow{\mathrm{d}} N$, where (b_n) is regularly varying with index $1/2$ and N is centered Gaussian. Since $b_n/a_n^2\to 0$ we again conclude that the right-hand side of (4.20) converges to zero in probability.

Part (2). We again appeal to (4.20). Let $\gamma<\min(2,\alpha)$. Then we have for $i\ne j$, using the independence of $(X_{it}X_{jt})$ conditional on $((\sigma_{it},\sigma_{jt}))$ and that the distribution of Z is centered if its first absolute moments exists, that

$$a_n^{-2\gamma}\mathbb{E}\Big[\big|S_{ij}\big|^\gamma\mid((\sigma_{it},\sigma_{jt}))\Big] \;\le\; c\,\frac{n}{a_n^{2\gamma}}\,\frac{1}{n}\sum_{t=1}^n(\sigma_{it}\sigma_{jt})^\gamma(\mathbb{E}|Z|^\gamma)^2\,,$$

cf. von Bahr and Esséen [115] and Petrov [94], 2.6.20 on p. 82. In view of the moment condition (4.13) we have $\mathbb{E}[(\sigma_i\sigma_j)^\gamma]<\infty$ and $n/a_n^{2\gamma}\to 0$ if we choose γ sufficiently close to $\min(2,\alpha)$. Then the right-hand side converges to zero in view of the ergodic theorem. An application of the conditional Markov inequality of order γ yields $a_n^{-2}S_{ij}\xrightarrow{\mathbb{P}}0$. This proves the theorem. □

Corollary 4.7. *Assume that* (X_{it}) *is either*

1. *an iid field of regularly varying random variables with index* $\alpha\in(0,4)$ *and* $\mathbb{E}[X]=0$ *if* $\mathbb{E}[|X|]<\infty$*, or*
2. *a stochastic volatility model of regularly varying random variables with index* $\alpha\in(0,4)\setminus\{2\}$ *satisfying the conditions of Theorem 4.6(2).*

Then

$$a_n^{-2}\max_{i=1,\ldots,p}\big|\lambda_{(i)}-S_{(i)}\big|\xrightarrow{\mathbb{P}}0\,,$$

where (λ_i) are the eigenvalues of $\mathbf{X}^n(\mathbf{X}^n)'$, $\lambda_{(1)} \ge \cdots \ge \lambda_{(p)}$ are their ordered values and $S_{(1)} \ge \cdots \ge S_{(p)}$ are the ordered values of $S_1, \dots, S_p$ defined in (4.4). In particular, we have

$$a_n^{-2}\big(\lambda_{(1)} - c_n, \dots, \lambda_{(p)} - c_n\big) \xrightarrow{\mathrm{d}} \big(\xi_{(1),\alpha/2}, \dots, \xi_{(p),\alpha/2}\big)\,, \tag{4.21}$$

where (c_n) is defined in (4.15) for $\alpha \neq 2$ and in Remark 4.4 for $\alpha = 2$, $(\xi_{i,\alpha/2})$ are iid $\alpha/2$-stable random variables given in Theorem 4.3 for the stochastic volatility model and in Remark 4.4 for the iid field, and $\xi_{(1),\alpha/2} \ge \cdots \ge \xi_{(p),\alpha/2}$ are their ordered values.

Proof. We have by Weyl's inequality (see Bhatia [20]) and Theorem 4.6,

$$a_n^{-2} \max_{i=1,\dots,p} \big|\lambda_{(i)} - S_{(i)}\big| \quad \le \quad a_n^{-2} \|\mathbf{X}^n(\mathbf{X}^n)' - \operatorname{diag}(\mathbf{X}^n(\mathbf{X}^n)')\|_2 \xrightarrow{\mathbb{P}} 0\,. \tag{4.22}$$

If (X_{it}) is an iid random field (see Remark 4.4) or a stochastic volatility model satisfying the conditions of Theorem 4.6(2) we have (4.14). Then (4.22) implies (4.21). □

Remark 4.8. If $\alpha \in (2,4)$ we have $\mathbb{E}[X^2] < \infty$. Therefore (4.21) reads as

$$\frac{n}{a_n^2}\,\Big(\frac{\lambda_{(i)}}{n} - \mathbb{E}[X^2]\Big)_{i=1,\dots,p} \xrightarrow{\mathrm{d}} (\xi_{(i),\alpha/2})_{i=1,\dots,p}\,. \tag{4.23}$$

We notice that $n/a_n^2 \to \infty$ for $\alpha \in (2,4)$ since (n/a_n^2) is regularly varying with index $1 - 2/\alpha$. In particular, if $\operatorname{tr}(\mathbf{X}^n(\mathbf{X}^n)')$ denotes the trace of $\mathbf{X}^n(\mathbf{X}^n)'$ we have for $i \le p$,

$$\frac{\lambda_{(i)}}{\operatorname{tr}(\mathbf{X}^n(\mathbf{X}^n)')} \quad = \quad \frac{\lambda_{(i)}/n}{(\lambda_1 + \cdots + \lambda_p)/n} \xrightarrow{\mathbb{P}} \frac{1}{p}\,. \tag{4.24}$$

The joint asymptotic distribution of the ordered eigenvalues $(\lambda_{(i)})$ is easily calculated from the distribution of a totally skewed $\alpha/2$-stable random variable $\xi_{1,\alpha/2}$; in particular, the limit of $(a_n^{-2}(\lambda_{(1)} - c_n))$ has the distribution of $\max(\xi_{1,\alpha/2}, \dots, \xi_{p,\alpha/2})$.

For applications, it is more natural to replace the random variables X_{it} by their mean-centered versions $X_{it} - \overline{X}_i$, where $\overline{X}_i = (1/n)\sum_{t=1}^n X_{it}$, instead of assuming that they have mean zero. The previous results remain valid for the sample-mean centered random variables X_{it}, also in the case when X has infinite first moment.

4.3.4 Some applications: Limit results for ordered eigenvalues and eigenvectors of the sample covariance matrix

In what follows, we assume the conditions of Corollary 4.7.

4.3.4.1 Spacings

Using the joint convergence of the normalized ordered eigenvalues $(\lambda_{(i)})$ we can calculate the limit of the spectral gaps:

$$\Big(\frac{\lambda_{(i)} - \lambda_{(i+1)}}{a_n^2}\Big)_{i=1,\dots,p-1} \xrightarrow{\mathrm{d}} \big(\xi_{(i),\alpha/2} - \xi_{(i+1),\alpha/2}\big)_{i=1,\dots,p-1}\,. \tag{4.25}$$

We notice that the ordered values $\xi_{(i),\alpha/2}$ and linear functionals thereof (such as $\xi_{(i),\alpha/2} - \xi_{(i+1),\alpha/2}$) are again jointly regularly varying with index $\alpha/2$. This is due to the continuous mapping theorem for regularly varying vectors; see Hult and Lindskog [74, 75], cf. Jessen and Mikosch [81].

4.3.4.2 Trace

For the trace of $\mathbf{X}^n(\mathbf{X}^n)'$ we have

$$\begin{aligned} a_n^{-2}\big(\mathrm{tr}(\mathbf{X}^n(\mathbf{X}^n)') - p\,c_n\big) &= a_n^{-2}\sum_{i=1}^p (S_i - c_n) \\ &= a_n^{-2}\sum_{i=1}^p (\lambda_i - c_n) \stackrel{\mathrm{d}}{\to} \xi_{1,\alpha/2} + \cdots + \xi_{p,\alpha/2} \stackrel{d}{=} p^{2/\alpha}\xi_{1,\alpha/2}\,. \end{aligned}$$

Moreover, we have the joint convergence of the normalized and centered $(\lambda_{(i)})$ and $\mathrm{tr}(\mathbf{X}^n(\mathbf{X}^n)') = \lambda_1 + \cdots + \lambda_p$. In particular, we have the self-normalized limit relations

$$\Big(\frac{\lambda_{(i)} - c_n}{\mathrm{tr}(\mathbf{X}^n(\mathbf{X}^n)') - p\,c_n}\Big)_{i=1,\ldots,p} \quad \stackrel{\mathrm{d}}{\to} \quad \Big(\frac{\xi_{(i),\alpha/2}}{\xi_{1,\alpha/2} + \cdots + \xi_{p,\alpha/2}}\Big)_{i=1,\ldots,p}\,,$$

and for $\alpha \in (2,4)$, by the strong law of large numbers,

$$\frac{np}{a_n^2}\Big(\frac{\lambda_{(i)} - c_n}{\mathrm{tr}(\mathbf{X}^n(\mathbf{X}^n)')}\Big)_{i=1,\ldots,p} \quad \stackrel{\mathrm{d}}{\to} \quad \frac{\xi_{(i),\alpha/2}}{\mathbb{E}[X^2]}\,.$$

4.3.4.3 Determinant

Since $\lambda_i - c_n$ are the eigenvalues of $\mathbf{X}^n(\mathbf{X}^n)' - c_n\mathbf{I}_p$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix, we obtain for the determinant

$$\begin{aligned} \det\big(a_n^{-2}(\mathbf{X}^n(\mathbf{X}^n)' - c_n\,\mathbf{I}_p)\big) &= \prod_{i=1}^p a_n^{-2}(\lambda_{(i)} - c_n) \\ &\stackrel{\mathrm{d}}{\to} \xi_{(1),\alpha/2}\cdots\xi_{(p),\alpha/2} = \xi_{1,\alpha/2}\cdots\xi_{p,\alpha/2}\,. \end{aligned}$$

For $\alpha \in (2,4)$, we also have

$$\begin{aligned} \frac{1}{a_n^2 c_n^{p-1}}\big(\det(\mathbf{X}^n(\mathbf{X}^n)') - c_n^p\big) &= \sum_{i=1}^p a_n^{-2}\big(\lambda_{(i)} - c_n\big)\prod_{j=1}^{i-1}\frac{\lambda_{(j)}}{c_n} \\ &\stackrel{\mathrm{d}}{\to} \sum_{i=1}^p \xi_{(i),\alpha/2} = \sum_{i=1}^p \xi_{i,\alpha/2} \stackrel{d}{=} p^{2/\alpha}\,\xi_{1,\alpha/2}\,, \end{aligned}$$

where we used (4.23).

4.3.4.4 Eigenvectors

It is also possible to localize the eigenvectors of the matrix $a_n^{-2}\mathbf{X}^n(\mathbf{X}^n)'$. Since this matrix is approximated by its diagonal in spectral norm, one may expect that the unit eigenvectors of the original matrix are close to the canonical basis vectors. We can write

$$a_n^{-2}\mathbf{X}^n(\mathbf{X}^n)'\mathbf{e}_{L_j} = a_n^{-2}S_{(j)}\,\mathbf{e}_{L_j} + \varepsilon_n\,\mathbf{W}\,,$$

where $\mathbf{W}$ is a unit vector orthogonal to $\mathbf{e}_{L_j}$, L_j is the index of $S_{(j)} = S_{L_j}$ and

$$\varepsilon_n = a_n^{-2}\|\big(\mathbf{X}^n(\mathbf{X}^n)' - S_{(j)}\big)\mathbf{e}_{L_j}\|_{\ell_2} \stackrel{\mathbb{P}}{\to} 0\,,$$

from Theorem 4.6 and by equivalence of all matrix norms. According to Proposition A.1 in Benaych-Georges and Peché [19], there is an eigenvalue $a_n^{-2}\lambda_{(j)}$ of $a_n^{-2}\mathbf{X}^n(\mathbf{X}^n)'$ in some ε_n-neighborhood of $a_n^{-2}S_{(j)}$. Define

$$\Omega_n = \{a_n^{-2}|\lambda_{(j)} - \lambda_{(l)}| > d_n \,, l \neq j\}\,,$$

for $d_n = k\varepsilon_n$ for any fixed $k > 1$. Then $\lim_{n\to\infty} \mathbb{P}(\Omega_n) = 1$ because of (4.25) and $d_n \stackrel{\mathbb{P}}{\to} 0$. Hence, for large n, $a_n^{-2}\lambda_{(j)}$ and $a_n^{-2}\lambda_{(l)}$ have distance at least d_n with high probability. Another application of Proposition A.1 in [19] yields that the unit eigenvector $\mathbf{V}$ associated with $a_n^{-2}\lambda_{(j)}$ satisfies the relation

$$\begin{aligned}
& \limsup_{n\to\infty} \mathbb{P}\big(\|\mathbf{V} - V_{L_j}\mathbf{e}_{L_j}\|_{\ell_2} > \delta\big) \\
\leq\; & \limsup_{n\to\infty} \mathbb{P}\big(\{\|\mathbf{V} - V_{L_j}\mathbf{e}_{L_j}\|_{\ell_2} > \delta\} \cap \Omega_n\big) + \limsup_{n\to\infty} \mathbb{P}(\Omega_n^c) \\
\leq\; & \limsup_{n\to\infty} \mathbb{P}\big(\{2\,\varepsilon_n/(d_n - \varepsilon_n) > \delta\} \cap \Omega_n\big) \\
=\; & \mathbf{1}_{\{2/(k-1)>\delta\}}\,.
\end{aligned}$$

For any fixed $\delta > 0$, the right-hand side is zero for sufficiently large k. Since both $\mathbf{V}$ and $\mathbf{e}_{L_j}$ are unit eigenvectors this means that $\|\mathbf{V} - \mathbf{e}_{L_j}\|_{\ell_2} \stackrel{\mathbb{P}}{\to} 0$.

4.3.4.5 Sample correlation matrix

In Remark 4.8 we mentioned that we can replace the variables X_{it} by their sample-mean centered versions $X_{it} - \overline{X}_i$ without changing the asymptotic theory. Similarly, one may be interested in transforming the X_{it} as follows:

$$\widetilde{X}_{it} = \frac{X_{it} - \overline{X}_i}{\widehat{\sigma}_i}\,, \qquad \widehat{\sigma}_i^2 = \sum_{t=1}^n (X_{it} - \overline{X}_i)^2\,.$$

Then the matrix

$$\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)' = \Big(\sum_{t=1}^n \widetilde{X}_{it}\widetilde{X}_{jt}\Big)_{i,j=1,\ldots,p}\,,$$

is the sample correlation matrix. We write $\widetilde{\lambda}_i$, $i = 1, \ldots, p$, for the eigenvalues of $\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)'$ and $\widetilde{\lambda}_{(1)} \geq \cdots \geq \widetilde{\lambda}_{(p)}$ for their ordered values.

We notice that the entries of this matrix are all bounded in modulus by one. In particular, the diagonal consists of ones. We do not have a complete limit theory for the eigenvalues $\widetilde{\lambda}_i$. We restrict ourselves to iid (X_{it}) to explain the differences.

Lemma 4.9. *Assume that (X_{it}) is an iid field of random variables.*

1. *If $\mathbb{E}[X^2] < \infty$ then*

$$\sqrt{n} \max_{i=1,\ldots,p} |\widetilde{\lambda}_i - 1| = O_{\mathbb{P}}(1)\,.$$

2. *If X is regularly varying with index $\alpha \in (0, 2)$ then*

$$\frac{a_n^2}{b_n} \max_{i=1,\ldots,p} |\widetilde{\lambda}_i - 1| = O_{\mathbb{P}}(1)\,,$$

where (a_n) and (b_n) are chosen $\mathbb{P}(|X| > a_n) \sim \mathbb{P}(|X_1 X_2| > b_n) \sim n^{-1}$ for iid copies X_1, X_2 of X.

Remark 4.10. Notice that the lemma implies $\widetilde{\lambda}_i \xrightarrow{\mathbb{P}} 1$ for $i = 1, \ldots, p$, and the analog of relation (4.24) remains valid.

Proof. Part(1) We assume without loss of generality that $1 = \mathbb{E}[X^2]$. Then by classical limit theory,

$$\begin{aligned}\sqrt{n}\big(\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)' - \operatorname{diag}(\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)')\big) &= \sqrt{n}\big(\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)' - \mathbf{I}_p\big) \\ &= \Big(\mathbf{1}(i \neq j)\frac{n^{-1/2}\sum_{t=1}^n (X_{it} - \overline{X}_i)(X_{jt} - \overline{X}_j)}{(\widehat{\sigma}_i/\sqrt{n})(\widehat{\sigma}_j/\sqrt{n})}\Big) \\ &\xrightarrow{\mathrm{d}} \big(N_{ij}\mathbf{1}(i \neq j)\big)\,,\end{aligned}$$

where N_{ij}, $1 \le i < j \le n$, are iid $N(0,1)$ and $N_{ij} = N_{ji}$. By Weyl's inequality,

$$\sqrt{n} \max_{i=1,\ldots,p} \Big|\widetilde{\lambda}_{(i)} - 1\Big| \le \sqrt{n}\|\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)' - \mathbf{I}_p\|_2 = O_{\mathbb{P}}(1)\,.$$

Part(2) If X is regularly varying with index $\alpha \in (0,2)$, we have that $(a_n^{-2}\widehat{\sigma}_i^2)$ converges to a vector of iid positive $\alpha/2$-stable random variables (ξ_i), while for every $i \neq j$, $b_n^{-1}\sum_{t=1}^n (X_{it} - \overline{X}_i)(X_{jt} - \overline{X}_j) \xrightarrow{\mathrm{d}} \xi_{ij}$ and the limit ξ_{ij} is α-stable. Then by Weyl's inequality

$$\frac{a_n^2}{b_n} \max_{i=1,\ldots,p} \Big|\widetilde{\lambda}_{(i)} - 1\Big| \le \frac{a_n^2}{b_n}\|\widetilde{\mathbf{X}}^n(\widetilde{\mathbf{X}}^n)' - \mathbf{I}_p\|_2 = O_{\mathbb{P}}(1)\,.$$

□

4.4 Case (2): σ dominates the tail

In this section we assume the conditions of Case (2); see Section 4.2.2. Our goal is to derive results analogous to Case (1): regular variation of (X_{it}), infinite variance limits for S_{ij} and limit theory for the eigenvalues of the corresponding sample covariance matrices. It turns out that this case offers a wider spectrum of possible limit behaviors and that we have to further distinguish our assumptions about the distribution of η. So, in addition to (4.9) we assume that either

$$\mathbb{E}[\mathrm{e}^{\eta\alpha}] = \infty \tag{4.26}$$

or

$$\lim_{x\to\infty} \frac{\mathbb{P}(\eta_1 + \eta_2 > x)}{\mathbb{P}(\eta_1 > x)} = c \in (0,\infty) \quad \Leftrightarrow \quad \lim_{x\to\infty} \frac{\mathbb{P}(\mathrm{e}^{\eta_1} \cdot \mathrm{e}^{\eta_2} > x)}{\mathbb{P}(\mathrm{e}^{\eta_1} > x)} = c \in (0,\infty) \tag{4.27}$$

hold, where η_1 and η_2 are independent copies of η.

Remark 4.11. Following Cline [34], we call the distribution of a random variable η *convolution equivalent* if e^{η} is regularly varying and relation (4.27) holds. The assumptions (4.26) and (4.27) are mutually exclusive, since the only possible finite limit c in (4.27) is given by $c = 2\mathbb{E}[\mathrm{e}^{\eta\alpha}]$; see Davis and Resnick [47]. There are, however, regularly varying distributions of e^{η} which satisfy $\mathbb{E}[\mathrm{e}^{\eta\alpha}] < \infty$ but not (4.27). An example is given in Cline [34], p. 538; see also Lemma 4.24(3) for a necessary and sufficient condition ensuring (4.27).

As we will see later, relations (4.26) and (4.27) cause rather distinct limit behavior of the sample covariance matrix. In particular, (4.27) allows for non-vanishing off-diagonal elements of the normalized sample covariance matrices, in contrast to Case (1).

For notational simplicity, define

$$\psi = \max_{k,l} \psi_{kl} \qquad \text{and} \qquad \Lambda = \{(k,l) : \psi_{kl} = \psi\}\,.$$

Recall that for convenience we assume that $\psi = 1$; if the latter condition does not hold we can replace (without loss of generality) the random variables η_{kl} by $\psi\eta_{kl}$ and the coefficients ψ_{kl} by ψ_{kl}/ψ. For given (i,j), we define

$$\psi^{ij} = \max_{k,l} \left(\psi_{kl} + \psi_{k+i-j,l}\right). \tag{4.28}$$

Notice that $1 \le \psi^{ij} \le 2$. For $d \ge 1$, we write $\mathbf{i} = (i_1,\dots,i_d), \mathbf{j} = (j_1,\dots,j_d)$ for elements of $\mathbb{Z}^d$. For given $\mathbf{i}$ and $\mathbf{j}$ we also define

$$\psi^{\mathbf{i},\mathbf{j}} = \max_{1\le l\le d} \psi^{i_l,j_l}.$$

4.4.1 Regular variation

We start by showing that the volatility sequences are regularly varying.

Proposition 4.12. *Under the aforementioned conditions and conventions (including that either (4.26) or (4.27) hold),*

1. *each of the sequences $(\sigma_{it})_{t\in\mathbb{Z}}$, $i = 1,2,\dots$, is regularly varying with index α,*
2. *each of the sequences $(\sigma_{it}\sigma_{jt})_{t\in\mathbb{Z}}$, $i,j = 1,2,\dots$, is regularly varying with corresponding index α/ψ^{ij},*
3. *For $d \ge 1$ and $\mathbf{i},\mathbf{j} \in \mathbb{Z}$, the d-variate sequence $((\sigma_{i_k,t}\sigma_{j_k,t})_{1\le k\le d})_{t\in\mathbb{Z}}$ is regularly varying with index $\alpha/\psi^{\mathbf{i},\mathbf{j}}$.*

Remark 4.13. Part (3) of the proposition possibly includes degenerate cases in the sense that for some choices of (i_k,j_k), $(\sigma_{i_k,t}\sigma_{j_k,t})$ is regularly varying with index $\alpha/\psi^{i_k,j_k} > \alpha/\psi^{\mathbf{i},\mathbf{j}}$.

Part (3) implies (2) in the case $d = 1$. Part (2) implies (1) by setting $i = j$ and observing that, by non-negativity of σ, regular variation of (σ_{it}^2) with index $\alpha/2$ is equivalent to regular variation of (σ_{it}) with index α.

Proof. To give some intuition we start with the proof of the marginal regular variation of σ, although it is just a special case of (1). We have

$$\sigma_{it} = \mathrm{e}^{\sum_{(k,l)\in\Lambda} \eta_{i-k,t-l}}\, \mathrm{e}^{\sum_{(k,l)\notin\Lambda} \psi_{kl}\eta_{i-k,t-l}} =: \sigma_{it,\Lambda}\sigma_{it,\Lambda^c}. \tag{4.29}$$

We first verify that $\sigma = \sigma_\Lambda\sigma_{\Lambda^c}$ is regularly varying with index α. Since $|\Lambda| < \infty$ by our assumptions, and in view of Embrechts and Goldie [57], Corollary on p. 245, cf. also Lemma 4.24(1) below, the product σ_Λ is regularly varying with index α. The random variable σ_{Λ^c} is independent of σ_Λ. Similarly to Mikosch and Rezapour [89] (see also the end of this proof for a similar argumentation) one can show that σ_{Λ^c} has moment of order $\alpha+\varepsilon$ for sufficiently small positive ε. Therefore, by Breiman's lemma [27],

$$\mathbb{P}(\sigma > x) \sim \mathbb{E}[\sigma_{\Lambda^c}^\alpha]\, \mathbb{P}(\sigma_\Lambda > x)\,, \qquad x \to \infty\,.$$

This proves regular variation with index α of the marginal distributions of (σ_{it}).

In the remainder of the proof we focus on (3). For a given choice of $\mathbf{i}, \mathbf{j}, \mathbf{t} \in \mathbb{Z}^d$, we write

$$\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}} = \{(m,n) : \psi_{i_l-m,t_l-n} + \psi_{j_l-m,t_l-n} = \psi^{\mathbf{i},\mathbf{j}} \text{ for some } 1 \le l \le d\}. \tag{4.30}$$

We will show that the random vector $(\sigma_{i_1,t_1}\sigma_{j_1,t_1}, \ldots, \sigma_{i_d,t_d}\sigma_{j_d,t_d}) =: \boldsymbol{\sigma}'$ is regularly varying with index $\alpha/\psi^{\mathbf{i},\mathbf{j}}$ which proves (3). Note that

$$\begin{aligned} \sigma_{i,t}\sigma_{j,t} &= \prod_{(k,l)} \exp(\psi_{kl}\eta_{i-k,t-l}) \prod_{(k',l')} \exp(\psi_{k'l'}\eta_{j-k',t-l'}) \\ &= \prod_{(m,n)} \exp((\psi_{i-m,t-n} + \psi_{j-m,t-n})\eta_{m,n}) \end{aligned}$$

and write

$$\boldsymbol{\sigma} = \mathbf{A}\,\mathbf{Z} \tag{4.31}$$

where

$$\begin{aligned} \mathbf{A} &= \operatorname{diag}\left(\begin{pmatrix} \prod_{(m,n)\in\Lambda^c_{\mathbf{i},\mathbf{j},\mathbf{t}}} \mathrm{e}^{\,\eta_{m,n}(\psi_{i_1-m,t_1-n}+\psi_{j_1-m,t_1-n})} \\ \vdots \\ \prod_{(m,n)\in\Lambda^c_{\mathbf{i},\mathbf{j},\mathbf{t}}} \mathrm{e}^{\,\eta_{m,n}(\psi_{i_d-m,t_d-n}+\psi_{j_d-m,t_d-n})} \end{pmatrix}'\right), \\ \mathbf{Z} &= \begin{pmatrix} \prod_{(m,n)\in\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}}} \mathrm{e}^{\,\eta_{m,n}(\psi_{i_1-m,t_1-n}+\psi_{j_1-m,t_1-n})} \\ \vdots \\ \prod_{(m,n)\in\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}}} \mathrm{e}^{\,\eta_{m,n}(\psi_{i_d-m,t_d-n}+\psi_{j_d-m,t_d-n})} \end{pmatrix}. \end{aligned}$$

where $\operatorname{diag}((a_1,\ldots,a_k))$ is any diagonal matrix with diagonal elements $a_1,\ldots,a_k$. We notice that $\mathbf{A}$ and $\mathbf{Z}$ are independent.

Consider iid copies (Y_j) of $\mathrm{e}^{\,\eta}$. There exist suitable numbers $(a_{ij})_{1\le i\le d, 1\le j\le p}$ with $p = |\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}}|$ the components of $\mathbf{Z}$ have representation in distribution $\prod_{j=1}^p Y_j^{a_{ij}}$, $1 \le i \le d$. By assumption, Y_j is regularly varying with index α and satisfies either assumption (4.52) or $\mathbb{E}[Y_j^\alpha] = \infty$. Furthermore, for each j there exists one $1 \le i \le d$ such that $a_{ij} = a_{\max} = \psi^{\mathbf{i},\mathbf{j}}$ by the definition of $\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}}$. An application of Proposition 4.26 shows that $\mathbf{Z}$ is regularly varying with index $\alpha/\psi^{\mathbf{i},\mathbf{j}}$ and limit measure $\mu_{\mathbf{Z}}$ which is given as μ in Proposition 4.26 (ii) (if (4.26) holds) or Proposition 4.26 (i) (if (4.27) holds). Now, choose $\epsilon, \delta > 0$ such that

$$\frac{\psi_{i_l-m,t_l-n} + \psi_{j_l-m,t_l-n}}{\psi^{\mathbf{i},\mathbf{j}}}(1+\delta) < 1-\epsilon, \quad (m,n) \in \Lambda^c_{\mathbf{i},\mathbf{j},\mathbf{t}}, \ \ 1 \le l \le d,$$

which is possible by the definition of $\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}}$ and the summability constraint on the coefficients. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\|\mathbf{A}\|_{\mathrm{op}}^{\alpha(1+\delta)/\psi^{\mathbf{i},\mathbf{j}}}\right]\\
\leq &\sum_{l=1}^{d}\prod_{(m,n)\in\Lambda^c_{\mathbf{i},\mathbf{j},\mathbf{t}}}\mathbb{E}\left[\mathrm{e}^{\eta_{m,n}\alpha(1+\delta)(\psi_{i_l-m,t_l-n}+\psi_{j_l-m,t_l-n})/\psi^{\mathbf{i},\mathbf{j}}}\right]\\
\leq &\sum_{l=1}^{d}\prod_{(m,n)\in\Lambda^c_{\mathbf{i},\mathbf{j},\mathbf{t}}}\mathbb{E}\left[\mathrm{e}^{\eta_{m,n}\alpha(1-\epsilon)}\right]^{(1+\delta)(\psi_{i_l-m,t_l-n}+\psi_{j_l-m,t_l-n})/((1-\epsilon)\psi^{\mathbf{i},\mathbf{j}})}<\infty,
\end{aligned}$$

where we used Jensen's inequality for the penultimate step and the summability condition of the coefficients for the final one. Thus we have verified all conditions of the multivariate Breiman lemma in Basrak et al. [13], implying that $\boldsymbol{\sigma}$ inherits regular variation from $\mathbf{Z}$ with corresponding index $\alpha/\psi^{\mathbf{i},\mathbf{j}}$ and limit measure $\mu_{\boldsymbol{\sigma}}(\cdot)=\mathbb{E}[\mu_{\mathbf{Z}}(\mathbf{A}^{-1}\cdot)]$. □

Proposition 4.14. *Assume that the aforementioned conditions (including either (4.26) or (4.27)) hold and that in addition $\mathbb{E}[|Z|^{\alpha+\delta}]<\infty$ for some $\delta>0$. Then the following statements hold:*

1. *Each of the sequences $(X_{it})_{t\in\mathbb{Z}}$, $i\in\mathbb{Z}$, is regularly varying with index α.*
   If (4.26) holds then the corresponding spectral tail process satisfies $\Theta_t^i=0$ a.s., $t\geq 1$, and $\mathbb{P}(\Theta_0^i=\pm1)=\mathbb{E}[Z_\pm^\alpha]/\mathbb{E}[|Z|^\alpha]$.
   If (4.27) holds, then for any Borel set $B=B_0\times\cdots\times B_n\subset\mathbb{R}^{n+1}$,

$$\begin{aligned}
&\mathbb{P}((\Theta_t^i)_{t=0,\dots,n}\in B) \qquad (4.32)\\
&=\sum_{(u,v)\in\Lambda_i^{(0)}}\frac{1}{|\Lambda_i^{(0)}|}\frac{\mathbb{E}\left[\mathbf{1}\left(\left(\mathbf{1}((u,v)\in\Lambda_i^{(t)})\frac{X_{it}}{|X_{i0}|}\right)_{t=0,\dots,n}\in B\right)|X_{i0}|^\alpha\right]}{\mathbb{E}[|X_{i0}|^\alpha]},
\end{aligned}$$

   where $\Lambda_i^{(t)}=\{(u,v):\psi_{i-u,t-v}=1\}, t=0,\dots,n$.

2. *Each of the sequences $(X_{it}X_{jt})_{t\in\mathbb{Z}}$, $i,j\in\mathbb{Z}$, is regularly varying with index α/ψ^{ij}.*
   If (4.26) holds then the corresponding spectral tail process satisfies $\Theta_t^{ij}=0$ a.s., $t\geq1$, and $\mathbb{P}(\Theta_0^{ij}=\pm1)=\mathbb{E}[(Z_iZ_j)_\pm^{\alpha/\psi^{ij}}]/\mathbb{E}[|Z_iZ_j|^{\alpha/\psi^{ij}}]$.
   If (4.27) holds, then for any Borel set $B=B_0\times\cdots\times B_n\subset\mathbb{R}^{n+1}$,

$$\begin{aligned}
&\mathbb{P}((\Theta_t^{ij})_{t=0,\dots,n}\in B) \qquad (4.33)\\
&=\sum_{(u,v)\in\Lambda_{i,j}^{(0)}}\frac{1}{|\Lambda_{i,j}^{(0)}|}\frac{\mathbb{E}\left[\mathbf{1}\left(\left(\mathbf{1}((u,v)\in\Lambda_{i,j}^{(t)})\frac{X_{it}X_{jt}}{|X_{i0}X_{j0}|}\right)_{t=0,\dots,n}\in B\right)|X_{i0}X_{j0}|^{\alpha/\psi^{ij}}\right]}{\mathbb{E}[|X_{i0}X_{j0}|^{\alpha/\psi^{ij}}]},
\end{aligned}$$

   where $\Lambda_{i,j}^{(t)}=\{(u,v):\psi_{i-u,t-v}+\psi_{j-u,t-v}=\psi^{ij}\}, t=0,\dots,n$.

3. *For $d\geq1$ and $\mathbf{i},\mathbf{j}\in\mathbb{Z}^d$, the d-variate sequence $((X_{i_kt}X_{j_kt})_{1\leq k\leq d})_{t\in\mathbb{Z}}$ is jointly regularly varying with index $\alpha/\psi^{\mathbf{i},\mathbf{j}}$.*

Remark 4.15. 1. Equation (4.32) shows that in this case the distribution of $(\Theta_t^i)_{t\geq 0}$ is a mixture of $|\Lambda_i^{(0)}|$ distributions, where each distribution gets the weight $1/|\Lambda_i^{(0)}|$. Heuristically speaking, a distribution in this mixture that corresponds to a specific $(u,v) \in \Lambda_i^{(0)}$ has interpretation as the distribution of $(X_{it}/|X_{i0}|)_{t\geq 0}$, given that we have seen an extreme observation of $|X_{i0}|$ caused by an extreme realization of $\mathrm{e}^{\eta_{u,v}}$. The variables $\mathrm{e}^{\eta_{u,v}}$, $(u,v) \in \Lambda_i^{(0)}$, are those which have a maximum exponent (equal to 1) in the product $\prod_{(u,v)} \exp(\psi_{i-u,-v}\eta_{u,v}) = \sigma_{i0}$. They are therefore the factors which are most likely to make σ_{i0}, hence X_{i0}, extreme.

   An analogous interpretation can be derived from (4.33) for the distribution of $(\Theta_t^{ij})_{t\geq 0}$.

2. Note that for fixed i, j, the inner indicator functions in (4.32) and (4.33) are positive only for finitely many t. Hence there are only finitely many $t \geq 1$ such that $\mathbb{P}(\Theta_t^i \neq 0) > 0$ and $\mathbb{P}(\Theta_t^{(ij)} \neq 0) > 0$.

3. Using similar techniques as in the proof of cases (1) and (2) below, one can also give an explicit expression for the resulting d-dimensional spectral tail process of $((X_{i_k t} X_{j_k t})_{1\leq k\leq d})_{t\in\mathbb{Z}}$ in (3). However, due to its complexity, we refrain from stating it here.

Proof. We start by showing that all mentioned sequences are regularly varying. Exemplarily, we show this for case (2). Very similar arguments can be used for the two other cases. For $n \geq 0$ write

$$(X_{it}X_{jt})'_{t=0,\dots,n} = \operatorname{diag}\left((Z_{it}Z_{jt})'_{t=0,\dots,n}\right) \cdot (\sigma_{it}\sigma_{jt})'_{t=0,\dots,n} .$$

Since $\psi^{ij} \geq 1$ our moment assumption on Z implies that $\mathbb{E}[|Z|^{\alpha/\psi^{ij}+\delta}] < \infty$ for some $\delta > 0$. Then Proposition 4.12 allows us to apply the aforementioned multivariate Breiman lemma, yielding the regular variation of the vector $(X_{it}X_{jt})_{t=0,\dots,n}$ with index α/ψ^{ij}. From the first definition given in Section 4.2.3, this implies the regular variation of the sequence.

As for the derivation of the explicit form of the spectral tail process in (1) and (2), we restrict ourselves to derive the distribution of the spectral tail process $(\Theta_t^{ij})_{t\geq 0}$ in part (2); part (1) is similar.

If $\mu_n^{\boldsymbol{\sigma}^{ij}}$ denotes the vague limit measure of $(\sigma_{i,0}\sigma_{j,0},\dots,\sigma_{i,n}\sigma_{j,n})'$ the multivariate Breiman lemma yields the vague limit measure $\mu_n^{\mathbf{X}^{ij}}$ of $(X_{i,0}X_{j,0},\dots,X_{i,n}X_{j,n})'$ given by

$$\begin{aligned}\mu_n^{\mathbf{X}^{ij}}(B) &= c\,\mathbb{E}\left[\mu_n^{\boldsymbol{\sigma}^{ij}}(\times_{t=0}^n(B_t/(Z_{it}Z_{jt})))\right] \\ &= c\,\mathbb{E}\left[\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}\left(\times_{t=0}^n\left(\frac{B_t Z_{it}^{-1}Z_{jt}^{-1}}{\prod_{(u,v)\in\Lambda_{i,j,n}^c}\mathrm{e}^{\eta_{u,v}(\psi_{i-u,t-v}+\psi_{j-u,t-v})}}\right)\right)\right] \qquad (4.34)\end{aligned}$$

for any $\mu_n^{\mathbf{X}^{ij}}$-continuity Borel set $B = \times_{t=0}^n B_t \in [-\infty,\infty]^{n+1}\setminus\{\mathbf{0}\}$ bounded away from $\mathbf{0}$, $\Lambda_{i,j,n}$ is equal to $\Lambda_{\mathbf{i},\mathbf{j},\mathbf{t}}$ as defined in (4.30) with $\mathbf{i} = (i,\dots,i)$, $\mathbf{j} = (j,\dots,j)$, $\mathbf{t} = (0,\dots,n)$, and $\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}$ is the limit measure of the regularly varying vector

$$\Big(\prod_{(u,v)\in\Lambda_{i,j,n}}\mathrm{e}^{\eta_{u,v}\,(\psi_{i-u,t-v}+\psi_{j-u,t-v})}\Big)_{t=0,\dots,n}, \qquad (4.35)$$

see the proof of Proposition 4.12. The distribution of the tail process of $(X_{it}X_{jt})$ (cf. Section 4.2.3) is then determined by

$$\begin{aligned}\mathbb{P}((Y\Theta_t^{ij})_{t=0,\dots,n}\in B) &= \lim_{x\to\infty}\frac{\mathbb{P}((X_{it}X_{jt}/x)_{t=0,\dots,n}\in B, |X_{i0}X_{j0}|/x>1)}{\mathbb{P}(|X_{i0}X_{j0}|/x>1)} \\ &= \frac{\mu_n^{\mathbf{X}^{ij}}\big(B\cap\big([-\infty,\infty]\backslash[-1,1]\times[-\infty,\infty]^n\big)\big)}{\mu_n^{\mathbf{X}^{ij}}\big([-\infty,\infty]\backslash[-1,1]\times[-\infty,\infty]^n\big)}.\end{aligned} \tag{4.36}$$

The concrete forms of $\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}$, hence of $\mu_n^{\mathbf{X}^{ij}}$, now depend on whether (4.26) or (4.27) holds.

We first assume (4.26). Note that $\Lambda_{i,j,n}=\cup_{t=0}^n\Lambda_{i,j}^{(t)}$, where $\Lambda_{i,j}^{(t)}=\{(u,v):\psi_{i-u,t-v}+\psi_{j-u,t-v}=\psi^{ij}\}$. Indeed, we easily see that $\Lambda_{i,j}^{(t)}=\Lambda_{i,j}^{(0)}+(0,t)$, $t=1,\dots,n$. We apply Proposition 4.26(ii) to derive the specific form of the limit measure $\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}$ of (4.35). Each component of this vector contains $|\Lambda_{i,j}^{(0)}|$ factors with maximal exponent ψ^{ij}. For the t-th component, those are the factors $\exp(\eta_{u,v}(\psi_{i-u,t-v}+\psi_{j-u,t-v}))$, $(u,v)\in\Lambda_{ij}^{(t)}$. Hence $p_{\text{eff}}=|\Lambda_{i,j}^{(0)}|$ and $P_{\text{eff}}=\{\Lambda_{i,j}^{(0)}+(0,t),t=0,\dots,n\}$. By (4.57), the measure $\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}$, up to a constant multiple, is given by

$$\begin{aligned}&\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}(B)\\ &= c\sum_{s=0}^n\int_0^\infty \mathbb{P}\Big[\Big(\mathbf{1}(\psi_{i-u,t-v}+\psi_{j-u,t-v}=\psi^{ij}\,\forall\,(u,v)\in\Lambda_{i,j}^{(s)})z^{\psi^{ij}} \\ &\qquad \prod_{(u,v)\in\Lambda_{i,j,n}\backslash\Lambda_{i,j}^{(s)}}\mathrm{e}^{\eta_{u,v}(\psi_{i-u,t-v}+\psi_{j-u,t-v})}\Big)_{0\le t\le n}\in B\Big]\nu_\alpha(dz)\\ &= c\sum_{s=0}^n\int_0^\infty \mathbb{P}\Big[\Big(\mathbf{1}(t=s)z^{\psi^{ij}} \\ &\qquad \prod_{(u,v)\in\Lambda_{i,j,n}\backslash\Lambda_{i,j}^{(s)}}\mathrm{e}^{\eta_{u,v}(\psi_{i-u,t-v}+\psi_{j-u,t-v})}\Big)_{0\le t\le n}\in B\Big]\nu_\alpha(dz),\end{aligned}$$

where $\nu_\alpha(dx)=\alpha x^{-\alpha-1}dx$. The s-th measure in the sum above is concentrated on the s-th axis. Therefore the limit measure $\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}$ is concentrated on the axes. By (4.34), this implies that $\mu_n^{\mathbf{X}^{ij}}$ is concentrated on the axes as well. Therefore $\mu_n^{\mathbf{X}^{ij}}(B\cap([-\infty,\infty]\backslash[-1,1])\times[-\infty,\infty]^n)=0$ as soon as one $B_i, 1\le i\le n$, in $B=\times_{i=0}^n B_i$ is bounded away from 0. With (4.36) this gives $Y\Theta_t^{ij}=0$ a.s. for $t\ge1$ and therefore $\Theta_t^{ij}=0$ a.s. for $t\ge1$. The law of Θ_0^{ij} follows from the univariate Breiman lemma.

Next assume (4.27). By Proposition 4.26(i), the vague limit measure $\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}$ is up to a constant given by

$$\begin{aligned}&\tilde{\mu}_n^{\boldsymbol{\sigma}^{ij}}(B)\\ &= \sum_{(u,v)\in\Lambda_{i,j,n}}\int_0^\infty \mathbb{P}\Big[\Big(\mathbf{1}((u,v)\in\Lambda_{i,j}^{(t)})z^{\psi^{ij}} \\ &\qquad \prod_{\substack{(\tilde{u},\tilde{v})\in\Lambda_{i,j,n}\\(\tilde{u},\tilde{v})\neq(u,v)}}\mathrm{e}^{(\psi_{i-\tilde{u},t-\tilde{v}}+\psi_{j-\tilde{u},t-\tilde{v}})\eta_{\tilde{u},\tilde{v}}}\Big)_{t=0,\dots,n}\in B\Big]\nu_\alpha(dz).\end{aligned}$$

For sets B such that $B \cap (\{0\} \times [-\infty,\infty]^n) = \emptyset$ it suffices thereby to sum only over $(u,v) \in \Lambda_{i,j}^{(0)}$ instead over all $(u,v) \in \Lambda_{i,j,n} = \cup_{t=0}^{n} \Lambda_{i,j}^{(t)}$. For these sets we have by Breiman's lemma (cf. (4.34)),

$$\begin{aligned}
&\mu_n^{\mathbf{X}^{ij}}(B)/c \\
= \quad & \sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \int_0^\infty \mathbb{P}\Big[(\mathbf{1}((u,v)\in\Lambda_{i,j}^{(t)})z^{\psi^{ij}} \\
& \prod_{(\tilde u,\tilde v)\neq(u,v)} \mathrm{e}^{(\psi_{i-\tilde u,t-\tilde v}+\psi_{j-\tilde u,t-\tilde v})\eta_{\tilde u,\tilde v}} Z_{it}Z_{jt})_{t=0,\dots,n} \in B\Big]\nu_\alpha(dz) \\
= \quad & \sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \int_0^\infty \mathbb{P}\Big((\mathbf{1}((u,v)\in\Lambda_{i,j}^{(t)})z^{\psi^{ij}} X_{it}X_{jt}\mathrm{e}^{-\psi^{ij}\eta_{u,v}})_{t=0,\dots,n} \in B\Big)\nu_\alpha(dz),
\end{aligned}$$

where we used that if $(u,v) \in \Lambda_{i,j}^{(t)}$, then

$$\prod_{(\tilde u,\tilde v)\neq(u,v)} \mathrm{e}^{(\psi_{i-\tilde u,t-\tilde v}+\psi_{j-\tilde u,t-\tilde v})\eta_{\tilde u,\tilde v}} = \frac{\sigma_{it}\sigma_{jt}}{\mathrm{e}^{(\psi_{i-u,t-v}+\psi_{j-u,t-v})\eta_{u,v}}} = \frac{\sigma_{it}\sigma_{jt}}{\mathrm{e}^{\psi^{ij}\eta_{u,v}}}.$$

Fubini's Theorem and a substitution finally simplify this expression to

$$\begin{aligned}
& \sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \mathbb{E}\left[\int_0^\infty \mathbf{1}\left(\left(\mathbf{1}((u,v)\in\Lambda_{i,j}^{(t)})z^{\psi^{ij}} X_{it}X_{jt}\mathrm{e}^{-\psi^{ij}\eta_{u,v}}\right)_{t=0,\dots,n} \in B\right)\nu_\alpha(dz)\right] \\
= \quad & \sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \mathbb{E}\Bigg[\int_0^\infty \mathbf{1}\left(\left(\mathbf{1}((u,v)\in\Lambda_{i,j}^{(t)})y\frac{X_{it}X_{jt}}{|X_{i0}X_{j0}|}\right)_{t=0,\dots,n} \in B\right) \\
& |X_{i0}X_{j0}|^{\alpha/\psi^{ij}}\mathrm{e}^{-\alpha\eta_{u,v}}\nu_{\frac{\alpha}{\psi^{ij}}}(dy)\Bigg].
\end{aligned}$$

Note that the range of the inner integral in the last expression can be changed from $(0,\infty)$ to $(1,\infty)$, if $B \cap [-1,1] \times [-\infty,\infty]^n = \emptyset$. Therefore, by writing

$$\tilde B_0 = B_0 \setminus [-1,1], \quad \tilde B_t = B_t, \quad t \ge 1, \quad \tilde B = \times_{t=0}^{n} \tilde B_t,$$

we get from (4.36) that

$$\begin{aligned}
&\mathbb{P}\big((Y\Theta_t^{ij})_{t=0,\dots,n} \in B\big) \\
= \quad & \frac{\mu_n^{\mathbf{X}^{ij}}(\tilde B)}{\mu_n^{\mathbf{X}^{ij}}(([-\infty,\infty]\setminus[-1,1]) \times [-\infty,\infty]^n)} \\
= \quad & \frac{\sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \mathbb{E}\left[\int_1^\infty \mathbf{1}_B\left(y\frac{X_{it}X_{jt}}{|X_{i0}X_{j0}|}\mathbf{1}_{\Lambda_{i,j}^{(t)}}(u,v)\right)_{t=0,\dots,n} |X_{i0}X_{j0}|^{\alpha/\psi^{ij}}\mathrm{e}^{-\alpha\eta_{u,v}}\nu_{\frac{\alpha}{\psi^{ij}}}(dy)\right]}{\sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \mathbb{E}\left[|X_{i0}X_{j0}|^{\alpha/\psi^{ij}}\mathrm{e}^{-\alpha\eta_{u,v}}\right]} \\
= \quad & \sum_{(u,v)\in\Lambda_{i,j}^{(0)}} \frac{1}{|\Lambda_{i,j}^{(0)}|} \frac{\mathbb{E}\left[\mathbf{1}_B\left(y\frac{X_{it}X_{jt}}{|X_{i0}X_{j0}|}\mathbf{1}_{\Lambda_{i,j}^{(t)}}(u,v)\right)_{t=0,\dots,n} |X_{i0}X_{j0}|^{\alpha/\psi^{ij}}\right]}{\mathbb{E}\left[|X_{i0}X_{j0}|^{\alpha/\psi^{ij}}\right]},
\end{aligned}$$

where Y is a Pareto(α/ψ^{ij}) random variable, independent of all other random variables in the expression. For the last equation, we expanded both numerator and denominator by multiplying with $\mathbb{E}(e^{\alpha\eta_{u,v}})$, noting that for $(u,v) \in \Lambda_{i,j}^{(0)}$ the random variable $e^{\alpha\eta_{u,v}}$ is independent both of the indicator function and of $|X_{i0}X_{j0}|^{\alpha/\psi^{ij}} e^{-\alpha\eta_{u,v}}$. From the law of the tail process $(Y\Theta_t^{ij})$ we can now see that the law of the spectral tail process (Θ_t^{ij}) satisfies (4.33). □

4.4.2 Infinite variance stable limit theory for the stochastic volatility model and its product processes

In the following result we provide central limit theory with infinite variance stable limits for the sums S_{ij}; see (4.4).

Theorem 4.16. *We consider the stochastic volatility model* (4.2) *and assume the special form of* (σ_{it}) *given in* (4.8) *with* $\psi = 1$. *For given* (i,j)*, define a sequence* (b_n) $n\,\mathbb{P}(|X_{i0}X_{j0}| > b_n) \to 1$ *as* $n \to \infty$*. Assume the following conditions:*

1. *The conditions of Proposition 4.14 hold, ensuring that* $\mathbb{E}[|Z|^{\alpha/\psi^{ij}+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$ *and* $(X_{it}X_{jt})$ *is regularly varying with index* α/ψ^{ij} *and spectral tail process* (Θ_h^{ij}).
2. $(\sigma_{it}\sigma_{jt})$ *is* α*-mixing with rate function* (α_h) *and there exists* $\delta > 0$ *such that* $\alpha_n = o(n^{-\delta})$.
3. *Either*

(i) $\alpha/\psi^{ij} < 1$, *or*

(ii) $i \neq j$, $\alpha/\psi^{ij} \in [1,2)$ *and* Z *is symmetric, or*

(iii) $i = j$, $\alpha/\psi^{ii} = \alpha/2 \in (1,2)$ *and the mixing rate in (2) satisfies* $\sup_n n \sum_{h=r_n}^{\infty} \alpha_h < \infty$ *for some integer sequence* (r_n) $nr_n/b_n^2 \to 0$ *as* $n \to \infty$.

Then

$$b_n^{-1}(S_{ij} - c_n) \xrightarrow{\mathrm{d}} \xi_{ij,\alpha/\psi^{ij}}\,, \tag{4.37}$$

where $\xi_{ij,\alpha/\psi^{ij}}$ *is a totally skewed to the right* α/ψ^{ij}*-stable random variable and*

$$c_n = \begin{cases} n\,\mathbb{E}[X^2] & i = j \text{ and } \alpha \in (2,4)\,, \\ 0 & i \neq j \text{ or } \alpha/\psi^{ij} < 1\,, \end{cases}$$

Remark 4.17.
1. If (α_h) decays at an exponential rate one can choose $r_n = C\log n$ for a sufficiently large constant C. Then $\sup_n n \sum_{h=r_n}^{\infty} \alpha_h < \infty$ and $nr_n/b_n^2 \to 0$ hold. These conditions are also satisfied if $\alpha_h \le cn^{-(1+\gamma)}$ for some $\gamma > 0$, $r_n = Cn^{\xi}$ for some $\xi > 0$ and $1/\gamma \le \xi < 2\psi^{ij}/\alpha - 1$.
2. The sequence $(X_{it}X_{jt})$ inherits α-mixing from $(\sigma_{it}\sigma_{jt})$; see Remark 4.5.
3. It is possible to prove joint convergence for $1 \le i,j \le p$ in (4.37). Due to different tail behavior for distinct (i,j) the normalizing sequences $(b_n) = (b_n^{ij})$ typically increase to infinity at different rates. Then it is only of interest to consider the

joint convergence of those S_{ij} whose summands $X_{it}X_{jt}$ have the same tail index α/ψ^{ij}. More precisely, it suffices to consider those S_{ij} with the property that $X_{it}X_{jt}$ is tail-equivalent to X_{it}^2. The joint convergence follows in a similar way as in the proof below, by observing that Theorem 4.22 is a multivariate limit result. The joint limit of S_{ij} in (4.37) with equivalent tails of index $\tilde{\alpha}$ (say) is jointly $\tilde{\alpha}$-stable with possible dependencies in the limit vector.

4. The strongest normalization is needed for $S_i = S_{ii}$. Recall that the summands X_{it}^2 of S_i are regularly varying with index $\alpha/2$, i.e., $\psi^{ii} = 2$. Let (a_n) be $n\,\mathbb{P}(|X| > a_n) \to 1$. Under the conditions of Theorem 4.16, we have that $a_n^{-2}(S_i - c_n) \stackrel{\mathrm{d}}{\to} \xi_{i,\alpha/2}$, $i = 1,\ldots,p$ for a jointly $\alpha/2$-stable limit. If $\alpha/2 < \alpha/\psi^{ij}$ for some $i \neq j$, then $b_n/a_n^2 \to 0$, hence $a_n^{-2}S_{ij} \stackrel{\mathbb{P}}{\to} 0$. It is possible that $X_{it}X_{jt}$ is regularly varying with index $\alpha/2$ but nevertheless $b_n/a_n^2 \to 0$; see Example 4.18 which deals with the case $\mathbb{E}[\mathrm{e}^{\,\alpha\eta}] = \infty$.

Proof. We apply Theorem 4.22 to the sequence $(X_{it}X_{jt})$, cf. also Remark 4.23.

(1) The regular variation condition on $(X_{it}X_{jt})$ with index α/ψ^{ij} is satisfied by assumption. Moreover, $\Theta_h = 0$ for sufficiently large h; see Remark 4.15.

(2) The assumption about the mixing coefficients in condition (2) implies that for a sufficiently small $\varepsilon \in (0,1)$ and $m_n = n^{1-\varepsilon}$ there exists an integer sequence $l_n = o(m_n)$ such that $k_n\alpha_{l_n} \to 0$. For this choice of m_n and l_n, the proof of the mixing condition for the sums of the truncated variables

$$\underline{S}_{ij} = \sum_{t=1}^{n} X_{it}X_{jt}\mathbf{1}(|X_{it}X_{jt}| > \varepsilon b_n)$$

is now analogous to the proof of the corresponding property in Theorem 4.3.

(3) We want to show that

$$\lim_{l\to\infty} \limsup_{n\to\infty} n \sum_{t=l}^{m_n} \mathbb{P}\big(|X_{it}X_{jt}| > b_n\,, |X_{i0}X_{j0}| > b_n\big) = 0 \tag{4.38}$$

for $m_n = n^{1-\varepsilon}$ as above. Write

$$\sigma_{it}\sigma_{jt} = \prod_{(m,n)} \exp((\psi_{i-m,t-n} + \psi_{j-m,t-n})\eta_{m,n})$$

and set $\Lambda_{\varepsilon,t} = \{(m,n) : \psi_{i-m,t-n} + \psi_{j-m,t-n} \geq 8^{-1}\psi^{ij}\varepsilon\}$, $t \in \mathbb{Z}$. Without loss of generality we assume that l is so large that $\Lambda_{\varepsilon,t} \cap \Lambda_{\varepsilon,0}$ is empty for all $t \geq l$. Then write for $t \geq l$,

$$\sigma_{it}\sigma_{jt} = \sigma_{it,jt,\Lambda_{\varepsilon,t}} \cdot \sigma_{it,jt,\Lambda_{\varepsilon,0}} \cdot \sigma_{it,jt,\Lambda^c_{\varepsilon,0,t}}, \quad \sigma_{i0}\sigma_{j0} = \sigma_{i0,j0,\Lambda_{\varepsilon,0}} \cdot \sigma_{i0,j0,\Lambda_{\varepsilon,t}} \cdot \sigma_{i0,j0,\Lambda^c_{\varepsilon,0,t}},$$

where

$$\sigma_{it_1,jt_1,\Lambda_{\varepsilon,t_2}} = \prod_{(m,n)\in\Lambda_{\varepsilon,t_2}} \exp((\psi_{i-m,t_1-n} + \psi_{j-m,t_1-n})\eta_{m,n}).$$

We conclude that $(\sigma_{it,jt,\Lambda_{\varepsilon,t}},\sigma_{it,jt,\Lambda_{\varepsilon,0}},\,\sigma_{i0,j0,\Lambda_{\varepsilon,0}},\,\sigma_{i0,j0,\Lambda_{\varepsilon,t}})$ and $(\sigma_{it,jt,\Lambda^c_{\varepsilon,0,t}},\sigma_{i0,j0,\Lambda^c_{\varepsilon,0,t}})$ are independent. We have

$$\begin{aligned}&\mathbb{P}\big(|X_{it}X_{jt}|>b_n\,,|X_{i0}X_{j0}|>b_n\big)\\ &\quad\le\ \mathbb{P}\big(\max(|Z_{i0}Z_{j0}|,|Z_{it}Z_{jt}|)\,\max(\sigma_{it,jt,\Lambda^c_{\varepsilon,0,t}},\sigma_{i0,j0,\Lambda^c_{\varepsilon,0,t}})\\ &\qquad\ \min(\sigma_{i0,j0,\Lambda_{\varepsilon,0}}\sigma_{i0,j0,\Lambda_{\varepsilon,t}},\sigma_{it,jt,\Lambda_{\varepsilon,t}}\sigma_{it,jt,\Lambda_{\varepsilon,0}})>b_n\big)\,.\end{aligned}$$

The distribution of $\max(\sigma_{it,jt,\Lambda^c_{\varepsilon,0,t}},\sigma_{i0,j0,\Lambda^c_{\varepsilon,0,t}})$ is stochastically dominated uniformly for $t\ge l$ by a distribution which has moment of order $8\alpha/(\psi^{ij}\varepsilon)>2\alpha/\psi^{ij}$. Furthermore,

$$\begin{aligned}&\min(\sigma_{i0,j0,\Lambda_{\varepsilon,0}}\sigma_{i0,j0,\Lambda_{\varepsilon,t}},\sigma_{it,jt,\Lambda_{\varepsilon,t}}\sigma_{it,jt,\Lambda_{\varepsilon,0}})\\ \le\ &\min\Big(\prod_{(m,n)\in\Lambda_{\varepsilon,0}\cup\Lambda_{\varepsilon,t}}\exp((\psi_{i-m,-n}+\psi_{j-m,-n})(\eta_{m,n})_+),\\ &\qquad\prod_{(m,n)\in\Lambda_{\varepsilon,0}\cup\Lambda_{\varepsilon,t}}\exp((\psi_{i-m,t-n}+\psi_{j-m,t-n})(\eta_{m,n})_+)\Big)\\ \le\ &\min\Big(\prod_{(m,n)\in\Lambda_{\varepsilon,0}}\exp(\psi^{ij}(\eta_{m,n})_+)\prod_{(m',n')\in\Lambda_{\varepsilon,t}}\exp(8^{-1}\psi^{ij}\varepsilon(\eta_{m',n'})_+),\\ &\qquad\prod_{(m',n')\in\Lambda_{\varepsilon,t}}\exp(\psi^{ij}(\eta_{m',n'})_+)\prod_{(m,n)\in\Lambda_{\varepsilon,0}}\exp(8^{-1}\psi^{ij}\varepsilon(\eta_{m,n})_+)\Big)\\ \le\ &\min\Big(\prod_{(m,n)\in\Lambda_{\varepsilon,0}}\exp((\psi^{ij}+8^{-1}\psi^{ij}\varepsilon)(\eta_{m,n})_+),\prod_{(m,n)\in\Lambda_{\varepsilon,t}}\exp((\psi^{ij}+8^{-1}\psi^{ij}\varepsilon)(\eta_{m,n})_+)\Big).\end{aligned}$$

The right-hand side is regularly varying with index $2\alpha/(\psi^{ij}(1+8^{-1}\varepsilon))$. A stochastic domination argument and an application of Breiman's lemma show that uniformly for $l\le t\le m_n$,

$$m_n\,n\,\mathbb{P}\big(|X_{it}X_{jt}|>b_n\,,|X_{i0}X_{j0}|>b_n\big)=n^{2-\varepsilon}o\left(b_n^{-2\alpha/(\psi^{ij}(1+4^{-1}\epsilon))}\right)=n^{2-\varepsilon}o(n^{-2/(1+2^{-1}\varepsilon)})=o(1)$$

which yields (4.38).

(4) We check the vanishing small values condition. For any fixed δ, we write

$$\begin{aligned}\overline{X_{it}X_{jt}}&=X_{it}X_{jt}\mathbf{1}(|X_{it}X_{jt}|\le\delta b_n)\,,\qquad i\ne j\,,\\ \overline{X^2_{it}}&=X^2_{it}\mathbf{1}(X^2_{it}\le\delta b_n)-\mathbb{E}[X^2_{it}\mathbf{1}(X^2_{it}\le\delta b_n)]\,,\\ \overline{S}_{ij}&=\sum_{t=1}^n\overline{X_{it}X_{jt}}\,,\qquad \overline{S}_i=\overline{S}_{ii}\,.\end{aligned}$$

Assume $\alpha/\psi^{ij}\in[1,2)$, $i\ne j$. Then, by symmetry of the random variables Z_{it} and Karamata's theorem for any $\gamma>0$ as $n\to\infty$, $\mathbb{E}[\overline{S}_{ij}]=0$ and

$$\begin{aligned}\mathbb{P}(|\overline{S}_{ij}|>\gamma b_n)&\le(\gamma b_n)^{-2}\mathrm{var}(\overline{S}_{ij})\\ &=n\,(\gamma b_n)^{-2}\mathbb{E}[(\overline{X_{it}X_{jt}})^2]\\ &\sim\gamma^{-2}\,\delta^{2-\alpha}\,,\end{aligned}$$

and the right-hand side converges to zero as $\delta\downarrow 0$.

For $i = j$ and $\alpha/\psi^{ii} > 1$ we need a different argument. We have by Čebyshev's inequality,

$$\begin{aligned}\mathbb{P}(|\overline{S}_i| > \gamma\, b_n) &\leq \gamma^{-2} b_n^{-2} \mathrm{var}\big(\overline{S}_i\big)\\ &= \gamma^{-2}\,(n/b_n^2) \sum_{|h|<n} (1-h/n)\,\mathrm{cov}(\overline{X_{i0}^2}, \overline{X_{ih}^2})\,.\end{aligned}$$

For $|h| \leq h_0$ for any fixed h_0, $(n/b_n^2)|\mathrm{cov}(\overline{X_{i0}^2}, \overline{X_{ih}^2})|$ vanishes by letting first $n \to \infty$ and then $\delta \downarrow 0$. This follows by Karamata's theorem. Standard bounds for the covariance function of an α-mixing sequence (see Doukhan [49], p. 3) yield

$$(n/b_n^2) \sum_{r_n \leq |h| < n} |\mathrm{cov}(\overline{X_{i0}^2}, \overline{X_{ih}^2})| \quad \leq \quad c\,\delta^2 n \sum_{r_n \leq |h| < n} \alpha_h\,,$$

where $r_n \to \infty$ is chosen $\sup_n n \sum_{r_n \leq |h| < \infty} \alpha_h < \infty$ and $nr_n/b_n^2 \to 0$. The right-hand side converges to zero by first letting $n \to \infty$ and then $\delta \downarrow 0$. It remains to show that

$$I_n = (n/b_n^2) \sum_{h_0 < |h| \leq r_n} (1-h/n)\,\mathrm{cov}(\overline{X_{i0}^2}, \overline{X_{ih}^2})$$

is asymptotically negligible. We have

$$\begin{aligned}|I_n| &\leq (n/b_n^2) \sum_{h_0 < |h| \leq r_n} \mathbb{E}[X_{i0}^2\, X_{ih}^2 \mathbf{1}(X_{i0}^2 \leq \delta b_n, X_{ih}^2 \leq \delta b_n)] + c\, n\, r_n/b_n^2\\ &\leq (n/b_n^2) \sum_{h_0 < |h| \leq r_n} \mathbb{E}[X_{i0}^2\, X_{ih}^2] + o(1)\,,\end{aligned}$$

where we used that $n\, r_n/b_n^2 \to 0$. We will show that the summands on the right-hand side are uniformly bounded by a constant if h_0 is sufficiently large. Then $\lim_{n\to\infty} I_n = 0$.

We observe that by Hölder's inequality,

$$\begin{aligned}\mathbb{E}[X_{i0}^2\, X_{ih}^2] &= c\,\mathbb{E}[\sigma_{i0}^2\, \sigma_{ih}^2]\\ &= c\,\mathbb{E}\big[\mathrm{e}^{2\sum_{(k,l)\in\Gamma_\xi} \psi_{kl}(\eta_{i-k,-l}+\eta_{i-k,h-l})}\,\mathrm{e}^{2\sum_{(k,l)\notin\Gamma_\xi} \psi_{kl}(\eta_{i-k,-l}+\eta_{i-k,h-l})}\big]\\ &\leq c\,\big(\mathbb{E}\big[\mathrm{e}^{2r\sum_{(k,l)\in\Gamma_\xi} \psi_{kl}(\eta_{i-k,-l}+\eta_{i-k,h-l})}\big]\big)^{1/r}\\ &\qquad \big(\mathbb{E}\big[\mathrm{e}^{2s\sum_{(k,l)\notin\Gamma_\xi} \psi_{kl}(\eta_{i-k,-l}+\eta_{i-k,h-l})}\big]\big)^{1/s}\,,\end{aligned}$$

where $\Gamma_\xi = \{(k,l) : \psi_{ik} > \xi\}$ for some positive $\xi, s, t\, 1/r + 1/s = 1$. Since σ_{i0}^2 has moments up to order $\alpha/\psi^{ii} \in (1,2)$ and $(\eta_{i-k,-l})_{(k,l)\in\Gamma_\xi}$ and $(\eta_{i-k,h-l})_{(k,l)\in\Gamma_\xi}$ are independent for sufficiently large h we can choose $r > 1$ close to one $\mathbb{E}\big[\mathrm{e}^{2r\sum_{(k,l)\in\Gamma_\xi} \psi_{kl}(\eta_{i-k,-l}+\eta_{i-k,h-l})}\big]$ is finite. This implies that we choose s sufficiently large. On the other hand, for fixed s we can make ξ so small that $\mathbb{E}\big[\mathrm{e}^{2s\sum_{(k,l)\notin\Gamma_\xi} \psi_{kl}(\eta_{i-k,-l}+\eta_{i-k,h-l})}\big]$ is finite and uniformly bounded for sufficiently large h. Fine tuning ξ and s, we may conclude that $\lim_{n\to\infty} I_n = 0$ as desired.

By Theorem 4.22 and Remark 4.23 the result now follows; see also the end of the proof of Theorem 4.3 for the form of the resulting limit law. □

Example 4.18. *We assume that* $\mathbb{E}[\mathrm{e}^{\alpha\eta}] = \infty$*, hence* $\mathrm{e}^{2\eta}$ *does not have a finite* $\alpha/2$*-th moment. Using Lemma 4.24(5), calculation shows that for* $i \neq j$ *with* $\psi^{ij} = 2$*,*

$$\lim_{x\to\infty} \frac{\mathbb{P}(|X_{i0}\,X_{j0}| > x)}{\mathbb{P}(X^2 > x)} = 0 \tag{4.39}$$

Define (a_n) $n\,\mathbb{P}(|X| > a_n) \to 1$*. We may conclude from* (4.39) *and Theorem 4.16 that for* $i \neq j$ *we have* $a_n^{-2} S_{ij} \xrightarrow{\mathbb{P}} 0$ *although both* $X_{i0}\,X_{j0}$ *and* X^2 *are regularly varying with index* $\alpha/2$*.*

By Theorem 4.16 and Remark 4.17 we conclude that

$$a_n^{-2}(S_i - c_n)_{i=1,\dots,p} \xrightarrow{\mathrm{d}} (\xi_{i,\alpha/2})_{i=1,\dots,p}\,, \tag{4.40}$$

where the limit vector consists of $\alpha/2$*-stable components. The spectral tail process* $(\mathbf{\Theta}_h)_{h\geq 1}$ *of the sequence* $\mathbf{X}_t = (X_{1t},\dots,X_{pt})'$*,* $t = 1,2,\dots$*, vanishes. This follows by an argument similar to the proofs of Propositions 4.14 and 4.26 under condition* (4.26)*. A similar argument also yields that*

$$\lim_{x\to\infty} \frac{\mathbb{P}(|X_{i0}| > x\,, |X_{j0}| > x)}{\mathbb{P}(|X| > x)} = 0\,, \qquad i \neq j\,.$$

Therefore the the distribution of $\mathbf{\Theta}_0$ *is concentrated on the axes and has the same form as* $\mathbf{\Theta}_0^{(2)}$ *in* (4.17)*. As in the proof of Theorem 4.3 this implies that the limit random vector in* (4.40) *has iid components.*

We conclude that the limit theory for S_{ij}*,* $1 \leq i,j \leq p$*, are very essentially the same in Case (1) and in Case (2) when the additional condition* $\mathbb{E}[\mathrm{e}^{\alpha\eta}] = \infty$ *holds.*

Example 4.19. *Assume that* (4.27) *holds. We may conclude from Theorem 4.16 that* $a_n^{-2} S_{ij} \xrightarrow{\mathbb{P}} 0$ *for* $i \neq j$ *if* $\psi^{ij} < 2$*. The crucial difference to the previous case appears when* $\psi^{ij} = 2$ *for some* $i \neq j$*. In this case, not only the* $(a_n^{-2}(S_i - c_n))$*,* $i = 1,2,\dots,$ *have totally skewed to the right* $\alpha/2$*-stable limits but we also have* $a_n^{-2} S_{ij} \xrightarrow{\mathrm{d}} \xi_{ij,\alpha/2}$ *for non-degenerate* $\alpha/2$*-stable* $\xi_{ij,\alpha/2}$*. From* (4.28) *we conclude that if* $\psi^{ij} = 2$ *appears then* $\psi^{i'j'} = 2$ *for all* $|i' - j'| = |i - j|$*. This means that non-degenerate limits may appear not only on the diagonal of the matrix* $a_n^{-2}(S_{ij} - c_n)$ *but also along full sub-diagonals.*

In this case, the distribution of $\mathbf{\Theta}_0$ *from the spectral tail process of the sequence* $\mathbf{X}_t = (X_{1t},\dots,X_{pt})'$ *does not have to be concentrated on the axes—in contrast to Example 4.18. This implies that the limiting* $\alpha/2$*-stable random variables* $\xi_{i,\alpha/2}, i = 1,\dots,p$*, are in general not independent. However, similar to the arguments at the end of the proof of Theorem 4.3, one can show that the distribution of the limiting random vector* $(\xi_{i,\alpha/2})_{i=1,\dots,p}$ *is the convolution of distributions of* $\alpha/2$*-stable random vectors which concentrate on hyperplanes of* $\mathbb{R}^p$ *of dimension less or equal than* $|\{(m,n) : \psi_{mn} = 1\}|$*.*

4.4.3 The eigenvalues of the sample covariance matrix of a multivariate stochastic volatility model

In this section we provide some results for the eigenvalues of the sample covariance matrix $\mathbf{X}^n(\mathbf{X}^n)'$ under the conditions of Theorem 4.16. We introduce the sets

$$\Gamma_p = \{(i,j) : 1 \leq i,j \leq p \text{ such that } \psi^{ij} = 2\}\,, \qquad \Gamma_p^c = \{(i,j) : 1 \leq i,j \leq p\}\backslash\Gamma_p$$

and let (a_n) be $n\,\mathbb{P}(|X| > a_n) \to 1$.

Theorem 4.20. *Assume that the conditions of Theorem 4.16 hold for* (X_{it}, X_{jt}), $1 \le i, j \le p$, *and* $\alpha \in (0,4)$. *Then*

$$a_n^{-2} \big\| \mathbf{X}^n (\mathbf{X}^n)' - \widetilde{\mathbf{X}}^n \big\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,,$$

where $\widetilde{\mathbf{X}}^n$ *is a* $p \times p$ *matrix with entries*

$$\widetilde{X}_{ij} = \sum_{t=1}^n X_{it} X_{jt} \mathbf{1}((i,j) \in \Gamma_p)\,, \qquad 1 \le i, j \le p\,.$$

Moreover, if $\mathbb{E}[\mathrm{e}^{\alpha\eta}] = \infty$ *we also have*

$$a_n^{-2} \big\| \mathbf{X}^n (\mathbf{X}^n)' - \operatorname{diag}(\mathbf{X}^n (\mathbf{X}^n)') \big\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,.$$

Proof. We have

$$a_n^{-4} \big\| \mathbf{X}^n (\mathbf{X}^n)' - \widetilde{\mathbf{X}}^n \big\|_2^2 \le \sum_{(i,j) \in \Gamma_p^c} \big(a_n^{-2} S_{ij} \big)^2\,.$$

For $(i,j) \in \Gamma_p^c$ we have $i \ne j$ and the sequence $(X_{it} X_{jt})$ is regularly varying with index $\alpha/\psi^{ij} > \alpha/2$. In view of Theorem 4.16 the right-hand side converges to zero in probability.

In the case when $\mathbb{E}[\mathrm{e}^{\alpha\eta}] = \infty$ we learned in Example 4.18 that $a_n^{-2} S_{ij} \xrightarrow{\mathbb{P}} 0$ whenever $i \ne j$. This concludes the proof. □

For any $p \times p$ non-negative definite matrix $\mathbf{A}$ write $\lambda_i(\mathbf{A})$, $i = 1, \dots, p$, for its eigenvalues and $\lambda_{(1)}(\mathbf{A}) \ge \dots \ge \lambda_{(p)}(\mathbf{A})$ for their ordered values. For the eigenvalues of $\mathbf{X}^n (\mathbf{X}^n)'$ we keep the previous notation (λ_i),

Corollary 4.21. *Assume the conditions of Theorem 4.20 and* $\alpha \in (0,4) \backslash \{2\}$. *Then*

$$a_n^{-2} \max_{i=1,\dots,p} \big| \lambda_{(i)} - \lambda_{(i)}(\widetilde{\mathbf{X}}^n) \big| \xrightarrow{\mathbb{P}} 0\,. \tag{4.41}$$

and

$$\begin{aligned} & a_n^{-2} \Big(\lambda_{(i)} - n\,\mathbb{E}[X^2] 1(\alpha \in (2,4)) \Big)_{i=1,\dots,p} \\ \xrightarrow{\mathrm{d}} \quad & \Big(\lambda_{(i)} \big((\xi_{kl,\alpha/2} \mathbf{1}((k,l) \in \Gamma_p))_{1 \le k,l \le p} \big) \Big)_{i=1,\dots,p}\,, \end{aligned} \tag{4.42}$$

where $(\xi_{ij,\alpha/2})_{(i,j) \in \Gamma_p}$ *are jointly* $\alpha/2$*-stable (possibly degenerate for* $i \ne j$*) random variables. Moreover, in the case when* $\mathbb{E}[\mathrm{e}^{\alpha\eta}] = \infty$ *we have*

$$a_n^{-2} \Big(\lambda_{(i)} - n\,\mathbb{E}[X^2] 1(\alpha \in (2,4)) \Big)_{i=1,\dots,p} \xrightarrow{\mathrm{d}} \big(\xi_{(i),\alpha/2} \big)_{i=1,\dots,p}\,, \tag{4.43}$$

where $(\xi_{i,\alpha/2})_{i=1,\dots,p}$ *are iid totally skewed to the right* $\alpha/2$*-stable random variables with order statistics* $\xi_{(1),\alpha/2} \ge \dots \ge \xi_{(p),\alpha/2}$.

Proof. Relation (4.41) is an immediate consequence of Theorem 4.20 and Weyl's inequality; see Bhatia [20]. We conclude from Theorem 4.16 and Remark 4.17(3) that

$$a_n^{-2} \big(S_{ij} - n\,\mathbb{E}[X^2]\, \mathbf{1}(\alpha \in (2,4)) \big)_{(i,j) \in \Gamma_p} \xrightarrow{\mathrm{d}} \big(\xi_{ij,\alpha/2} \big)_{(i,j) \in \Gamma_p}\,. \tag{4.44}$$

Then (4.42) follows. Relation (4.43) is a special case of (4.42). If $\mathbb{E}[\mathrm{e}^{\alpha\eta}] = \infty$ then, in view of Example 4.18, only the diagonal elements in (4.44) have non-degenerate iid $\alpha/2$-stable limits. □

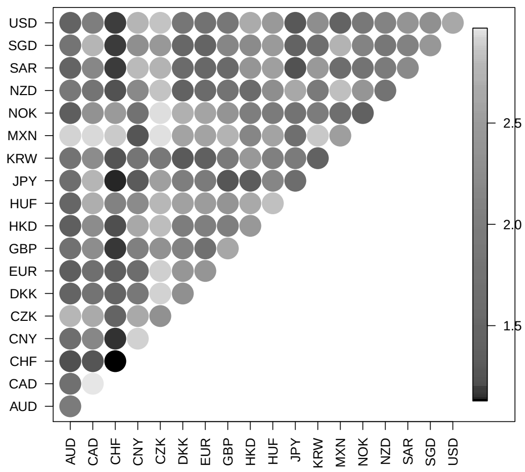
Some conclusions

By virtue of this corollary and in view of Section 4.3.3 the results for the eigenvalues in Case (1) and in Case (2) when $\mathbb{E}[e^{\alpha\eta}] = \infty$ are very much the same. Moreover, the results in Section 4.3.4 remain valid in the latter case.

If (4.27) holds, Case (2) is quite different from Case (1); see Example 4.19. In this case not only the diagonal of the matrix $\mathbf{X}^n(\mathbf{X}^n)'$ determines the asymptotic behavior of its eigenvalues and eigenvectors. Indeed, if $\psi^{ij} = 2$ for some $i \neq j$, then at least two sub-diagonals of $\mathbf{X}^n(\mathbf{X}^n)'$ have non-degenerate $\alpha/2$-limits and these sub-diagonals together with the diagonal determine the asymptotic behavior of the eigenspectrum. The limiting diagonal elements are dependent in contrast to Case (1). This fact and the presence of sub-diagonals are challenges if one wants to calculate the limit distributions of the eigenvalues and eigenvectors.

4.5 Simulations and data example

In this section we illustrate the behavior of sample covariance matrices for moderate sample sizes for the models discussed in Sections 4.3 and 4.4 and we compare them with a real-life data example. These data consist of 1567 daily log-returns of foreign exchange (FX) rates from 18 currencies against the Swedish Kroner (SEK) from January 4th 2010 to April 1st 2016, as made available by the Swedish National Bank. To start with, the Hill estimators of the tail indices $\alpha_{ij}, 1 \leq i, j, \leq 18$, of the cross products $X_{it}X_{jt}, 1 \leq i, j, \leq 18$, are visualized in Figure 4.1.



In particular, the Hill estimators on the diagonal (corresponding to the series $X_{it}^2, 1 \leq i \leq 18$) of the values $\alpha_i/2$, where α_i is the tail index of the ith currency, are of similar size although not identical. Even if all series had the same tail index the Hill estimator exhibits high statistical uncertainty which even increases for serially dependent data, cf. Drees [50].

Figure 4.1: Estimated tail indices of cross products for the FX rates of 18 currencies against SEK. The indices are derived by Hill estimators with threshold equal to the 97%-quantile of $n = 1567$ observations.

A way to make the data more homogeneous in their tails is to rank-transform their marginals to the same distribution. We do, however, refrain from such a transformation to keep the correlation structure of the original data unchanged.

It is clearly visible that some off-diagonal components of the matrix have an estimated tail index which is comparable to the on-diagonal elements. This implies that the tails of the corresponding off-diagonal entries $S_{ij}, i \neq j$, of the sample covariance matrix may be of a similar magnitude as the on-diagonal entries S_i. This is in stark contrast to the asymptotic behavior of the models analyzed in Section 4.3.

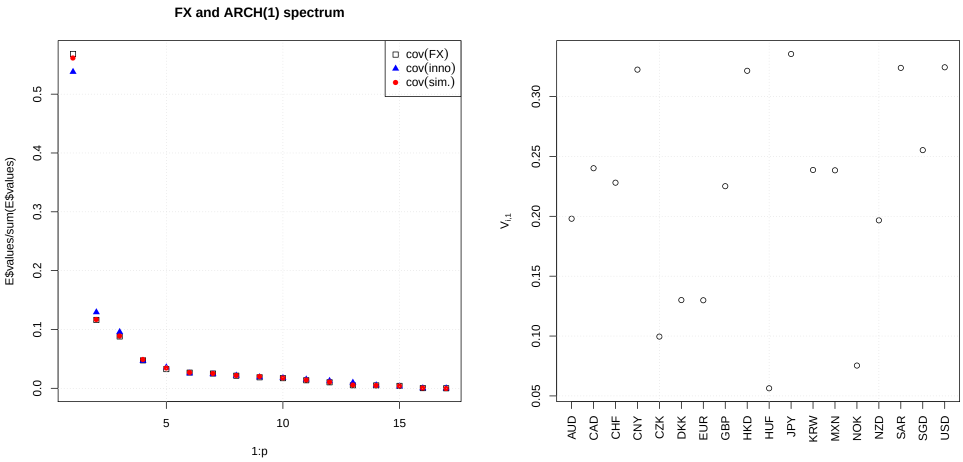


Figure 4.2: Normalized and ordered eigenvalues (left) and eigenvector corresponding to largest eigenvalue (right) of real and simulated data, with $n = 1567, p = 18$. Based on FX rate data of 18 foreign currencies against SEK.

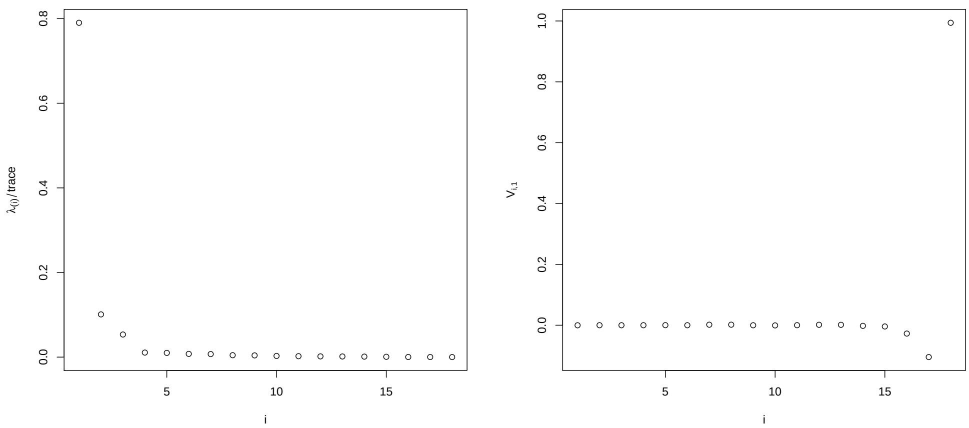


Figure 4.3: Based on a stochastic volatility model with heavy-tailed innovation sequence.

Figure 4.2 shows the ordered eigenvalues of the sample covariance matrix (normalized by its trace) and the eigenvector of the FX rate data corresponding to the largest eigenvalue. There exists a notable spectral gap between the largest and second largest

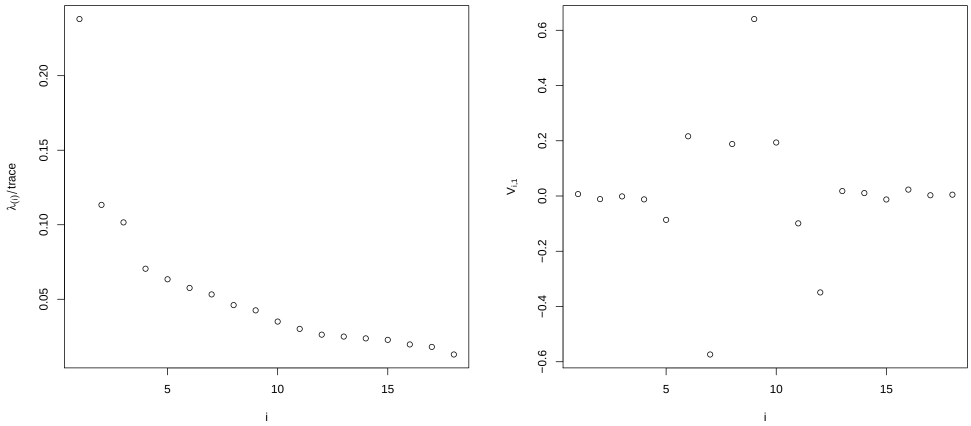


Figure 4.4: Based on a stochastic volatility model with heavy-tailed volatility sequence that satisfies assumptions of Example 4.18

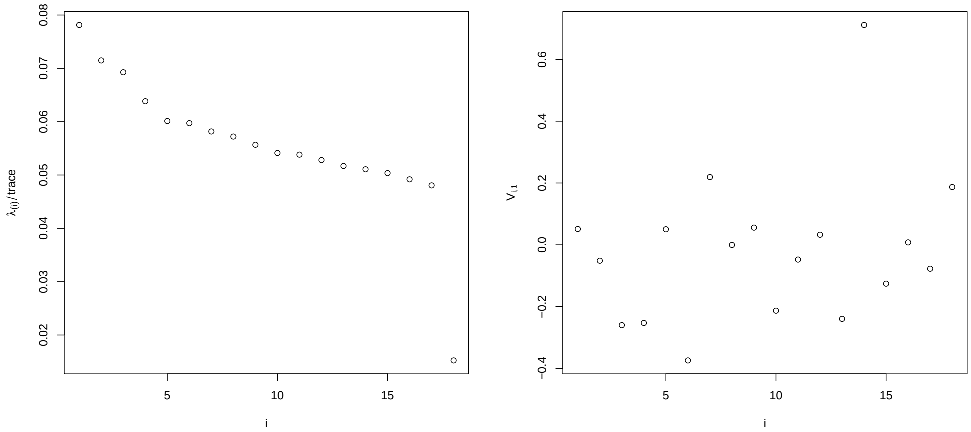


Figure 4.5: Based on a stochastic volatility model with heavy-tailed volatility sequence that satisfies assumptions of Example 4.19

eigenvalues and the unit eigenvector corresponding to the largest eigenvector has all positive and non-vanishing components. For comparison and to illustrate the variety of the models discussed above we also plot corresponding realizations of three model specifications from Sections 4.3 and 4.4. In all cases we choose $p = 18$ and $n = 1567$ in accordance with the data example. We assume throughout a moving average structure in the log-volatility process $\log \sigma_{it}$ in (4.2). More specifically,

$$\sigma_{it} = \exp(\sum_{k=1}^{18} \eta_{i-k,t}), \quad 1 \le i \le 18, \ t \in \mathbb{Z}. \tag{4.45}$$

In accordance with the model properties discussed in Section 4.3, we first assume iid standard Gaussian $\eta_{i,t}$ and iid Z_{it} with a Student-t distribution with $t = 3$ degrees of freedom. Figure 4.3 shows the normalized eigenvalues and the first unit eigenvector from a realization of this model. We notice a relatively large gap between the first and second eigenvalue and, in accordance with Section 4.3.4.4, we see that the first unit eigenvector is relatively close to a unit basis vector. Figure 4.4 shows the corresponding realizations for the model (4.45) with a specification according to Example 4.18, i.e., Exponential(3)-distributed iid $\eta_{i,t}$ (meaning that $\mathbb{P}(\eta_{i,t} > x) = \exp(-3x), x \geq 0$, which implies $\alpha = 3$ and $\mathbb{E}[\mathrm{e}^{3\eta}] = \infty$) and iid standard Gaussian Z_{it}. Compared to the first simulated model, we see a slower decay in the magnitude of the ordered eigenvalues and a more spread out first unit eigenvector. This observation illustrates that although the limit behavior of this model and the one analyzed before should be very similar (cf. Example 4.18), convergence to the prescribed limit appears slower for the heavy-tailed volatility sequence than for the heavy-tailed innovations. Finally, Figure 4.5 shows a simulation drawn from (4.45) where the $\eta_{i,t}$ are iid such that $\mathbb{P}(\eta_{i,t} > x) \sim x^{-2}\exp(-3x), x \to \infty$, and the Z_{it} are iid standard Gaussian. Again, $\alpha = 3$, but direct calculations show that the distribution of $\eta_{i,t}$ is convolution equivalent, i.e., it satisfies (4.27) instead of (4.26). The graphs are in line with the analysis in Example 4.19 and illustrate a very spread out dominant eigenvector. We note that while none of the three very simple models analyzed in the simulations above is able to fully describe the behavior of the analyzed data, the two models with heavy-tailed volatility and light-tailed innovations are able to explain a non-concentrated first unit eigenvector of the sample covariance matrix and therefore non-negligible dependence between components as seen in the data.

Acknowledgements

Thomas Mikosch's and Xiaolei Xie's research is partly supported by the Danish Research Council Grant DFF-4002-00435 "Large random matrices with heavy tails and dependence". Parts of the paper were written when Mohsen Rezapour visited the Department of Mathematics at the University of Copenhagen December 2015–January 2016. He would like to thank the Department of Mathematics for hospitality.

4.6 Some α-stable limit theory

In this paper, we make frequently use of Theorem 4.3 in Mikosch and Wintenberger [90] which we quote for convenience:

Theorem 4.22. *Let* $(\mathbf{Y}_t)$ *be an* $\mathbb{R}^p$*-valued strictly stationary sequence,* $\mathbf{S}_n = \mathbf{Y}_1 + \cdots + \mathbf{Y}_n$ *and* (a_n) *be* $n\,\mathbb{P}(\|\mathbf{Y}\| > a_n) \to 1$*. Also write for* $\varepsilon > 0$, $\overline{\mathbf{Y}}_t = \mathbf{Y}_t \mathbf{1}(\|\mathbf{Y}_t\| \leq \varepsilon a_n)$, $\underline{\mathbf{Y}}_t = \mathbf{Y}_t - \overline{\mathbf{Y}}_t$ *and*

$$\overline{\mathbf{S}}_{l,n} = \sum_{t=1}^{l} \overline{\mathbf{Y}}_t \qquad \underline{\mathbf{S}}_{l,n} = \sum_{t=1}^{l} \underline{\mathbf{Y}}_t \,.$$

Assume the following conditions:

1. $(\mathbf{Y}_t)$ *is regularly varying with index* $\alpha \in (0,2) \setminus \{1\}$ *and spectral tail process* $(\boldsymbol{\Theta}_j)$*.*

2. *A mixing condition holds: there exists an integer sequence $m_n \to \infty$ $k_n = [n/m_n] \to \infty$ and*

$$\mathbb{E}\mathrm{e}^{\,i\mathbf{t}'\underline{\mathbf{S}}_n/a_n} - \Big(\mathbb{E}\mathrm{e}^{\,i\mathbf{t}'\underline{\mathbf{S}}_{m_n,n}/a_n}\Big)^{k_n} \to 0\,, \qquad n \to \infty\,, \qquad \mathbf{t} \in \mathbb{R}^p\,. \tag{4.46}$$

3. *An anti-clustering condition holds:*

$$\lim_{l\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\max_{t=l,\ldots,m_n} \|\mathbf{Y}_t\| > \delta a_n \mid \|\mathbf{Y}_0\| > \delta a_n\big) = 0\,, \qquad \delta > 0 \tag{4.47}$$

for the same sequence (m_n) as in (2).

4. *If $\alpha \in (1,2)$, in addition $\mathbb{E}[\mathbf{Y}] = \mathbf{0}$ and the vanishing small values condition holds:*

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\big(a_n^{-1}\|\overline{\mathbf{S}}_n - \mathbb{E}[\overline{\mathbf{S}}_n]\| > \delta\big) = 0\,, \qquad \delta > 0 \tag{4.48}$$

and $\sum_{i=1}^{\infty} \mathbb{E}[\|\mathbf{\Theta}_i\|] < \infty$.

Then $a_n^{-1}\mathbf{S}_n \xrightarrow{\mathrm{d}} \xi_\alpha$ for an α-stable $\mathbb{R}^p$-valued vector ξ_α with log characteristic function

$$\int_0^\infty \mathbb{E}\big[\mathrm{e}^{\,i\,y\,\mathbf{t}'\sum_{j=0}^{\infty}\mathbf{\Theta}_j} - \mathrm{e}^{\,i\,y\,\mathbf{t}'\sum_{j=1}^{\infty}\mathbf{\Theta}_j} - i\,y\,\mathbf{t}'\mathbf{1}_{(1,2)}(\alpha)\big]\,d(-y^\alpha)\,, \qquad \mathbf{t} \in \mathbb{R}^p\,. \tag{4.49}$$

Remark 4.23. *If we additionally assume that $\mathbf{Y}$ is symmetric, which implies $\mathbb{E}[\overline{\mathbf{Y}}] = \mathbf{0}$, then the statement of the theorem also holds for $\alpha = 1$.*

4.7 (Joint) Tail behavior for products of regularly varying random variables

In this paper, we make frequently use of the tail behavior of products of non-negative independent random variables X and Y. In particular, we are interested in conditions for the existence of the limit

$$\lim_{x\to\infty} \frac{\mathbb{P}(XY > x)}{\mathbb{P}(X > x)} = q\,. \tag{4.50}$$

for some $q \in [0,\infty]$. We quote some of these results for convenience.

Lemma 4.24. *Let X and Y be independent random variables.*

1. *If X and Y are regularly varying with index $\alpha > 0$ then XY is regularly varying with the same index.*

2. *If X is regularly varying with index $\alpha > 0$ and $\mathbb{E}[Y^{\alpha+\varepsilon}] < \infty$ for some $\varepsilon > 0$ then (4.50) holds with $q = \mathbb{E}[Y^\alpha]$.*

3. *If X and Y are iid regularly varying with index $\alpha > 0$ and $\mathbb{E}[Y^\alpha] < \infty$, then (4.50) holds with $q = 2\mathbb{E}[|Y|^\alpha]$ iff*

$$\lim_{M\to\infty} \limsup_{x\to\infty} \frac{\mathbb{P}(XY > x, M < Y \le x/M)}{\mathbb{P}(X > x)} = 0\,. \tag{4.51}$$

4. *If X and Y are regularly varying with index $\alpha > 0$, $\mathbb{E}[Y^\alpha + X^\alpha] < \infty$, $\lim_{x\to\infty} \mathbb{P}(Y > x)/\mathbb{P}(X > x) = 0$ and (4.51) holds, then (4.50) holds with $q = \mathbb{E}[|Y|^\alpha]$.*

5. *Assume that $\mathbb{E}[|Y|^\alpha] = \infty$. Then (4.50) holds with $q = \infty$.*

Proof. (1) This is proved in Embrechts and Goldie [57].
(2) This is Breiman's [27] result.
(3) This is Proposition 3.1 in Davis and Resnick [46].
(4) This part is proved similarly to (3); we borrow the ideas from [46]. For $M > 0$ we have the following decomposition

$$\begin{aligned}\frac{\mathbb{P}(XY > x)}{\mathbb{P}(X > x)} &= \frac{\mathbb{P}(XY > x, Y \le M)}{\mathbb{P}(X > x)} + \frac{\mathbb{P}(XY > x, M < Y \le x/M)}{\mathbb{P}(X > x)} + \frac{\mathbb{P}(XY > x, Y > x/M)}{\mathbb{P}(X > x)} \\ &\sim \mathbb{E}[Y^\alpha \mathbf{1}(Y \le M)] + \frac{\mathbb{P}(XY > x, M < Y \le x/M)}{\mathbb{P}(X > x)} + \mathbb{E}[(X \wedge M)^\alpha] \frac{\mathbb{P}(Y > x)}{\mathbb{P}(X > x)} \\ &= \mathbb{E}[Y^\alpha \mathbf{1}(Y \le M)] + \frac{\mathbb{P}(XY > x, M < Y \le x/M)}{\mathbb{P}(X > x)} + o(1)\,.\end{aligned}$$

Here we applied Breiman's result twice. The second term vanishes by virtue of (4.51). Thus $q = \mathbb{E}[Y^\alpha]$.
(5) The same argument as for (4) yields as $x \to \infty$,

$$\frac{\mathbb{P}(XY > x)}{\mathbb{P}(X > x)} \ge \frac{\mathbb{P}(XY > x, Y \le M)}{\mathbb{P}(X > x)} \sim \mathbb{E}[Y^\alpha \mathbf{1}(Y \le M)]\,.$$

Then (4.50) with $q = \infty$ is immediate. □

Lemma 4.25. *Let $Y_1, \ldots, Y_p \ge 0$ be iid regularly varying random variables with index $\alpha > 0$. Assume that*

$$\lim_{t\to\infty} \frac{\mathbb{P}(Y_1 \cdot Y_2 > t)}{\mathbb{P}(Y_1 > t)} = c \in (0, \infty)\,. \tag{4.52}$$

Then for any $a_1, \ldots, a_p \ge 0$ such that $a_{\max} := \max_{j=1,\ldots,p} a_j > 0$ and any $v > 0$ we have

$$\lim_{t\to\infty} \frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i} > vt)}{\mathbb{P}(Y_1^{a_{\max}} > t)} = \sum_{j:a_j = a_{\max}} \lim_{s\to 0} \lim_{t\to\infty} \frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i} > vt, Y_j^{a_{\max}} > st)}{\mathbb{P}(Y_1^{a_{\max}} > t)} \tag{4.53}$$

and

$$\lim_{s\to 0} \limsup_{t\to\infty} \frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i} > vt, \max_{j=1,\ldots,p} Y_j^{a_{\max}} \le st)}{\mathbb{P}(Y_1^{a_{\max}} > t)} = 0. \tag{4.54}$$

Proof. In view of Davis and Resnick [47] the only possible value for c in (4.52) is $2\mathbb{E}[Y_1^\alpha]$ (which implies that $\mathbb{E}[Y_1^\alpha] < \infty$). Furthermore, we note that the product $\prod_{j:a_j = a_{\max}} Y_j^{a_j}$ is regularly varying with index $-\alpha/a_{\max}$; see Embrechts and Goldie [57], Corollary on p. 245. By Breiman's lemma this implies that

$$\begin{aligned}&\lim_{t\to\infty} \frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i} > vt)}{\mathbb{P}(Y_1^{a_{\max}} > t)} \\ =\ &\lim_{t\to\infty} \frac{\mathbb{P}(Y_1^{a_{\max}} > vt)}{\mathbb{P}(Y_1^{a_{\max}} > t)} \frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i} > vt)}{\mathbb{P}(Y_1^{a_{\max}} > vt)} \\ =\ &v^{-\alpha/a_{\max}} \Big(\prod_{j:a_j \ne a_{\max}} \mathbb{E}[Y_j^{\alpha a_j/a_{\max}}] \Big) \lim_{t\to\infty} \frac{\mathbb{P}(\prod_{j:a_j = a_{\max}} Y_j^{a_{\max}} > vt)}{\mathbb{P}(Y_1^{a_{\max}} > vt)}.\end{aligned}$$

By Lemma 2.5 in Embrechts and Goldie [56] (cf. also Chover, Ney and Wainger [32]) this equals

$$v^{-\alpha/a_{max}}\Big(\prod_{j:a_j\neq a_{\max}}\mathbb{E}[Y_j^{\alpha a_j/a_{\max}}]\Big)|\{j:a_j=a_{\max}\}|\,\mathbb{E}[Y_1^{\alpha}]^{|\{j:a_j=a_{\max}\}|-1}.$$

On the other hand, we have

$$\begin{aligned}
&\sum_{j:a_j=a_{\max}}\lim_{s\to 0}\lim_{t\to\infty}\frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i}>vt, Y_j^{a_{\max}}>st)}{\mathbb{P}(Y_1^{a_{\max}}>t)}\\
=&\sum_{j:a_j=a_{\max}}\lim_{s\to 0}\lim_{t\to\infty}\frac{\mathbb{P}(Y_j^{a_{\max}}\min(s^{-1},v^{-1}\prod_{k\neq j}Y_k^{a_k})>t)}{\mathbb{P}(Y_j^{a_{\max}}>t)}\\
=&\sum_{j:a_j=a_{\max}}\lim_{s\to 0}\mathbb{E}[(\min(s^{-1},v^{-1}\prod_{k\neq j}Y_k^{a_k}))^{\alpha/a_{\max}}]\\
=&\ v^{-\alpha/a_{\max}}\sum_{j:a_j=a_{\max}}\prod_{k\neq j}\mathbb{E}[Y_k^{\alpha a_k/a_{\max}}]\\
=&\ v^{-\alpha/a_{\max}}\Big(\prod_{j:a_j\neq a_{\max}}\mathbb{E}[Y_j^{\alpha a_j/a_{\max}}]\Big)|\{j:a_j=a_{\max}\}|\,\mathbb{E}[Y_1^{\alpha}]^{|\{j:a_j=a_{\max}\}|-1},
\end{aligned}$$

where we applied Breiman's lemma in the second step to the bounded random variable $\min(s^{-1},\ v^{-1}\prod_{k\neq j}Y_k^{a_k})$, and the monotone convergence theorem in the penultimate step. This proves (4.53). To prove (4.54) note that for $s>0$,

$$\begin{aligned}
&\frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i}>vt)}{\mathbb{P}(Y_1^{a_{\max}}>t)}\\
\geq\ &\frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i}>vt, \max_{j=1,\ldots,p}Y_j^{a_{\max}}\leq st)}{\mathbb{P}(Y_1^{a_{\max}}>t)}\\
&+\sum_{j:a_j=a_{\max}}\frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i}>vt, Y_j^{a_{\max}}>st)}{\mathbb{P}(Y_1^{a_{\max}}>t)}\\
&-\frac{\mathbb{P}(\prod_{i=1}^p Y_i^{a_i}>vt, Y_{j_1}^{a_{\max}}>st, Y_{j_2}^{a_{\max}}>st \text{ for some } j_1\neq j_2)}{\mathbb{P}(Y_1^{a_{\max}}>t)},\quad s>0.
\end{aligned}$$

The last summand on the right-hand side converges to 0 as $t\to\infty$ by independence of the $Y_j's$. Moreover, the left-hand term and the second term on the right-hand side become equal by first $t\to\infty$ and then $s\to 0$, in view of (4.53). Therefore the first right-hand term vanishes by first $t\to\infty$ and then $s\to 0$. This proves the statement. □

Proposition 4.26. *Let $Y_1,\ldots,Y_p\geq 0$ be iid regularly varying with index α and $(a_{ij})\in[0,\infty)^{n\times p}, n,p\geq 1$, be such that $\max_{1\leq i\leq n}a_{ik}=a_{\max}:=\max_{i,j}a_{ij}>0$ for any $1\leq k\leq p$.*

(i) Assume that (4.52) holds. Then the random vector

$$\mathbf{Y}:=\big(\prod_{j=1}^p Y_j^{a_{ij}}\big)_{1\leq i\leq n} \tag{4.55}$$

is regularly varying with index $\alpha/a_{\max}$. Furthermore, up to a constant the limit measure μ of $\mathbf{Y}$ is given by $\sum_{j=1}^p \mu_j$, where for any Borel set $B \in [0,\infty]^n$ bounded away from $\mathbf{0}$ and $\nu_\alpha(dz) = \alpha z^{-\alpha-1}dz$,

$$\mu_j(B) \;=\; \int_0^\infty \mathbb{P}\Big(\Big(\mathbf{1}(a_{ij} = a_{\max}) z^{a_{\max}} \prod_{k\neq j} Y_k^{a_{ik}}\Big)_{1\le i\le n} \in B\Big)\nu_\alpha(dz). \quad (4.56)$$

(ii) Assume that $\mathbb{E}[Y_1^\alpha] = \infty$. Set

$$\begin{aligned} p_{\mathit{eff}} &:= \max_i |\{1 \le j \le p : a_{ij} = a_{\max}\}|\,, \\ P_{\mathit{eff}} &:= \{A \subset \{1,\dots,p\} : |A| = p_{\mathit{eff}} \wedge \exists\, i : \forall\, j \in A : a_{ij} = a_{\max}\}\,. \end{aligned}$$

Then the random vector $\mathbf{Y}$ in (4.55) is regularly varying with index $\alpha/a_{\max}$. Furthermore, up to a constant the limit measure μ of $\mathbf{Y}$ is equal to $\sum_{A\in P_{\mathit{eff}}} \mu_A$, where for any Borel set $B \in [0,\infty]^n$ bounded away from $\mathbf{0}$,

$$\mu_A(B) \;=\; \int_0^\infty \mathbb{P}\Big(\Big(\mathbf{1}(a_{ij} = a_{\max}\,\forall\, j \in A) z^{a_{\max}} \prod_{k\notin A} Y_k^{a_{ik}}\Big)_{1\le i\le n} \in B\Big)\nu_\alpha(dz)\,. \quad (4.57)$$

Proof. (i) Let $B \in [0,\infty]^n$ be a Borel set bounded away from $\mathbf{0}$. For $s > 0$ we have

$$\begin{aligned} \frac{\mathbb{P}(\mathbf{Y} \in tB)}{\mathbb{P}(Y_1^{a_{\max}} > t)} \;=\; & \frac{\mathbb{P}(\mathbf{Y} \in tB, \max_{j=1,\dots,p} Y_j^{a_{\max}} \le st)}{\mathbb{P}(Y_1^{a_{\max}} > t)} + \sum_{j=1}^p \frac{\mathbb{P}(\mathbf{Y} \in tB, Y_j^{a_{\max}} > st)}{\mathbb{P}(Y_1^{a_{\max}} > t)} \\ & - \frac{\mathbb{P}(\mathbf{Y} \in tB, Y_{j_1}^{a_{\max}} > st, Y_{j_2}^{a_{\max}} > st, \text{ for some } j_1 \neq j_2)}{\mathbb{P}(Y_1^{a_{\max}} > t)}. \end{aligned} \quad (4.58)$$

Since B is bounded away from $\mathbf{0}$, there exists $v > 0$ and $1 \le i \le n$ such that $B \subset \{(x_1,\dots,x_n) \in [0,\infty]^n : x_i > v\}$. From Lemma 4.25, (4.54) the first summand in (4.58) therefore tends to 0 by first $t \to \infty$ and then $s \to 0$. Furthermore, the third summand converges to zero as $t \to \infty$ by independence of the $Y_j's$. We are thus left to show

$$\lim_{s\searrow 0}\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{Y} \in tB, Y_j^{a_{\max}} > st)}{\mathbb{P}(Y_1^{a_{\max}} > t)} = \mu_j(B), \quad 1 \le j \le p,$$

with μ_j as in (4.56). For $s > 0$ write

$$\begin{aligned} & \lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{Y} \in tB, Y_j^{a_{\max}} > st)}{\mathbb{P}(Y_1^{a_{\max}} > t)} \\ =\; & s^{-\alpha/a_{\max}} \lim_{t\to\infty} \mathbb{P}(\mathbf{Y} \in tB \mid Y_j^{a_{\max}} > st) \\ =\; & s^{-\alpha/a_{\max}} \lim_{t\to\infty} \mathbb{P}\left(\left(\left(\frac{Y_j^{a_{\max}}}{st}\right)^{\frac{a_{ij}}{a_{\max}}} s^{\frac{a_{ij}}{a_{\max}}} t^{\frac{a_{ij}}{a_{\max}}-1} \prod_{k\neq j} Y_k^{a_{ik}}\right)_{1\le i\le n} \in B \;\middle|\; Y_j^{a_{\max}} > st\right) \\ =\; & s^{-\alpha/a_{\max}} \int_1^\infty \mathbb{P}\left(\left(\mathbf{1}(a_{ij} = a_{\max}) sy \prod_{k\neq j} Y_k^{a_{ik}}\right)_{1\le i\le n} \in B\right) \nu_{\alpha/a_{\max}}(dy). \end{aligned}$$

Substituting sy by z in the integral finally gives

$$\lim_{s\searrow 0}\lim_{t\to\infty}\frac{\mathbb{P}(\mathbf{Y}\in tB, Y_j^{a_{\max}}>st)}{\mathbb{P}(Y_1^{a_{\max}}>t)}$$
$$= \int_0^\infty \mathbb{P}\left(\left(\mathbf{1}(a_{ij}=a_{\max})z^{a_{\max}}\prod_{k\neq j}Y_k^{a_{ik}}\right)_{1\le i\le n}\in B\right)\nu_\alpha(dz).$$

(ii) Note first that under our assumptions for any $1\le n_1<n_2\le p$,

$$\begin{aligned}\lim_{t\to\infty}\frac{\mathbb{P}(\prod_{j=1}^{n_2}Y_j>t)}{\mathbb{P}(\prod_{j=1}^{n_1}Y_j>t)} &= \lim_{t\to\infty}\int_0^\infty\frac{\mathbb{P}(\prod_{j=1}^{n_1}Y_j>t/y)}{\mathbb{P}(\prod_{j=1}^{n_1}Y_j>t)}P^{\prod_{j=n_1+1}^{n_2}Y_j}(dy)\\ &\ge \mathbb{E}\left[\prod_{j=n_1+1}^{n_2}Y_j^\alpha\right]=\infty \qquad (4.59)\end{aligned}$$

by Fatou's lemma and the regular variation of $\prod_{j=1}^{n_1}Y_j$. Write now

$$\mathbf{Y}=\sum_{\substack{1\le i\le n\\ |\{j:a_{ij}=a_{\max}\}|=p_{\text{eff}}}}\prod_{j=1}^p Y_j^{a_{ij}}\mathbf{e}_i+\sum_{\substack{1\le i\le n\\ |\{j:a_{ij}=a_{\max}\}|<p_{\text{eff}}}}\prod_{j=1}^p Y_j^{a_{ij}}\mathbf{e}_i, \qquad (4.60)$$

where $\mathbf{e}_i$ stands for the i-th unit vector. The first sum can also be written as

$$\sum_{A\in P_{\text{eff}}}\text{diag}((\mathbf{1}(a_{ij}=a_{\max}\,\forall\, j\in A)\prod_{k\notin A}Y_k^{a_{ik}})_{1\le i\le n})\prod_{j\in A}Y_j^{a_{\max}}=:\sum_{A\in P_{\text{eff}}}\mathbf{Y}^A, \qquad (4.61)$$

where for each summand the random matrix and the random factor are independent and for the non-zero entries of the matrix we have $a_{ik}<a_{\max}$ since $k\notin A$. Thus, by the multivariate version of Breiman's lemma each $\mathbf{Y}^A$ is a multivariate regularly varying vector with limit measure μ_A (up to a constant multiplier) as in (4.57) and normalizing function $P(\prod_{i=1}^{p_{\text{eff}}}Y_i^{a_{\max}}>x)$. Furthermore, for $A,A'\in P_{\text{eff}}$ with $A\neq A'$ and i,i' such that $a_{ij}=a_{\max}\,\forall\, j\in A$ and $a_{i'j}=a_{\max}\,\forall\, j\in A'$ we have

$$\begin{aligned}&\frac{\mathbb{P}(\mathbf{Y}_i^A>x,\mathbf{Y}_{i'}^{A'}>x)}{\mathbb{P}(\prod_{i=1}^{p_{\text{eff}}}Y_i^{a_{\max}}>x)} \qquad (4.62)\\ =\;&\frac{\mathbb{P}((\prod_{j\in A\cap A'}Y_j)^{a_{\max}}\prod_{j\in(A\cap A')^c}Y_j^{a_{ij}}>x,(\prod_{j\in A\cap A'}Y_j)^{a_{\max}}\prod_{j\in(A\cap A')^c}Y_j^{a_{i'j}}>x)}{\mathbb{P}(\prod_{i=1}^{p_{\text{eff}}}Y_i^{a_{\max}}>x)}.\end{aligned}$$

By Janßen and Drees [78], Theorem 4.2 (in connection with Remark 4.3 (ii) and the minor change that our random variables are regularly varying with index α instead of 1), the numerator behaves asymptotically like $\mathbb{P}((\prod_{j\in A\cap A'}Y_j)^{a_{\max}}>x)$, since $\kappa_0=a_{\max}^{-1},\kappa_j=0, j\in(A\cap A')^c$ is the unique non-negative optimal solution to

$$\kappa_0+\sum_{j\in(A\cap A')^c}\kappa_j\to\min!$$

under

$$\kappa_0a_{\max}+\sum_{j\in(A\cap A')^c}\kappa_ja_{ij}\ge 1,\quad \kappa_0a_{\max}+\sum_{j\in(A\cap A')^c}\kappa_ja_{i'j}\ge 1.$$

This is because $\min(a_{ij}, a_{i'j}) < a_{\max}$ and $\max(a_{ij}, a_{i'j}) \leq a_{\max}$ for all $j \in (A \cap A')^c$. Since $A \neq A'$, we have $|A \cap A'| < p_{\text{eff}}$ and thus, by (4.59), the expression (4.62) converges to 0 as $x \to \infty$. Therefore, each component of $\mathbf{Y}^A$ is asymptotically independent of each component of $\mathbf{Y}^{A'}$ and thus the sum in (4.61) is multivariate regularly varying with limit measure $\sum_{A \in P_{\text{eff}}} \mu_A$ and normalizing function $\mathbb{P}(\prod_{i=1}^{p_{\text{eff}}} Y_i^{a_{\max}} > x)$. Since the second sum in (4.60) consists by (4.59) only of random vectors for which $\mathbb{P}(\| \prod_{j=1}^{p} Y_j^{a_{ij}} \mathbf{e}_i \| > x) = \mathbb{P}(\prod_{j=1}^{p} Y_j^{a_{ij}} > x) = o(\mathbb{P}(\prod_{i=1}^{p_{\text{eff}}} Y_i^{a_{\max}} > x))$, we have that $\mathbf{Y}$ is regularly varying with index $\alpha / a_{\max}$ and limit measure $\sum_{A \in P_{\text{eff}}} \mu_A$ by Lemma 3.12 in Jessen and Mikosch [81]. $\square$

Chapter 5

Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series

RICHARD DAVIS, JOHANNES HEINY,
THOMAS MIKOSCH & XIAOLEI XIE
Extremes 19, 3 (2016), 517–547.

Abstract

We provide some asymptotic theory for the largest eigenvalues of a sample covariance matrix of a p-dimensional time series where the dimension $p = p_n$ converges to infinity when the sample size n increases. We give a short overview of the literature on the topic both in the light- and heavy-tailed cases when the data have finite (infinite) fourth moment, respectively. Our main focus is on the heavy-tailed case. In this case, one has a theory for the point process of the normalized eigenvalues of the sample covariance matrix in the iid case but also when rows and columns of the data are linearly dependent. We provide limit results for the weak convergence of these point processes to Poisson or cluster Poisson processes. Based on this convergence we can also derive the limit laws of various functionals of the ordered eigenvalues such as the joint convergence of a finite number of the largest order statistics, the joint limit law of the largest eigenvalue and the trace, limit laws for successive ratios of ordered eigenvalues, etc. We also develop some limit theory for the singular values of the sample autocovariance matrices and their sums of squares. The theory is illustrated for simulated data and for the components of the S&P 500 stock index.

keywords: Regular variation, sample covariance matrix, dependent entries, largest eigenvalues, trace, point process convergence, cluster Poisson limit, infinite variance stable limit, Fréchet distribution. **subject class:** Primary 60B20; Secondary 60F05 60F10 60G10 60G55 60G70

5.1 Estimation of the largest eigenvalues

5.1.1 The light-tailed case

One of the exciting new areas of statistics is concerned with analyses of large data sets. For such data one often studies the dependence structure via covariances and correlations. In this paper we focus on one aspect: the estimation of the eigenvalues of the covariance matrix of a multivariate time series when the dimension p of the series increases with the sample size n. In particular, we are interested in limit theory for the largest eigenvalues of the sample covariance matrix. This theory is closely related to topics from classical extreme value theory such as maximum domains of attraction with the corresponding normalizing and centering constants for maxima; cf. Embrechts et al. [58], Resnick [96, 97]. Moreover, point process convergence with limiting Poisson and cluster Poisson processes enters in a natural way when one describes the joint convergence of the largest eigenvalues of the sample covariance matrix. Large deviation techniques find applications, linking extreme value theory with random walk theory and point process convergence. The objective of this paper is to illustrate some of the main developments in random matrix theory for the particular case of the sample covariance matrix of multivariate time series with independent or dependent entries. We give special emphasis to the heavy-tailed case when extreme value theory enters in a rather straightforward way.

Classical multivariate time series analysis deals with observations which assume values in a p-dimensional space where p is "relatively small" compared to the sample size n. With the availability of large data sets p can be "large" relative to n. One of the possible consequences is that standard asymptotics (such as the central limit theorem) break down and may even cause misleading results.

The dependence structure in multivariate data is often summarized by the covariance matrix which is typically estimated by its sample analog. For example, principal component analysis (PCA) extracts principal component vectors corresponding to the largest eigenvalues of the sample covariance matrix. The magnitudes of these eigenvalues provide an empirical measure of the importance of these components.

If p, n are fixed, a column of the $p \times n$ data matrix

$$\mathbf{X} = \mathbf{X}_n = \big(X_{it}\big)_{i=1,\dots,p;t=1,\dots,n}$$

represents an observation of a p-dimensional time series model with unknown parameters. In this section we assume that the real-valued entries X_{it} are iid, unless mentioned otherwise, and we write X for a generic element. One challenge is to infer information about the parameters from the eigenvalues $\lambda_1, \dots, \lambda_p$ of the *sample covariance matrix* $\mathbf{X}\mathbf{X}'$. In the notation we suppress the dependence of (λ_i) on n and p. If p and n are finite and the columns of $\mathbf{X}$ are iid and multivariate normal, Muirhead [91] derived a (rather complicated) formula for the joint distribution of the eigenvalues (λ_i).

For p fixed and $n \to \infty$, assuming $\mathbf{X}$ has centered normal entries and a diagonal covariance matrix Σ, Anderson [5] derived the joint asymptotic density of $(\lambda_1, \dots, \lambda_p)$. We quote from Johnstone [83]: "The classic paper by Anderson [5] gives the limiting joint distribution of the roots, but the marginal distribution of the largest eigenvalue is hard to extract even in the null case" (i.e., when the covariance matrix Σ is proportional to the identity matrix).

It turns out that limit theory for the largest eigenvalues becomes "easier" when the dimension p increases with n. Over the last 15 years there has been increasing interest in the case when $p = p_n \to \infty$ as $n \to \infty$. In most of the literature (exceptions are El Karoui [55], Davis et al. [44, 45] and Heiny and Mikosch [68]) one assumes that p and n grow at the same rate:

$$\frac{p}{n} \to \gamma \qquad \text{for some } \gamma \in (0, \infty). \tag{5.1}$$

In random matrix theory, the convergence of the *empirical spectral distributions* $(F_{n^{-1}\mathbf{X}\mathbf{X}'})$ of a sequence $(n^{-1}\mathbf{X}\mathbf{X}')$ of non-negative definite matrices is the principle object of study. The empirical spectral distribution $F_{n^{-1}\mathbf{X}\mathbf{X}'}$ is constructed from the eigenvalues via

$$F_{n^{-1}\mathbf{XX}'}(x) = \frac{1}{p}\,\#\{1 \le j \le p : n^{-1}\lambda_j \le x\}, \quad x \in \mathbb{R}, \quad n \ge 1.$$

In the literature convergence results for the sequence of empirical spectral distributions are established under the assumption that p and n grow at the same rate. Suppose that the iid entries Z_{it} have mean 0 and variance 1. If (5.1) holds, then, with probability one, $(F_{n^{-1}\mathbf{XX}'})$ converges to the celebrated Marčenko–Pastur law with absolutely continuous part given by the density,

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}\,, & \text{if } a \le x \le b, \\ 0\,, & \text{otherwise,} \end{cases} \tag{5.2}$$

where $a = (1-\sqrt{\gamma})^2$ and $b = (1+\sqrt{\gamma})^2$. The Marčenko–Pastur law has a point mass $1 - 1/\gamma$ at the origin if $\gamma > 1$, cf. Bai and Silverstein [9, Chapter 3]. The point mass at zero is intuitively explained by the fact that, with probability 1, $\min(p, n)$ eigenvalues λ_i are non-zero. When $n = (1/\gamma)\ p$ and $\gamma > 1$ one sees that the proportion of non-zero eigenvalues of the sample covariance matrix is $1/\gamma$ while the proportion of zero eigenvalues is $1 - 1/\gamma$.

While the finite second moment is the central assumption to obtain the Marčenko–Pastur law as the limiting spectral distribution, the finite fourth moment plays a crucial role when studying the largest eigenvalues

$$\lambda_{(1)} \ge \cdots \ge \lambda_{(p)} \tag{5.3}$$

of $\mathbf{XX}'$, where we suppress the dependence on n in the notation.

Assuming (5.1) and iid entries X_{it} with zero mean, unit variance and finite fourth moment, Geman [62] showed that

$$\frac{\lambda_{(1)}}{n} \overset{\text{a.s.}}{\to} \big(1+\sqrt{\gamma}\big)^2, \qquad n \to \infty\,. \tag{5.4}$$

Johnstone [83] complemented this strong law of large numbers by the corresponding central limit theorem in the special case of iid standard normal entries:

$$n^{2/3}\,\frac{(\sqrt{\gamma})^{1/3}}{\big(1+\sqrt{\gamma}\big)^{4/3}}\Big(\frac{\lambda_{(1)}}{n} - \big(1+\sqrt{\tfrac{p}{n}}\big)^2\Big) \overset{\text{d}}{\to} \text{TW}\,, \tag{5.5}$$

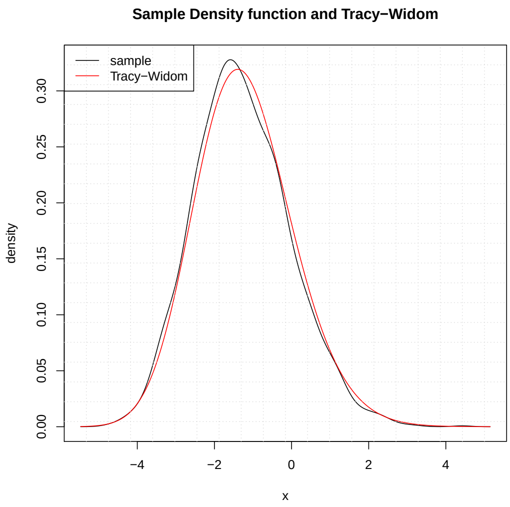
where the limiting random variable has a *Tracy–Widom distribution* of order 1. Notice that the centering $\big(1+\sqrt{\frac{p}{n}}\big)^2$ can in general not be replaced by $(1+\sqrt{\gamma})^2$. This distribution is ubiquitous in random matrix theory. Its distribution function F_1 is given by

$$F_1(s) = \exp\Big\{-\frac{1}{2}\int_s^\infty [q(x) + (x-s)q^2(x)]\,\mathrm{d}x\Big\}\,,$$

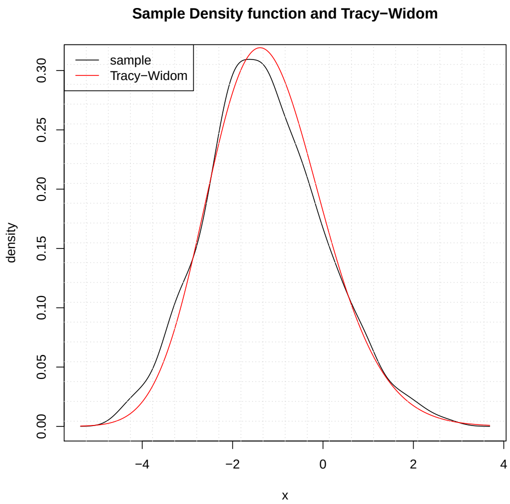
where $q(x)$ is the unique solution to the Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x)\,,$$

where $q(x) \sim \text{Ai}(x)$ as $x \to \infty$ and $\text{Ai}(\cdot)$ is the Airy kernel; see Tracy and Widom [112] for details. We notice that the rate $n^{2/3}$ compares favorably to the $\sqrt{n}$-rate in the classical



(a) Standard normal entries



(b) Entry distribution: $\mathbb{P}(X = \sqrt{3}) = \mathbb{P}(X = -\sqrt{3}) = 1/6$, $\mathbb{P}(X = 0) = 2/3$. Note $\mathbb{E}X = 0$, $\mathbb{E}[X^2] = 1$, $\mathbb{E}[X^3] = 0$ and $\mathbb{E}[X^4] = 3$, i.e., the first 4 moments of X match those of the standard normal distribution.

Figure 5.1: Sample density function of the largest eigenvalue compared with the Tracy–Widom density function. The data matrix $\mathbf{X}$ has dimension 200×1000. An ensemble of 2000 matrices is simulated.

central limit theorem for sums of iid finite variance random variables. The calculation of the spectrum is facilitated by the fact that the distribution of the classical Gaussian matrix ensembles is invariant under orthogonal transformations. The corresponding computation for non-invariant matrices with non-Gaussian entries is more complicated and was a major challenge for several years; a first step was made by Johansson [82]. Johnstone's result was extended to matrices $\mathbf{X}$ with iid non-Gaussian entries by Tao and Vu [108, Theorem 1.16]. Assuming that the first four moments of the entry distribution match those of the standard normal distribution, they showed (5.5) by employing *Lindeberg's replacement method*, i.e., the iid non-Gaussian entries are replaced step-by-step by iid Gaussian ones. This approach is well-known from summation theory for sequences of iid random variables. Tao and Vu's result is a consequence of the so-called *Four Moment Theorem*, which describes the insensitivity of the eigenvalues with respect to changes in the distribution of the entries. To some extent (modulo the strong moment matching conditions) it shows the universality of Johnstone's limit result (5.5). Later we will deal with entries with infinite fourth moment. In this case, the weak limit for the normalized largest eigenvalue $\lambda_{(1)}$ is distinct from the Tracy–Widom distribution: the classical Fréchet extreme value distribution appears. In Figure 5.1 we illustrate how the Tracy–Widom approximation works for Gaussian and non-Gaussian entries of $\mathbf{X}$ and in Figure 5.2 we also illustrate that this approach fails when $\mathbb{E}[X^4] = \infty$.

Figure 5.1 compares the sample density function of the properly normalized largest eigenvalue estimated from 2000 simulated sample covariance matrices $\mathbf{X}\mathbf{X}'$ ($n = 1000, p = 200$) with the Tracy–Widom density. If X has infinite fourth moment and further regularity conditions on the tail hold then the Tracy–Widom limiting law needs to be replaced by the Fréchet distribution; see Section 5.1.2 for details. Figure 5.2 illustrates this fact with a simulated ensemble whose entries are distributed according to the heavy-tailed distribution from (5.32) below with $\alpha = 1.6$.

5.1.2 The heavy-tailed case

So far we focused on "light-tailed" $\mathbf{X}$ in the sense that its entries have finite fourth moment. However, there is statistical evidence that the assumption of finite fourth moment may be violated when dealing with data from insurance, finance or telecommunications. We illustrate this fact in Figure 5.3 where we show the pairs (α_L, α_U) of lower and upper tail indices of $p = 478$ log-return series composing the S&P 500 index estimated from $n = 1,345$ daily observations from 01/04/2010 to 02/28/2015. This means we assume for every row $(X_{it})_{t=1,\ldots,n}$ of $\mathbf{X}$ that the tails behave like

$$\mathbb{P}(X_{it} > x) \sim c_U\, x^{-\alpha_U} \qquad \text{and} \qquad \mathbb{P}(X_{it} < -x) \sim c_L\, x^{-\alpha_L}\,, \qquad x \to \infty\,,$$

for non-negative constants c_L, c_U. We apply the Hill estimator (see Embrechts et al. [58], p. 330, de Haan and Ferreira [67], p. 69) to the time series of the gains and losses in a naive way, neglecting the dependence and non-stationarity in the data; we also omit confidence bands. From the figure it is evident that the majority of the return series have tail indices below four, corresponding to an infinite fourth moment. The behavior of the largest eigenvalue $\lambda_{(1)}$ changes dramatically when $\mathbf{X}$ has infinite fourth moment. Bai and Silverstein [10] proved for an $n \times n$ matrix $\mathbf{X}$ with iid centered entries that

$$\limsup_{n\to\infty} \frac{\lambda_{(1)}}{n} = \infty \qquad \text{a.s.} \tag{5.6}$$

This is in stark contrast to Geman's result (5.4).

In the heavy-tailed case it is common to assume a *regular variation condition:*

$$\mathbb{P}(X > x) \sim p_+ \frac{L(x)}{x^\alpha} \qquad \text{and} \qquad \mathbb{P}(X < -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \tag{5.7}$$

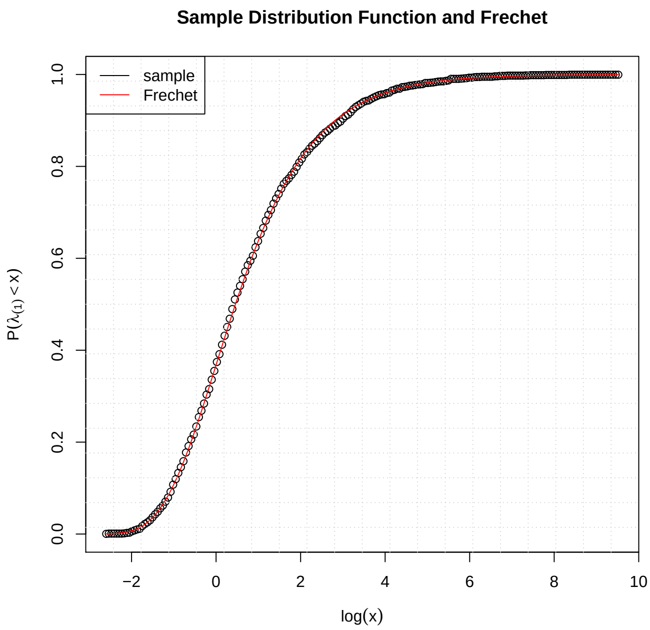


Figure 5.2: Sample distribution function of the largest eigenvalue $\lambda_{(1)}$ compared to the Fréchet distribution (solid line) with $\alpha = 1.6$. The data matrices have dimension 200×1000 and iid entries with infinite fourth moment. The results are based on 2000 replicates.

where $p_\pm$ are non-negative constants $p_+ + p_- = 1$ and L is a slowly varying function. In particular, if $\alpha < 4$ we have $\mathbb{E}[X^4] = \infty$. The regular variation condition on X (we will also refer to X as a regularly varying random variable) is needed for proving asymptotic theory for the eigenvalues of $\mathbf{XX}'$. This is similar to proving limit theory for sums of iid random variables with infinite variance stable limits; see for example Feller [61].

In (5.2) we have seen that the sequence $(F_{n^{-1}\mathbf{XX}'})$ of empirical spectral distributions converges to the Marčenko–Pastur law if the centered iid entries possess a finite second moment. Now we will discuss the situation when the entries are still iid and centered, but have an infinite variance. Here we assume the entries to be regularly varying with index $\alpha \in (0,2)$. Assuming (5.1) with $\gamma \in (0,1]$ in this infinite variance case, Belinschi et al. [16, Theorem 1.10] showed that the sequence $(F_{a_{n+p}^{-2}\mathbf{XX}'})$ converges with probability one to a non-random probability measure with density ρ_α^γ satisfying

$$\rho_\alpha^\gamma(x) x^{1+\alpha/2} \to \frac{\alpha\gamma}{2(1+\gamma)}, \quad x \to \infty,$$

see also Ben Arous and Guionnet [18, Theorem 1.6]. The normalization (a_k) is chosen $\mathbb{P}(|X| > a_k) \sim k^{-1}$ as $k \to \infty$. An application of the Potter bounds (see Bingham et al. [21, p. 25]) shows that $a_{n+p}^2/n \to \infty$.

It is interesting to note that there is a phase change in the extreme eigenvalues in going from finite to infinite fourth moment, while the phase change occurs for the empirical spectral distribution going from finite to infinite variance.

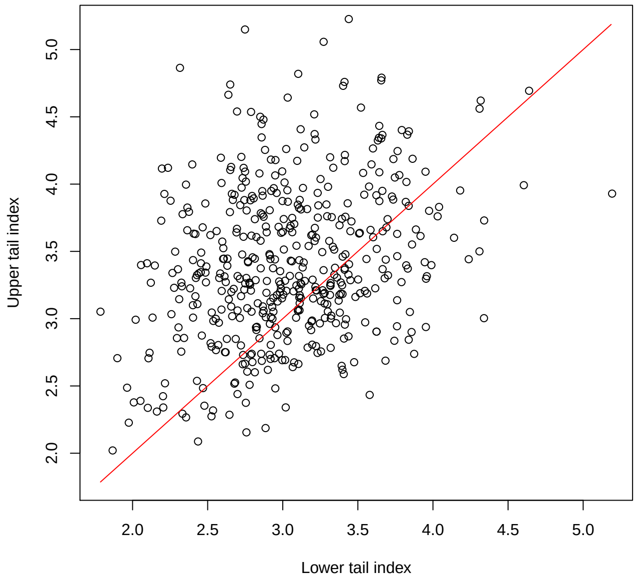


Figure 5.3: Tail indices of log-returns of 478 time series from the S&P 500 index. The values $(\hat{\alpha}_L, \hat{\alpha}_U)$ of the lower and upper tail indices are provided by Hill's estimator. We also draw the line $\hat{\alpha}_U = \hat{\alpha}_L$.

The theory for the largest eigenvalues of sample covariance matrices with heavy tails is less developed than in the light-tailed case. Pioneering work for $\lambda_{(1)}$ in the case of iid regularly varying entries X_{it} with index $\alpha \in (0,2)$ is due to Soshnikov [105, 106]. He showed the point process convergence

$$N_n = \sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}\lambda_i} \xrightarrow{\mathrm{d}} N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \qquad n \to \infty\,, \tag{5.8}$$

under the growth condition (5.1) on (p_n). Here

$$\Gamma_i = E_1 + \cdots + E_i\,, \qquad i \ge 1\,, \tag{5.9}$$

and (E_i) is an iid standard exponential sequence. In other words, N is a Poisson point process on $(0,\infty)$ with mean measure $\mu(x,\infty) = x^{-\alpha/2}$, $x > 0$. Convergence in distribution of point processes is understood in the sense of weak convergence in the space of point measures equipped with the vague topology; see Resnick [96, 97]. We can easily derive the limiting distribution of $a_{np}^{-2}\lambda_{(k)}$ for fixed $k \ge 1$ from (5.8):

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(a_{np}^{-2}\lambda_{(k)} \le x) &= \lim_{n\to\infty} \mathbb{P}(N_n(x,\infty) < k) = \mathbb{P}(N(x,\infty) < k) = \mathbb{P}(\Gamma_k^{-2/\alpha} \le x)\\ &= \sum_{s=0}^{k-1} \frac{\big(\mu(x,\infty)\big)^s}{s!}\, \mathrm{e}^{-\mu(x,\infty)}, \quad x > 0.\end{aligned}$$

In particular,

$$\frac{\lambda_{(1)}}{a_{np}^2} \xrightarrow{\mathrm{d}} \Gamma_1^{-\alpha/2}, \qquad n \to \infty,$$

where the limit has *Fréchet distribution* with parameter $\alpha/2$ and distribution function

$$\Phi_{\alpha/2}(x) = \mathrm{e}^{-x^{-\alpha/2}}, \qquad x > 0.$$

We mention that the tail balance condition (5.7) may be replaced in this case by the weaker assumption $\mathbb{P}(|X| > x) = L(x)x^{-\alpha}$ for a slowly varying function L. Indeed, it follows from the proof that only the squares X_{it}^2 contribute to the point process limits of the eigenvalues (λ_i). A consequence of the continuous mapping theorem and (5.8) is the joint convergence of the upper order statistics: for any $k \geq 1$,

$$a_{np}^{-2}\big(\lambda_{(1)}, \ldots, \lambda_{(k)}\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha}, \ldots, \Gamma_k^{-2/\alpha}\big), \qquad n \to \infty.$$

It follows from standard theory for point processes with iid points (e.g. Resnick [96, 97]) that (5.8) remains valid if we replace N_n by the point process $\sum_{i=1}^p \sum_{t=1}^n \varepsilon_{X_{it}^2/a_{np}^2}$. Then we also have for any $k \geq 1$,

$$a_{np}^{-2}\big(X_{(1),np}^2, \ldots, X_{(k),np}^2\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha}, \ldots, \Gamma_k^{-2/\alpha}\big), \qquad n \to \infty, \tag{5.10}$$

where

$$X_{(1),np}^2 \geq \cdots \geq X_{(np),np}^2$$

denote the order statistics of $(X_{it}^2)_{i=1,\ldots,p;t=1,\ldots,n}$.

Auffinger et al. [7] showed that (5.8) remains valid under the regular variation condition (5.7) for $\alpha \in (2,4)$, the growth condition (5.1) on (p_n) and the additional assumption $\mathbb{E}[X] = 0$. Of course, (5.10) remains valid as well. Davis et al. [45] extended these results to the case when the rows of $\mathbf{X}$ are iid linear processes with iid regularly varying noise. The Poisson point process convergence result of (5.8) remains valid in this case. Different limit processes can only be expected if there is dependence across rows and columns.

In what follows, we refer to the *heavy-tailed case* when we assume the regular variation condition (5.7) for some $\alpha \in (0,4)$.

5.1.3 Overview

The primary objective of this overview is to make a connection between extreme value theory and the behavior of the largest eigenvalues of sample covariance matrices from heavy-tailed multivariate time series. For time series that are linearly dependent through time and across rows, it turns out that the extreme eigenvalues are essentially determined by the extreme order statistics from an array of iid random variables. The asymptotic behavior of the extreme eigenvalues is then derived routinely from classical extreme value theory. As such, explicit joint distributions of the extreme order statistics can be given which yield a plethora of ancillary results. Convergence of the point process of extreme eigenvalues, properly normalized, plays a central role in establishing the main results.

In Section 5.2 we continue the study of the case when the data matrix $\mathbf{X}$ consists of iid heavy-tailed entries. We will consider power-law growth rates on the dimension

(p_n) that is more general than prescribed by (5.1). In Section 5.3 we introduce a model for X_{it} which allows for linear dependence across the rows and through time. The point process convergence of normalized eigenvalues is presented in Section 5.3.4. This result lays the foundation for new insight into the spectral behavior of the sample covariance matrix, which is the content of Section 5.4.1.

Sections 5.4.1 and 5.4.3 are devoted to *sample autocovariance matrices*. Motivated by [86], we study the eigenvalues of sums of transformed matrices and illustrate the results in two examples. These results are applied to the time series of S&P 500 in Section 5.4.2. Appendix 5.A contains useful facts about regular variation and point processes.

5.2 General growth rates for p_n in the iid heavy-tailed case

This section is based on ideas in Heiny and Mikosch [68] where one can also find detailed proofs.

Growth conditions on (p_n)

In many applications it is not realistic to assume that the dimension p of the data and the sample size n grow at the same rate. The aforementioned results of Soshnikov [105, 106] and Auffinger et al. [7] already indicate that the value γ in the growth rate (5.1) does not appear in the distributional limits. This obervation is in contrast to the light-tailed case; see (5.4) and (5.5). Davis et al. [44, 45] and Heiny and Mikosch [68] allowed for more general rates for $p_n \to \infty$ than linear growth in n. Recall that $p = p_n \to \infty$ is the number of rows in the matrix $\mathbf{X}_n$. We need to specify the growth rate of (p_n) to ensure a non-degenerate limit distribution of the normalized singular values of the sample autocovariance matrices. To be precise, we assume

$$p = p_n = n^{\beta}\ell(n), \quad n \ge 1, \qquad (C_p(\beta))$$

where ℓ is a slowly varying function and $\beta \ge 0$. If $\beta = 0$, we also assume $\ell(n) \to \infty$. Condition $C_p(\beta)$ is more general than the growth conditions in the literature; see [7, 44, 45].

Theorem 5.1. *Assume that $\mathbf{X} = \mathbf{X}_n$ has iid entries satisfying the regular variation condition (5.7) for some $\alpha \in (0,4)$. If $\mathbb{E}[|X|] < \infty$ we also suppose that $\mathbb{E}[X] = 0$. Let (p_n) be an integer sequence satisfying $C_p(\beta)$ with $\beta \ge 0$. In addition, we require*

$$\min(\beta, \beta^{-1}) \in (\alpha/2 - 1, 1] \qquad \textit{for } \alpha \in [2,4), \qquad (\widetilde{C}_\beta(\alpha))$$

Then

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}\lambda_i} \xrightarrow{\text{d}} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \qquad n \to \infty\,, \tag{5.11}$$

where the convergence holds in the space of point measures with state space $(0,\infty)$ equipped with the vague toplogy; see Resnick [97].

A discussion of the case $\beta \in [0,1]$

We mentioned earlier that in the heavy-tailed case, limit theory for the largest eigenvalues of the sample covariance matrix is rather insensitive to the growth rate of (p_n) and that the limits are essentially determined by the diagonal of this matrix. This is confirmed by the following result.

Proposition 5.2. *Assume that $\mathbf{X} = \mathbf{X}_n$ has iid entries satisfying the regular variation condition (5.7) for some $\alpha \in (0,4)$. If $\mathbb{E}[|X|] < \infty$ we also suppose that $\mathbb{E}[X] = 0$. Then for any sequence (p_n) satisfying $C_p(\beta)$ with $\beta \in [0,1]$ we have*

$$a_{np}^{-2} \|\mathbf{X}\mathbf{X}' - \operatorname{diag}(\mathbf{X}\mathbf{X}')\|_2 \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,,$$

where $\|\cdot\|_2$ denotes the spectral norm; see (5.22) for its definition.

Proposition 5.2 is not unexpected for two reasons:

- It is well-known from classical theory (see Embrechts and Veraverbeke [59]) that for any iid regularly varying non-negative random variables Y, Y' with index $\alpha' > 0$, $Y\,Y'$ is regularly varying with index α' while Y^2 is regularly varying with index $\alpha'/2$. Therefore X^2 and $X_{11}X_{12}$ are regularly varying with indices $\alpha/2$ and α, respectively.
- The aforementioned tail behavior is inherited by the entries of $\mathbf{X}\mathbf{X}'$ in the following sense. By virtue of Nagaev-type large deviation results for an iid regularly varying sequence (Y_i) with index $\alpha' \in (0,2)$ where we also assume that $\mathbb{E}[Y_0] = 0$ if $\mathbb{E}[|Y_0|] < \infty$ (see Theorem 5.21) we have that $\mathbb{P}(Y_1 + \cdots + Y_n > b_n)/(n\,\mathbb{P}(|Y_0| > b_n))$ converges to a non-negative constant provided $b_n/a_n' \to \infty$, where $\mathbb{P}(|Y_0| > a_n') \sim n^{-1}$ as $n \to \infty$. As a consequence of the tail behaviors of X_{it}^2 and $X_{it}X_{jt}$ for $i \neq j$ and Nagaev's results we have for (b_n) $b_n/a_n^2 \to \infty$,

  $$\frac{\mathbb{P}\big((\mathbf{X}\mathbf{X}')_{ij} > b_n\big)}{\mathbb{P}\big((\mathbf{X}\mathbf{X}')_{ii} - c_n > b_n\big)} \sim \frac{n\,\mathbb{P}(X_{11}X_{12} > b_n)}{n\,\mathbb{P}(X^2 > b_n)} \to 0\,, \qquad n \to \infty\,,$$

  where $c_n = 0$ or $n\,\mathbb{E}[X^2]$ according as $\alpha \in (0,2)$ or $\alpha \in (2,4)$. This means that the diagonal and off-diagonal entries of $\mathbf{X}\mathbf{X}'$ inherit the tails of X_{it}^2 and $X_{it}X_{jt}$, $i \neq j$, respectively, above the high threshold b_n.

Proposition 5.2 has some immediate consequences for the approximation of the eigenvalues of $\mathbf{X}\mathbf{X}'$ by those of $\operatorname{diag}(\mathbf{X}\mathbf{X}')$. Indeed, let C be a symmetric $p \times p$ matrix with eigenvalues $\lambda_1(C), \ldots, \lambda_p(C)$ and ordered eigenvalues

$$\lambda_{(1)}(C) \geq \cdots \geq \lambda_{(p)}(C)\,. \tag{5.12}$$

Then for any symmetric $p \times p$ matrices A, B, by *Weyl's inequality* (see Bhatia [20]),

$$\max_{i=1,\ldots,p} \big|\lambda_{(i)}(A+B) - \lambda_{(i)}(A)\big| \leq \|B\|_2\,.$$

If we now choose $A + B = \mathbf{X}\mathbf{X}'$ and $A = \operatorname{diag}(\mathbf{X}\mathbf{X}')$ we obtain the following result.

Corollary 5.3. *Under the conditions of Proposition 5.2,*

$$a_{np}^{-2} \max_{i=1,\ldots,p} \big|\lambda_{(i)} - \lambda_{(i)}(\operatorname{diag}(\mathbf{X}\mathbf{X}'))\big| \stackrel{\mathbb{P}}{\to} 0\,, \quad n \to \infty\,.$$

Thus the problem of deriving limit theory for the order statistics of $\mathbf{XX}'$ has been reduced to limit theory for the order statistics of the iid row-sums

$$D_i^{\to} = (\mathbf{XX}')_{ii} = \sum_{t=1}^{n} X_{it}^2 \,, \qquad i = 1, \ldots, p \,,$$

which are the eigenvalues of $\operatorname{diag}(\mathbf{XX}')$. This theory is completely described by the point processes constructed from the points $D_i^{\to}/a_{np}^2$ $i = 1, \ldots, p$. Necessary and sufficient conditions for the weak convergence of these point processes are provided by Lemma 5.22 which in combination with the Nagaev-type large deviation results of Theorem 5.21 yield the following result; see also Davis et al. [44].

Lemma 5.4. *Assume the conditions of Proposition 5.2 hold. Then*

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(D_i^{\to} - c_n)} \xrightarrow{\mathrm{d}} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}} \,, \qquad n \to \infty \,,$$

where (Γ_i) *is defined in* (5.9) *and* $c_n = 0$ *if* $\mathbb{E}[D^{\to}] = \infty$ *and* $c_n = \mathbb{E}[D^{\to}]$ *otherwise.*

In this result, centering is only needed for $\alpha \in [2, 4)$ when $n/a_{np}^2 \not\to 0$. Under the additional condition $\widetilde{C}_\beta(\alpha)$, $n/a_{np}^2 \to 0$ in view of the Potter bounds; see Bingham et al. [21, p. 25]. Combining Lemma 5.4 and Corollary 5.3, we conclude that Theorem 5.1 holds for $\beta \in [0, 1]$.

Extension to general β

Next we explain that it suffices to consider only the case $\beta \in [0, 1]$ and how to proceed when $\beta > 1$. The main reason is that the $p \times p$ sample covariance matrix $\mathbf{XX}'$ and the $n \times n$ matrix $\mathbf{X}'\mathbf{X}$ have the same rank and their non-zero eigenvalues coincide; see Bhatia [20, p. 64]. When proving limit theory for the eigenvalues of the sample covariance matrix one may switch to $\mathbf{X}'\mathbf{X}$ and vice versa, hereby interchanging the roles of p and n. By switching to $\mathbf{X}'\mathbf{X}$, one basically replaces β by β^{-1}. Since $\min(\beta, \beta^{-1}) \in [0, 1]$ for any $\beta \geq 0$, one can assume without loss of generality that $\beta \in [0, 1]$. This trick allows one to extend results for (p_n) satisfying $C_p(\beta)$ with $\beta \in [0, 1]$ to $\beta > 1$. We illustrate this approach by providing the direct analogs of Proposition 5.2 and Corollary 5.3.

Proposition 5.5. *Assume that* $\mathbf{X} = \mathbf{X}_n$ *has iid entries satisfying the regular variation condition* (5.7) *for some* $\alpha \in (0, 4)$*. If* $\mathbb{E}[|X|] < \infty$ *we also suppose that* $\mathbb{E}[X] = 0$*. Then for any sequence* (p_n) *satisfying* $C_p(\beta)$ *with* $\beta > 1$ *we have*

$$a_{np}^{-2} \|\mathbf{X}'\mathbf{X} - \operatorname{diag}(\mathbf{X}'\mathbf{X})\|_2 \xrightarrow{\mathbb{P}} 0 \,, \qquad n \to \infty \,,$$

where $\|\cdot\|_2$ *denotes the spectral norm.*

Note that for $\beta > 1$ we have $\lim_{n\to\infty} p/n = \infty$. This means that $\mathbf{X}'\mathbf{X}$ has asymptotically a much smaller dimension than $\mathbf{XX}'$ and therefore it is more convenient to work with $\mathbf{X}'\mathbf{X}$ when bounding the spectral norm.

Corollary 5.6. *Under the conditions of Proposition 5.5,*

$$a_{np}^{-2} \max_{i=1,\ldots,n} \big| \lambda_{(i)} - \lambda_{(i)}(\operatorname{diag}(\mathbf{X}'\mathbf{X})) \big| \xrightarrow{\mathbb{P}} 0 \,, \quad n \to \infty \,.$$

Now, Theorem 5.1 for $\beta > 1$ is a consequence of Corollary 5.6.

5.3 Introducing dependence between the rows and columns

For details on the results of this section, we refer to Davis et al. [44], Heiny and Mikosch [68] and Heiny et al. [69].

5.3.1 The model

When dealing with covariance matrices of a multivariate time series $(\mathbf{X}_n)$ it is rather natural to assume dependence between the entries X_{it}. In this section we introduce a model which allows for *linear dependence* between the rows and columns of $\mathbf{X}$:

$$X_{it} = \sum_{l\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} h_{kl} Z_{i-k,t-l}\,, \qquad i,t \in \mathbb{Z}\,, \tag{5.13}$$

where $(Z_{it})_{i,t\in\mathbb{Z}}$ is a field of iid random variables and $(h_{kl})_{k,l\in\mathbb{Z}}$ is an array of real numbers. Of course, linear dependence is restrictive in some sense. However, the particular dependence structure allows one to determine those ingredients in the sample covariance matrix which contribute to its largest eigenvalues. If the series in (5.13) converges a.s. (X_{it}) constitutes a strictly stationary random field. We denote generic elements of the Z- and X-fields by Z and X, respectively. We assume that Z is regularly varying in the sense that

$$\mathbb{P}(Z > x) \sim p_+ \frac{L(x)}{x^\alpha} \quad \text{and} \quad \mathbb{P}(Z \le -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \tag{5.14}$$

for some tail index $\alpha > 0$, constants $p_+, p_- \ge 0$ with $p_+ + p_- = 1$ and a slowly varying L. We will assume $\mathbb{E}[Z] = 0$ whenever $\mathbb{E}[Z^2] < \infty$. Moreover, we require the summability condition

$$\sum_{l\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} |h_{kl}|^\delta < \infty \tag{5.15}$$

for some $\delta \in (0, \min(\alpha/2, 1))$ which ensures the a.s. absolute convergence of the series in (5.13). Under the conditions (5.14) and (5.15), the marginal and finite-dimensional distributions of the field (X_{it}) are regularly varying with index α; see Embrechts et al. [58], Appendix A3.3. Therefore we also refer to (X_{it}) and (Z_{it}) as regularly varying fields.

The model (5.13) was introduced by Davis et al. [45], assuming the rows iid, and in the present form by Davis et al. [44].

5.3.2 Sample covariance and autocovariance matrices

From the field (X_{it}) we construct the $p \times n$ matrices

$$\mathbf{X}_n(s) = (X_{i,t+s})_{i=1,\dots,p;t=1,\dots,n}\,, \qquad s = 0, 1, 2, \dots\,, \tag{5.16}$$

As before, we will write $\mathbf{X} = \mathbf{X}_n(0)$. Now we can introduce the (non-normalized) *sample autocovariance matrices*

$$\mathbf{X}_n(0)\mathbf{X}_n(s)'\,, \qquad s = 0, 1, 2, \dots\,. \tag{5.17}$$

We will refer to s as the *lag*. For $s = 0$, we obtain the *sample covariance matrix*. In what follows, we will be interested in the asymptotic behavior (of functions) of the eigen- and singular values of the sample covariance and autocovariance matrices in the heavy-tailed case. Recall that the *singular values* of a matrix A are the square roots of the eigenvalues of the non-negative definite matrix AA' and its *spectral norm* $\|A\|_2$ is its largest singular value. We notice that $\mathbf{X}_n(0)\mathbf{X}_n(s)'$ is not symmetric and therefore its eigenvalues can be complex. To avoid this situation, we use the squares

$$\mathbf{X}_n(0)\mathbf{X}_n(s)'\mathbf{X}_n(s)\mathbf{X}_n(0)' \tag{5.18}$$

whose eigenvalues are the squares of the singular values of $\mathbf{X}_n(0)\mathbf{X}_n(s)'$. The idea of using the sample autocovariance matrices and functions of their squares (5.18) originates from a paper by Lam and Yao [86] who used a model different from (5.13). This idea is quite natural in the context of time series analysis.

In Theorem 5.7 below, we provide a general approximation result for the ordered singular values of the sample autocovariance matrices in the heavy-tailed case. This result is rather technical. To formulate it we introduce further notation. As before, $p = p_n$ is any integer sequence converging to infinity.

5.3.3 More notation

Important roles are played by the quantities $(Z_{it}^2)_{i=1,\dots,p;t=1,\dots,n}$ and their order statistics

$$Z_{(1),np}^2 \geq Z_{(2),np}^2 \geq \dots \geq Z_{(np),np}^2, \qquad n, p \geq 1\,. \tag{5.19}$$

As important are the row-sums

$$D_i^{\rightarrow} = D_i^{(n),\rightarrow} = \sum_{t=1}^n Z_{it}^2\,, \qquad i = 1,\dots,p; \quad n = 1,2,\dots\,, \tag{5.20}$$

with generic element $D^{\rightarrow}$ and their ordered values

$$D_{(1)}^{\rightarrow} = D_{L_1}^{\rightarrow} \geq \dots \geq D_{(p)}^{\rightarrow} = D_{L_p}^{\rightarrow}\,, \tag{5.21}$$

where we assume without loss of generality that $(L_1, \dots, L_p)$ is a permutation of $(1, \dots, p)$ for fixed n.

Finally, we introduce the column-sums

$$D_t^{\downarrow} = D_t^{(n),\downarrow} = \sum_{i=1}^p Z_{it}^2\,, \qquad t = 1,\dots,n; \quad p = 1,2,\dots\,,$$

with generic element $D^{\downarrow}$ and we also adapt the notation from (5.21) to these quantities.

Matrix norms

For any $p \times n$ matrix $\mathbf{A} = (a_{ij})$, we will use the following norms:

- *Spectral norm:*

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{(1)}(\mathbf{A}\mathbf{A}')}\,, \tag{5.22}$$

- *Frobenius norm:*

$$\|\mathbf{A}\|_F = \Big(\sum_{i=1}^p \sum_{j=1}^n |a_{ij}|^2\Big)^{1/2}\,.$$

We will frequently make use of the bound $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$. Standard references for matrix norms are [17, 20, 73, 102].

Singular values of the sample autocovariance matrices

Fix integers $n \ge 1$ and $s \ge 0$. We recycle the λ-notation for the singular values $\lambda_1(s), \ldots, \lambda_p(s)$ of the sample autocovariance matrix $\mathbf{X}_n(0)\mathbf{X}_n(s)'$, suppressing the dependence on n. Correspondingly, the order statistics are denoted by

$$\lambda_{(1)}(s) \ge \cdots \ge \lambda_{(p)}(s)\,. \tag{5.23}$$

When $s = 0$ we typically write λ_i instead of $\lambda_i(0)$.

The matrix $\mathbf{M}(s)$

We introduce some auxiliary matrices derived from the coefficient matrix $\mathbf{H} = (h_{kl})_{k,l\in\mathbb{Z}}$:

$$\mathbf{H}(s) = (h_{k,l+s})_{k,l\in\mathbb{Z}}, \qquad \mathbf{M}(s) = \mathbf{H}(0)\mathbf{H}(s)' \qquad s \ge 0\,.$$

Notice that

$$(\mathbf{M}(s))_{ij} = \sum_{l\in\mathbb{Z}} h_{i,l} h_{j,l+s}, \qquad i,j \in \mathbb{Z}. \tag{5.24}$$

We denote the ordered singular values of $\mathbf{M}(s)$ by

$$v_1(s) \ge v_2(s) \ge \cdots\,. \tag{5.25}$$

Let $r(s)$ be the rank of $\mathbf{M}(s)$ so that $v_{r(s)}(s) > 0$ while $v_{r(s)+1}(s) = 0$ if $r(s)$ is finite, otherwise $v_i(s) > 0$ for all i. We also write $r = r(0)$.

Under the summability condition (5.15) on (h_{kl}) for fixed $s \ge 0$,

$$\begin{aligned}\sum_{i=1}^{\infty}(v_i(s))^2 &= \|\mathbf{M}(s)\|_F^2 = \sum_{i,j\in\mathbb{Z}} \sum_{l_1,l_2\in\mathbb{Z}} h_{i,l_1} h_{j,l_1+s} h_{i,l_2} h_{j,l_2+s} \\ &\le c\Big(\sum_{l_1,l_2\in\mathbb{Z}} \sum_{i\in\mathbb{Z}} |h_{i,l_1} h_{i,l_2}|\Big)^2 \le c \sum_{l_1\in\mathbb{Z}} \sum_{i\in\mathbb{Z}} |h_{i,l_1}| < \infty\,.\end{aligned} \tag{5.26}$$

Therefore all singular values $v_i(s)$ are finite and the ordering (5.25) is justified.

Here and in what follows, we write c for any constant whose value is not of interest.

Normalizing sequence

We define (a_k) by

$$\mathbb{P}(|Z| > a_k) \sim k^{-1}\,, \qquad k \to \infty\,,$$

and choose the normalizing sequence for the singular values as (a_{np}^2) for suitable sequences $p = p_n \to \infty$.

Approximations to singular values

We will give approximations to the singular values $\lambda_i(s)$ in terms of the p largest ordered values for $s \ge 0$,

$$\begin{aligned}&\delta_{(1)}(s) \ge \cdots \ge \delta_{(p)}(s)\,,\\ &\gamma^{\rightarrow}_{(1)}(s) \ge \cdots \ge \gamma^{\rightarrow}_{(p)}(s)\,,\\ &\gamma^{\downarrow}_{(1)}(s) \ge \cdots \ge \gamma^{\downarrow}_{(n)}(s)\,,\end{aligned}$$

from the sets

$$\begin{aligned}
&\big\{Z^2_{(i),np} v_j(s)\,, i=1,\ldots,p\,; j=1,2,\ldots\big\}\,,\\
&\big\{D_i^{\rightarrow} v_j(s), i=1,\ldots,p\,; j=1,2,\ldots\big\}\,,\\
&\big\{D_t^{\downarrow} v_j(s), t=1,\ldots,n\,; j=1,2,\ldots\big\}\,,
\end{aligned}$$

respectively.

5.3.4 Approximation of the singular values

In the following result we povide some useful approximations to the singular values of the sample autocovariance matrices of the linear model (5.13).

Theorem 5.7. *Consider the linear process* (5.13) *under*

- *the regular variation condition* (5.14) *for some* $\alpha \in (0,4)$*,*
- *the centering condition* $\mathbb{E}[Z]=0$ *if* $\mathbb{E}[|Z|]<\infty$*,*
- *the summability condition* (5.15) *on the coefficient matrix* (h_{kl})*,*
- *the growth condition* $C_p(\beta)$ *on* (p_n) *for some* $\beta \ge 0$*.*

Then the following statements hold for $s \ge 0$*:*

1. *We consider two disjoint cases:* $\alpha \in (0,2)$ *and* $\beta \in (0,\infty)$*, or* $\alpha \in [2,4)$ *and* β *satisfying* $\widetilde{C}_\beta(\alpha)$*. Then*

$$a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)| \stackrel{\mathbb{P}}{\rightarrow} 0, \quad n \to \infty. \tag{5.27}$$

2. *Assume* $\beta \in [0,1]$*. If* $\alpha \in (0,2]$*,* $\mathbb{E}[Z^2]=\infty$ *or* $\alpha \in [2,4)$*,* $\mathbb{E}[Z^2]<\infty$ *and* $\beta \in (\alpha/2-1,1]$ *then*

$$a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \gamma^{\rightarrow}_{(i)}(s)| \stackrel{\mathbb{P}}{\rightarrow} 0, \quad n \to \infty.$$

Assume $\beta > 1$*. If* $\alpha \in (0,2]$*,* $\mathbb{E}[Z^2]=\infty$ *or* $\alpha \in [2,4)$*,* $\mathbb{E}[Z^2]<\infty$ *and* $\beta^{-1} \in (\alpha/2-1,1]$*. Then*

$$a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \gamma^{\downarrow}_{(i)}(s)| \stackrel{\mathbb{P}}{\rightarrow} 0, \quad n \to \infty.$$

Remark 5.8. The proof of Theorem 5.7 is given in Heiny et al. [69]. Part (2) of this result with more restrictive conditions on the growth rate of (p_n) is contained in Davis et al. [44]. These proofs are very technical and lengthy.

Remark 5.9. If we consider a random array (h_{kl}) independent of (X_{it}) and assume that the summability condition (5.15) holds a.s. then Theorem 5.7 remains valid conditionally on (h_{kl}), hence unconditionally in $\mathbb{P}$-probability; see also [44].

5.3.5 Point process convergence

Theorem 5.7 and arguments similar to the proofs in Davis et al. [44] enable one to derive the weak convergence of the point processes of the normalized singular values. Recall the representation of the points (Γ_i) of a unit rate homogeneous Poisson process on $(0,\infty)$ given in (5.9). For $s \ge 0$, we define the point processes of the normalized singular values:

$$N_n^{\lambda,s} = \sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(\lambda_{(i)}(0),\ldots,\lambda_{(i)}(s))}\,. \tag{5.28}$$

Theorem 5.10. *Assume the conditions of Theorem 5.7. Then $(N_n^{\lambda,s})$ converge weakly in the space of point measures with state space $(0,\infty)^{s+1}$ equipped with the vague topology. If either $\alpha \in (0,2]$, $\mathbb{E}[Z^2]=\infty$ and $\beta \ge 0$, or $\alpha \in [2,4)$, $\mathbb{E}[Z^2]<\infty$ and $\widetilde{C}_\beta(\alpha)$ hold then*

$$N_n^{\lambda,s} \xrightarrow{\rm d} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}(v_j(0),\ldots,v_j(s))}, \qquad n \to \infty. \tag{5.29}$$

Proof. Regular variation of Z^2 is equivalent to

$$n\,p\,\mathbb{P}(a_{np}^{-2}Z^2 \in \cdot) \xrightarrow{\rm v} \mu(\cdot), \tag{5.30}$$

where $\xrightarrow{\rm v}$ denotes vague convergence of Radon measures on $(0,\infty)$ and the measure μ is given by $\mu(x,\infty)=x^{-\alpha/2}$, $x>0$. In view of Resnick [96], Proposition 3.21, (5.30) is equivalent to the weak convergence of the following point processes:

$$\sum_{i=1}^{p}\sum_{t=1}^{n} \varepsilon_{a_{np}^{-2}Z_{it}^2} = \sum_{i=1}^{np} \varepsilon_{a_{np}^{-2}Z_{(i),np}^2} \xrightarrow{\rm d} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}} = \widetilde{N}\,, \quad n\to\infty\,,$$

where the limit $\widetilde{N}$ is a Poisson random measure (PRM) with state space $(0,\infty)$ and mean measure μ.

Since $a_{np}^{-2}Z_{(p),np}^2 \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$, the point processes $\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}Z_{(i),np}^2}$ converge weakly to the same PRM:

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}Z_{(i),np}^2} \xrightarrow{\rm d} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \quad n\to\infty\,. \tag{5.31}$$

A continuous mapping argument together with the fact that $\sum_{i=1}^{\infty}(v_i(s))^2<\infty$ (see (5.26)) shows that

$$\sum_{j=1}^{\infty}\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}Z_{(i),np}^2(v_j(0),\ldots,v_j(s))} \xrightarrow{\rm d} \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}(v_j(0),\ldots,v_j(s))}\,.$$

If the assumptions of part (1) of Theorem 5.7 are satisfied an application of (5.27) (also recalling the definition of $(\delta_{(i)}(s))$) shows that (5.31) remains valid with the points $(a_{np}^{-2}Z_{(i),np}^2(v_j(0),\ldots,v_j(s)))$ replaced by $(a_{np}^{-2}(\lambda_{(i)}(0),\ldots,\lambda_{(i)}(s))$.

The only cases which are not covered by Theorem 5.7(1) are $\alpha\in(0,2)$, $\beta=0$ and $\alpha=2$, $\mathbb{E}[Z^2]=\infty$, $\beta\ge 0$. In these cases we get from Theorem 5.21 that

$$p\,\mathbb{P}(a_{np}^{-2}D^{\rightarrow}>x) \sim p\,n\,\mathbb{P}(Z^2>a_{np}^2x) \to \mu(x,\infty)\,, \quad x>0\,,$$

i.e., $p\,\mathbb{P}(a_{np}^{-2}D^{\rightarrow} \in \cdot) \stackrel{v}{\rightarrow} \mu(\cdot)$. It follows from Lemma 5.22 that $\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}D_i^{\rightarrow}} \stackrel{\mathrm{d}}{\rightarrow} \widetilde{N}$. As before, a continuous mapping argument in combination with the approximation obtained in Theorem 5.7(2) justifies the replacement of the points $(a_{np}^{-2}D_{(i)}^{\rightarrow}(v_j(0),\ldots,v_j(s)))$ by $(a_{np}^{-2}(\lambda_{(i)}(0),\ldots,\lambda_{(i)}(s)))$ in the case $\beta \in [0,1]$. If $\beta > 1$ one has to work with the quantities $(D_i^{\downarrow})_{i=1,\ldots,n}$ instead of $(D_i^{\rightarrow})_{i=1,\ldots,p}$ and one may follow the same argument as above. This finishes the proof. □

5.4 Some applications

5.4.1 Sample covariance matrices

The sample covariance matrix $\mathbf{X}_n(0)\mathbf{X}_n(0)' = \mathbf{X}\mathbf{X}'$ is a non-negative definite matrix. Therefore its eigenvalues and singular values coincide. Moreover, $v_j = v_j(0)$, $j \geq 1$, are the eigenvalues of $\mathbf{M} = \mathbf{M}(0)$.

Theorem 5.7(1) yields an approximation of the ordered eigenvalues $(\lambda_{(i)})$ of $\mathbf{X}\mathbf{X}'$ by the quantities $(\delta_{(i)})$ which are derived from the order statistics of (Z_{it}^2). Part (2) of this result provides an approximation of $(\lambda_{(i)})$ by the quantities $(\gamma_{(i)}^{\rightarrow/\downarrow})$ which are derived from the order statistics of the partial sums $(D_i^{\rightarrow/\downarrow})$.

In the following example we illustrate the quality of the two approximations.

Example 5.11. We choose a Pareto-type distribution for Z with density

$$f_Z(x) = \begin{cases} \frac{\alpha}{(4|x|)^{\alpha+1}}\,, & \text{if } |x| > 1/4 \\ 1\,, & \text{otherwise.} \end{cases} \tag{5.32}$$

We simulated 20,000 matrices $\mathbf{X}_n$ for $n = 1{,}000$ and $p = 200$ whose iid entries have this density. We assume $\beta = 1$. Note that $\mathbf{M} = \mathbf{M}(0)$ has rank one and $v_1 = 1$. The estimated densities of the deviations $a_{np}^{-2}(\lambda_{(1)} - D_{(1)}^{\rightarrow})$ and $a_{np}^{-2}(\lambda_{(1)} - Z_{(1),np}^2)$ based on the simulations are shown in Figure 5.4. The approximation error is very small indeed. According to the theory,

$$a_{np}^{-2}\sup_i |D_{(i)}^{\rightarrow} - \lambda_{(i)}| + a_{np}^{-2}\sup_i |Z_{(i),np}^2 - \lambda_{(i)}| \stackrel{\mathbb{P}}{\rightarrow} 0\,,$$

but for finite n the $(D_{(i)}^{\rightarrow})$ sequence yields a better approximation to $(\lambda_{(i)})$. By construction, the considered differences have a tendency to be positive but Figure 5.4 also shows that the median of the approximation error for $a_{np}^{-2}(\lambda_{(1)} - D_{(1)}^{\rightarrow})$ is almost zero.

Theorem 5.10 and the continuous mapping theorem immediately yield results about the joint convergence of the largest eigenvalues of the matrices $a_{np}^{-2}\mathbf{X}_n\mathbf{X}_n'$ for $\alpha \in (0,2)$ and $\alpha \in (2,4)$ when β satisfies $\widetilde{C}_\beta(\alpha)$. For fixed $k \geq 1$ one gets

$$a_{np}^{-2}\big(\lambda_{(1)},\ldots,\lambda_{(k)}\big) \stackrel{\mathrm{d}}{\rightarrow} \big(d_{(1)},\ldots,d_{(k)}\big)\,,$$

where $d_{(1)} \geq \cdots \geq d_{(k)}$ are the k largest ordered values of the set $\{\Gamma_i^{-2/\alpha}v_j, i = 1,2,\ldots, j = 1,\ldots,r\}$. The continuous mapping theorem yields for $k \geq 1$,

$$\frac{\lambda_{(1)}}{\lambda_{(1)} + \cdots + \lambda_{(k)}} \stackrel{\mathrm{d}}{\rightarrow} \frac{d_{(1)}}{d_{(1)} + \cdots + d_{(k)}}\,, \quad n \to \infty\,. \tag{5.33}$$

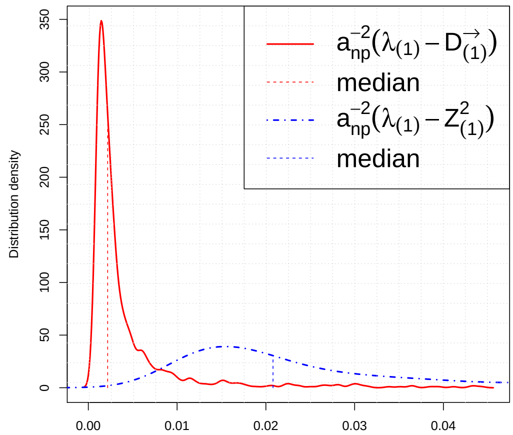


Figure 5.4: Density of the approximation errors for the eigenvalues of $a_{np}^{-2}\mathbf{X}\mathbf{X}'$. The entries of $\mathbf{X}$ are iid with density (5.32) and $\alpha = 1.6$.

An application of the continuous mapping theorem to the distributional convergence of the point processes in Theorem 5.10 in the spirit of Resnick [97], Theorem 7.1, also yields the following result; see Davis et al. [44] for a proof and a similar result in the case $\alpha \in (2, 4)$.

Corollary 5.12. *Assume the conditions of Theorem 5.7. If $\alpha \in (0, 2]$ and $\mathbb{E}[Z^2] = \infty$, then*

$$a_{np}^{-2}\Big(\lambda_{(1)}, \sum_{i=1}^{p} \lambda_i\Big) \xrightarrow{\rm d} \Big(v_1\, \Gamma_1^{-2/\alpha}, \sum_{j=1}^{r} v_j \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}\Big),$$

where $\Gamma_1^{-2/\alpha}$ is Fréchet $\Phi_{\alpha/2}$-distributed. and $\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}$ has the distribution of a positive $\alpha/2$-stable random variable. In particular,

$$\frac{\lambda_{(1)}}{\lambda_1 + \cdots + \lambda_p} \xrightarrow{\rm d} \frac{v_1}{\sum_{j=1}^{r} v_j} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}}, \qquad n \to \infty. \tag{5.34}$$

Remark 5.13. The ratio

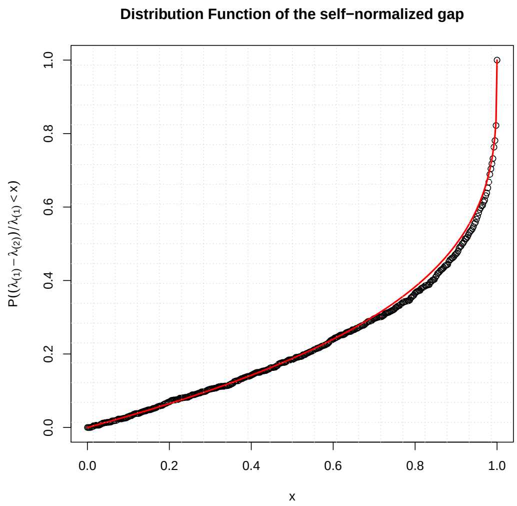
$$\frac{\lambda_{(1)} + \cdots + \lambda_{(k)}}{\lambda_1 + \cdots + \lambda_p}, \qquad k \ge 1,$$

plays an important role in PCA. It reflects the proportion of the total variance in the data that we can explain by the first k principal components. It follows from Corollary 5.12 that for fixed $k \ge 1$,

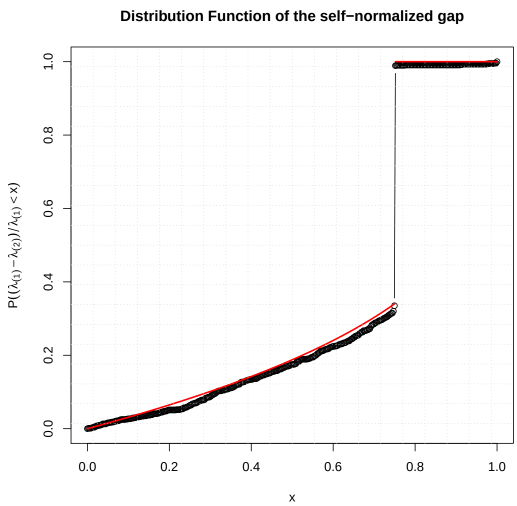
$$\frac{\lambda_{(1)} + \cdots \lambda_{(k)}}{\lambda_1 + \cdots + \lambda_p} \xrightarrow{\rm d} \frac{d_{(1)} + \cdots + d_{(k)}}{d_{(1)} + d_{(2)} + \cdots}.$$

Unfortunately, the limiting variable does in general not have a clean form. An exception is the case when $r = 1$; see Example 5.16. Also notice that the trace of $\mathbf{X}\mathbf{X}'$ coincides with $\lambda_1 + \cdots + \lambda_p$.

To illustrate the theory we consider a simple moving average example taken from Davis et al. [44].



(a) iid data



(b) data from model (5.35)

Figure 5.5: Distribution function of $(\lambda_{(1)} - \lambda_{(2)})/\lambda_{(1)}$ for iid data (left) and data generated from the model (5.35) (right). In each graph we compare the empirical distribution function (dotted line, based on 1000 simulations of 200×1000 matrices with Z-distribution (5.32)) with the theoretical curve (solid line).

Example 5.14. Assume that $\alpha \in (0, 2)$ and

$$X_{it} = Z_{it} + Z_{i,t-1} - 2(Z_{i-1,t} - Z_{i-1,t-1}), \quad i, t \in \mathbb{Z}. \tag{5.35}$$

In this case, the non-zero entries of $\mathbf{H}$ are

$$h_{00}=1, h_{01}=1, h_{10}=-2 \quad \text{and} \quad h_{11}=2.$$

Hence $\mathbf{M}=\mathbf{H}\mathbf{H}'$ has the positive eigenvalues $v_1=8$ and $v_2=2$. The limit point process in (5.29) is

$$N=\sum_{i=1}^{\infty}\varepsilon_{8\Gamma_i^{-2/\alpha}}+\sum_{i=1}^{\infty}\varepsilon_{2\Gamma_i^{-2/\alpha}}\,,$$

so that

$$a_{np}^{-2}\big(\lambda_{(1)},\lambda_{(2)}\big)\stackrel{\rm d}{\to}\big(8\Gamma_1^{-2/\alpha},2\Gamma_1^{-2/\alpha}\vee 8\Gamma_2^{-2/\alpha}\big)\,.$$

Using the fact that $U=\Gamma_1/\Gamma_2$ has a uniform distribution on $(0,1)$ we calculate

$$\mathbb{P}(2\Gamma_1^{-2/\alpha}>8\Gamma_2^{-2/\alpha})=\mathbb{P}(\Gamma_1/\Gamma_2<2^{-\alpha})=2^{-\alpha}\in(1/4,1).$$

In particular, we have for the normalized spectral gap

$$a_{np}^{-2}\big(\lambda_{(1)}-\lambda_{(2)}\big)\stackrel{\rm d}{\to}6\,\Gamma_1^{-2/\alpha}\mathbf{1}_{\{\Gamma_1 4^{\alpha/2}<\Gamma_2\}}+8\,\big(\Gamma_1^{-2/\alpha}-\Gamma_2^{-2/\alpha}\big)\mathbf{1}_{\{\Gamma_1 4^{\alpha/2}>\Gamma_2\}}$$

and for the self-normalized spectral gap (see also Example 5.15 for a detailed analysis)

$$\begin{aligned}\frac{\lambda_{(1)}-\lambda_{(2)}}{\lambda_{(1)}}&\stackrel{\rm d}{\to}\frac{6}{8}\,\mathbf{1}_{\{\Gamma_1 2^{\alpha}<\Gamma_2\}}+\big(1-(\Gamma_1/\Gamma_2)^{2/\alpha}\big)\mathbf{1}_{\{\Gamma_1 2^{\alpha}>\Gamma_2\}}\\&=\frac{3}{4}\,\mathbf{1}_{\{U2^{\alpha}<1\}}+\big(1-U^{2/\alpha}\big)\mathbf{1}_{\{U2^{\alpha}>1\}}=Y\,.\end{aligned}$$

The limit distribution of the spectral gap has an atom at $3/4$ with probability $2^{-\alpha}$, i.e., $\mathbb{P}(Y=3/4)=2^{-\alpha}$, and

$$\mathbb{P}(Y\le x)=1-(1-x)^{\alpha/2},\qquad x\in(0,3/4).$$

In the iid case the limit distribution of the self-normalized spectral gap has distribution function $F(x)=1-(1-x)^{\alpha/2}$ for $x\in[0,1]$. This means that the atom disappears if the entries are iid. Figure 5.5 compares the distribution function of Y with F for $\alpha=0.6$; the atom at $3/4$ is clearly visible.

Along the same lines, we also have

$$(a_{np}^{-2}\lambda_{(1)},\lambda_{(2)}/\lambda_{(1)})\stackrel{\rm d}{\to}(8\,\Gamma_1^{-2/\alpha},\frac{1}{4}\,\mathbf{1}_{\{U<2^{-\alpha}\}}+U^{2/\alpha}\,\mathbf{1}_{\{U\ge 2^{-\alpha}\}})$$

and hence the limit distribution of $\lambda_{(2)}/\lambda_{(1)}$ is supported on $[1/4,1)$ with mass of $2^{-\alpha}$ at $1/4$. The histogram of the ratio $\big(\lambda_{(2)}/\lambda_{(1)}\big)^{2/\alpha}$ based on 1000 replications from the model (5.35) with noise given by a t-distribution with $\alpha=1.5$ degrees of freedom, $n=1000$ and $p=200$ is displayed in Figure 5.6. Observing that $2^{-\alpha}=0.3536\ldots$, the histogram is remarkably close to what one would expect from a sample from the truncated uniform distribution, $2^{-\alpha}\,\mathbf{1}_{\{U<2^{-\alpha}\}}+U\,\mathbf{1}_{\{U\ge 2^{-\alpha}\}}$. The mass of the limiting discrete component of the ratio can be much larger if one conditions on $a_{np}^{-2}\lambda_{(1)}$ being large. Specifically, for any $\epsilon\in(0,1/4)$ and $x>0$,

$$\lim_{n\to\infty}\mathbb{P}(\epsilon<\lambda_{(2)}/\lambda_{(1)}\le 1/4|\lambda_{(1)}>a_{np}^2x)=\mathbb{P}(\Gamma_1/\Gamma_2\le 2^{-\alpha}|\Gamma_1<(x/8)^{-\alpha/2})=G(x)\,.$$

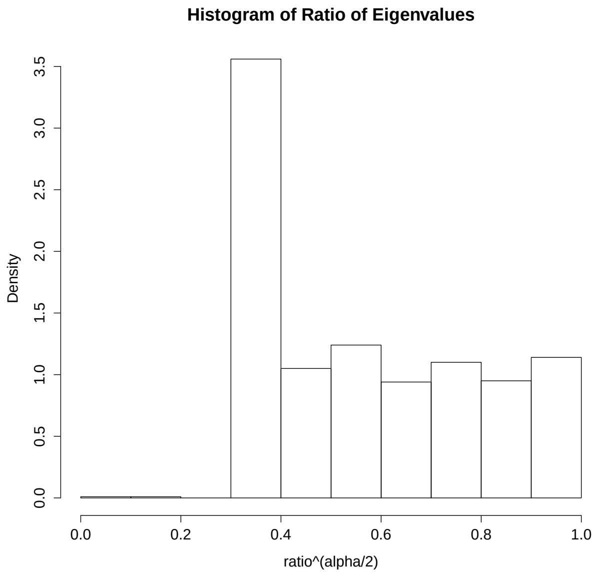


Figure 5.6: Histogram based on 1000 replications of $\big(\lambda_{(2)}/\lambda_{(1)}\big)^{2/\alpha}$ from model (5.35).

The function G approaches 1 as $x \to \infty$ indicating the speed at which the two largest eigenvalues get linearly related; see Figure 5.7 for a graph of G in the case $\alpha = 1.5$. In addition, from Remark 5.13, we also have

$$\frac{\lambda_{(1)}}{\lambda_1 + \cdots + \lambda_p} \xrightarrow{\mathrm{d}} \frac{4}{5} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}} .$$

Clearly, the limit random variable is stochastically smaller than what one would get in the iid case; see (5.34).

Example 5.15. The previous example also illustrates the behavior of the two largest eigenvalues in the general case when the rank r of the matrix $\mathbf{M}$ is larger than one. We have in general

$$\frac{\lambda_{(2)}}{\lambda_{(1)}} \xrightarrow{\mathrm{d}} \frac{v_2}{v_1} \mathbf{1}_{\{U<(v_2/v_1)^{\alpha/2}\}} + U^{2/\alpha} \mathbf{1}_{\{U\ge(v_2/v_1)^{\alpha/2}\}} .$$

In particular, the limiting *self-normalized spectral gap* has representation

$$\frac{\lambda_{(1)} - \lambda_{(2)}}{\lambda_{(1)}} \xrightarrow{\mathrm{d}} \frac{v_1 - v_2}{v_1} \mathbf{1}_{\{U<(v_2/v_1)^{\alpha/2}\}} + (1 - U^{2/\alpha}) \mathbf{1}_{\{U\ge(v_2/v_1)^{\alpha/2}\}} .$$

The limiting variable assumes values in $(0, 1 - v_2/v_1]$ and has an atom at the right endpoint. This is in contrast to the iid case and to the case when $r = 1$ (hence $v_2 = 0$) including the case of iid rows and the separable case; see Example 5.16.

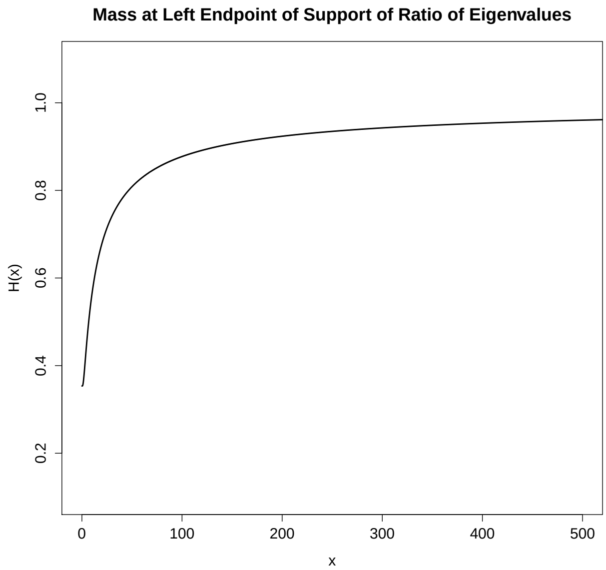


Figure 5.7: Graph of $G(x) = \mathbb{P}(\Gamma_1/\Gamma_2 \le 2^{-\alpha} | \Gamma_1 < (x/8)^{-\alpha/2})$ when $\alpha = 1.5$.

Example 5.16. We consider the separable case when $h_{kl} = \theta_k c_l$, $k, l \in \mathbb{Z}$, where (c_l), (θ_k) are real sequences such that the conditions on (h_{kl}) in Theorem 5.7 hold. In this case,

$$\mathbf{M} = \sum_{l \in \mathbb{Z}} c_l^2 \ \ (\theta_i \theta_j)_{i,j \in \mathbb{Z}} \,.$$

Note that $r = 1$ with the only non-negative eigenvalue

$$v_1 = \sum_{l \in \mathbb{Z}} c_l^2 \ \sum_{k \in \mathbb{Z}} \theta_k^2 \,.$$

In this case, the limiting point process in Theorem 5.10 is a PRM on $(0, \infty)$ with mean measure of (y, ∞) given by $(v_1/y)^{\alpha/2}$, $y > 0$. The normalized eigenvalues have similar asymptotic behavior as in the case of iid entries. For example, the log-spacings have the same limit as in the iid case for fixed k,

$$\big(\log \lambda_{(1)} - \log \lambda_{(2)}, \ldots, \log \lambda_{(k+1)} - \log \lambda_{(k)}\big) \stackrel{\mathrm{d}}{\to} -\frac{2}{\alpha} \big(\log(\Gamma_1/\Gamma_2), \ldots, \log(\Gamma_k/\Gamma_{k+1})\big) \,.$$

The same observation applies to the ratio of the largest eigenvalue and the trace in the case $\alpha \in (0, 2)$:

$$\frac{\lambda_{(1)}}{\operatorname{tr}(\mathbf{X}\mathbf{X}')} = \frac{\lambda_{(1)}}{\lambda_1 + \cdots + \lambda_p} \stackrel{\mathrm{d}}{\to} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}} \,.$$

We also mentioned in Example 5.15 that the distributional limit of the self-normalized spectral gap has no atom as in the iid case.

5.4.2 S&P 500 data

We conduct a short analysis of the largest eigenvalues of the univariate log-return time series which compose the S&P 500 stock index; see Section 5.1.2 for a description of the data. Although there is strong empirical evidence that these univariate series have power-law tails (see Figure 5.3) we do not expect that they have the same tail index. One way to proceed would be to ignore this fact because the tail indices are in a close range and the differences are due to large sampling errors for estimating such quantities. One could also collect time series with similar tail indices in the same group. In this case, the dimension p decreases. This grouping would be a rather arbitrary classification method. We have chosen a third way: to use rank transforms. This approach has its merits because it aims at standardizing the tails but it also has a major disadvantage: one destroys the covariance structure underlying the data.

Given a $p \times n$ matrix $(R_{it})_{i=1,\cdots,p;t=1,\cdots,n}$, we construct a matrix $\mathbf{X}$ via the rank transforms

$$X_{it} = -\Big[\log\Big(\frac{1}{n+1}\sum_{\tau=1}^{n}\mathbf{1}_{\{R_{i\tau}\le R_{it}\}}\Big)\Big]^{-1}, \qquad i=1,\ldots,p;t=1,\ldots,n\,.$$

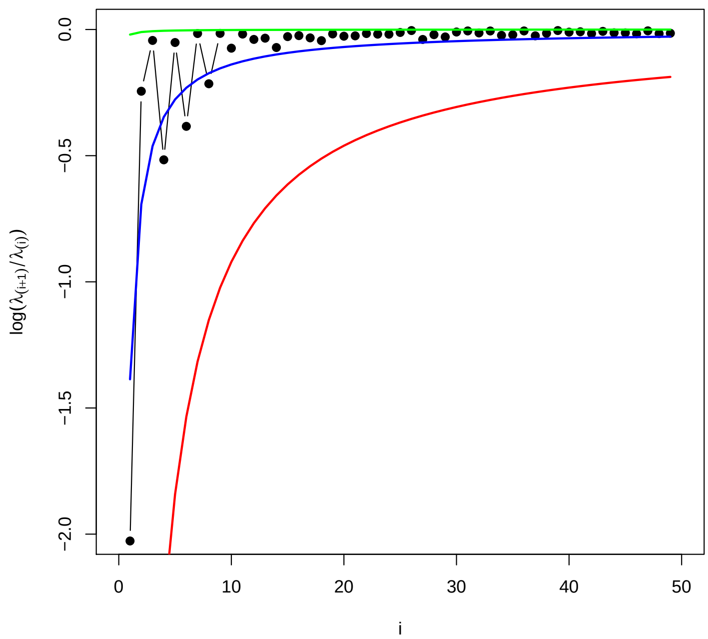


Figure 5.8: The logarithms of the ratios $\lambda_{(i+1)}/\lambda_{(i)}$ for the S&P 500 series after rank transform. We also show the 1, 50 and 99% quantiles (bottom, middle, top lines, respectively) of the variables $\log((\Gamma_i/\Gamma_{i+1})^2)$.

If the rows $R_{i1},\ldots,R_{in}$ were iid (or, more generally, stationary ergodic) with a continuous distribution then the averages under the logarithm would be asymptotically uniform

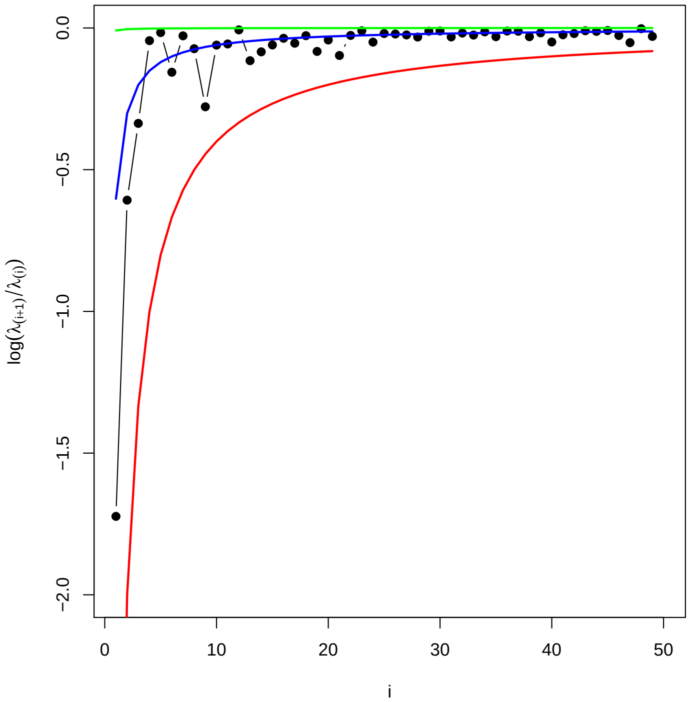


Figure 5.9: The logarithms of the ratios $\lambda_{(i+1)}/\lambda_{(i)}$ for the original (non-rank transformed) S&P 500 log-return data. We also show the 1, 50 and 99% quantiles (bottom, middle, top lines, respectively) of the variables $\log((\Gamma_i/\Gamma_{i+1})^{2/2.3})$; see also Figure 5.8 for comparison.

on $(0,1)$ as $n \to \infty$. Hence X_{it} would be asymptotically standard Fréchet Φ_1-distributed. In what follows, we assume that the aforementioned univariate time series of the S&P 500 index have undergone the rank transform and that their marginal distributions are close to Φ_1; we always use the symbol $\mathbf{X}$ for the resulting multivariate series.

In Figure 5.8 we show the ratios of the consecutive ordered eigenvalues $(\lambda_{(i+1)}/\lambda_{(i)})$ of the matrix $\mathbf{X}\mathbf{X}'$. This graph shows the rather surprising fact that the ratios are close to one even for small values i. We also show the 1, 50 and 99 % quantiles of the variables $((\Gamma_i/\Gamma_{i+1})^{2/\alpha})$ calculated from the formula

$$\mathbb{P}\big((\Gamma_i/\Gamma_{i+1})^{2/\alpha} \le x\big) = x^{i\cdot\alpha/2}, \quad x \in (0,1)\,. \tag{5.36}$$

For increasing i, the distribution is concentrated closely to 1, in agreement with the strong law of large numbers which yields $\Gamma_i/\Gamma_{i+1} \stackrel{\text{a.s.}}{\to} 1$ as $i \to \infty$. The asymptotic distributions (5.36) correspond to the case when the matrix $\mathbf{M}$ has rank $r = 1$. It includes the iid and separable cases; see Example 5.16. The shown asymptotic quantiles are in agreement with the rank $r = 1$ hypothesis.

For comparison, in Figure 5.9 we also show the ratios $(\lambda_{(i+1)}/\lambda_{(i)})$ for the non-rank transformed S&P 500 data and the 1, 50 and 99% quantiles of the variables $\log((\Gamma_i/\Gamma_{i+1})^{2/\alpha})$, where we choose $\alpha = 2.3$ motivated by the estimated tail indices in Figure 5.3. The two graphs in Figure 5.8 and Figure 5.9 are quite similar but the smallest ratios for the original data are slightly larger than for the rank-transformed data.

5.4.3 Sums of squares of sample autocovariance matrices

In this section we consider some additive functions of the squares of $\mathbf{A}_n(s) = \mathbf{X}_n(0)\mathbf{X}_n(s)'$ given by $\mathbf{A}_n(s)\mathbf{A}_n(s)'$ for $s = 0, 1, \ldots$. By definition of the singular values of a matrix (see (5.23)), the non-negative definite matrix $\mathbf{A}_n(s)\mathbf{A}_n(s)'$ has eigenvalues $(\lambda_i^2(s))_{i=1,\ldots,p}$.

The following result is a corollary of Theorem 5.7.

Proposition 5.17. *Consider the linear process* (5.13) *under the conditions of Theorem* 5.7*. Then the following statements hold for* $s \geq 0$*:*

(1) *We consider two disjoint cases:* $\alpha \in (0,2)$ *and* $\beta \in (0,\infty)$*, or* $\alpha \in [2,4)$ *and* β *satisfying* $\widetilde{C}_\beta(\alpha)$*. Then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - \delta_{(i)}^2(s)| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

(2) *Assume* $\beta \in [0,1]$*. If* $\alpha \in (0,2]$*,* $\mathbb{E}[Z^2] = \infty$ *or* $\alpha \in [2,4)$*,* $\mathbb{E}[Z^2] < \infty$ *and* $\beta \in (\alpha/2 - 1, 1]$*, then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\rightarrow}(s))^2| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

Assume $\beta > 1$*. If* $\alpha \in (0,2]$*,* $\mathbb{E}[Z^2] = \infty$ *or* $\alpha \in [2,4)$*,* $\mathbb{E}[Z^2] < \infty$ *and* $\beta^{-1} \in (\alpha/2 - 1, 1]$*. Then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\downarrow}(s))^2| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

To the best of our knowledge, sums of squares of sample autocovariance matrices were used first in the paper by Lam and Yao [86]; their time series model is quite different from ours.

Proof. Part (1). The proof follows from Theorem 5.7 if we can show that

$$a_{np}^{-2} \max_{i=1,\ldots,p} \big(\lambda_{(i)}(s) + \delta_{(i)}(s)\big) = O_{\mathbb{P}}(1) \quad n \to \infty\,.$$

We have by Theorem 5.10,

$$a_{np}^{-2} \max_{i=1\ldots,p} \lambda_{(i)}(s) = a_{np}^{-2}\lambda_{(1)}(s) \xrightarrow{\text{d}} c\,\xi_{\alpha/2}\,, \tag{5.37}$$

where $\xi_{\alpha/2}$ has a $\Phi_{\alpha/2}$ distribution. In view of Theorem 5.7(1) we also have

$$a_{np}^{-2} \max_{i=1\ldots,p} \delta_{(i)}(s) \xrightarrow{\text{d}} c\,\xi_{\alpha/2}\,.$$

Therefore, again using Theorem 5.7(1), we have

$$\begin{aligned} &a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - \delta_{(i)}^2(s)| \\ &\leq \Big[a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)|\Big] \Big[a_{np}^{-2} \max_{i=1,\ldots,p} \big(|\lambda_{(i)}(s)| + |\delta_{(i)}(s)|\big)\Big] \xrightarrow{\mathbb{P}} 0, \quad n \to \infty. \end{aligned}$$

This proves part (1).

Part (2). Now assume $\beta \in [0,1]$ and $\alpha \in (0,2]$, $\mathbb{E}[Z^2] = \infty$ or $\alpha \in [2,4)$, $\mathbb{E}[Z^2] < \infty$ and $\beta \in (\alpha/2-1,1]$. Then (5.37) is still true and we have by Theorem 5.7(2) and Theorem 5.10

$$a_{np}^{-2} \max_{i=1\ldots,p} \gamma_{(i)}^{\rightarrow}(s) \xrightarrow{\mathrm{d}} c\,\xi_{\alpha/2}\,.$$

We then have

$$\begin{aligned} & a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\rightarrow}(s))^2| \\ \leq\; & \big[a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \gamma_{(i)}^{\rightarrow}(s)|\big]\, \big[a_{np}^{-2} \max_{i=1,\ldots,p} \big(\lambda_{(i)}(s) + \gamma_{(i)}^{\rightarrow}(s)\big)\big] \xrightarrow{\mathbb{P}} 0\,, \qquad n\to\infty. \end{aligned}$$

The proof of the remaining part is similar and therefore omitted. □

Now, using Proposition 5.17 and a continuous mapping argument, we can show limit theory for the eigenvalues

$$w_{(1)}(s_0,s_1) \geq \cdots \geq w_{(p)}(s_0,s_1)\,, \qquad 0 \leq s_0 \leq s_1\,,$$

of the non-negative definite random matrices

$$\sum_{s=s_0}^{s_1} \mathbf{A}_n(s)\mathbf{A}_n(s)'\,. \tag{5.38}$$

Proposition 5.18. *Assume* $0 \leq s_0 \leq s_1$ *and the conditions of Theorem 5.7 hold. If* $\alpha \in (0,4)$ *and* $\beta \in (0,1] \cap (\alpha/2-1,1]$ *then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |w_{(i)}(s_0,s_1) - \omega_{(i)}(s_0,s_1)| \xrightarrow{\mathbb{P}} 0, \quad n\to\infty,$$

where $\omega_{(i)}(s_0,s_1)$ *are the ordered values of the set* $\{Z^4_{(i),np} v_j(s_0,s_1), i = 1,\ldots,p; j = 1,2,\ldots\}$ *and* $(v_j(s_0,s_1))$ *are the ordered eigenvalues of* $\sum_{s=s_0}^{s_1} \mathbf{M}(s)\mathbf{M}(s)'$.

Example 5.19. Recall the separable case from Example 5.16, i.e., $h_{kl} = \theta_k c_l$, $k,l \geq 0$, where (c_l), (θ_k) are real sequences such that the conditions on (h_{kl}) in Theorem 5.7 hold. Write $\Theta_{ij} = \theta_i\theta_j$. It is symmetric and has rank one; the only non-zero eigenvalue is $\gamma_\theta(0) = \sum_{k=0}^\infty \theta_k^2$. Hence Θ is non-negative definite. We get from (5.24) that

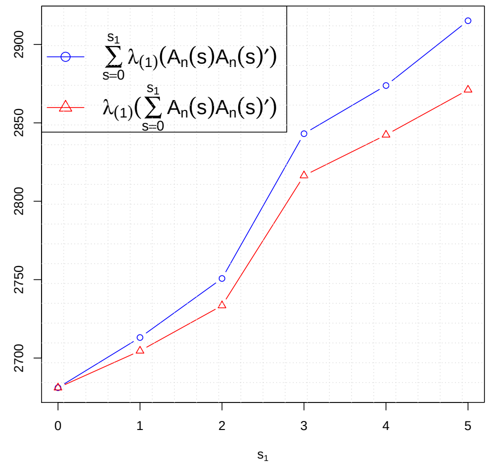
$$\mathbf{M}(s) = \gamma_c(s)\,\Theta, \quad s \geq 0\,,$$

where

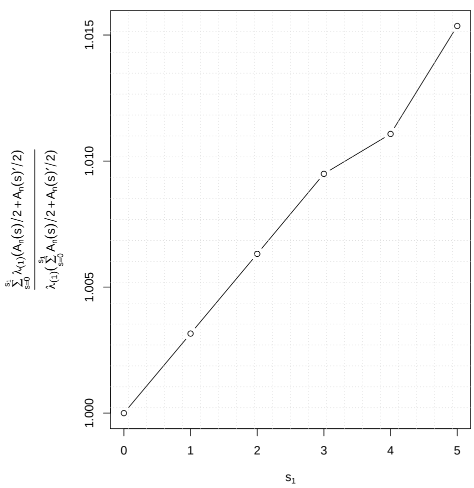
$$\gamma_c(s) = \sum_{l=0}^\infty c_l c_{l+s}\,, \qquad s \geq 0\,.$$

The matrix $\mathbf{M}(s)$ has the only non-zero eigenvalue $\gamma_c(s)\gamma_\theta(0)$. The factors $(\gamma_c(s))$ can be positive or negative; they constitute the autocovariance function of a stationary linear process with coefficients (c_l). Accordingly, $\mathbf{M}(s)$ is either non-negative or non-positive definite. Now we consider the non-negative definite matrix

$$\sum_{s=s_0}^{s_1} \mathbf{M}(s)\,\mathbf{M}(s)' = \sum_{s=s_0}^{s_1} \gamma_c^2(s)\,\Theta\Theta'\,.$$



(a)



(b)

Figure 5.10: The largest eigenvalues of the sums of the squared autocovariance matrices compared with the sums of the largest eigenvalues of these matrices for the S&P 500 data for different values s_1. The two values are surprisingly close to each other; mind the scale of the y-axis. We also show their ratios.

This matrix has rank 1 and its largest eigenvalue is given by

$$C_{c,\theta}(s_0, s_1) = \sum_{s=s_0}^{s_1} \gamma_c^2(s)\, \gamma_\theta^2(0)\,.$$

An application of Proposition 5.18 yields that the ordered eigenvalues of

$$a_{np}^{-4}\sum_{s=s_0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'$$

are uniformly approximated by the quantities

$$a_{np}^{-4}Z_{(i),np}^4 C_{c,\theta}(s_0,s_1)\,, \qquad i=1,\ldots,p\,. \tag{5.39}$$

Since

$$C_{c,\theta}(s_0,s_1)=\sum_{i=s_0}^{s_1}C_{c,\theta}(i,i)$$

one gets the remarkable property that

$$\begin{aligned} & a_{np}^{-4}\max_{i=1,\ldots,p}\Big|\lambda_{(i)}\Big(\sum_{s=s_0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'\Big)-Z_{(i),np}^4C_{c,\theta}(s_0,s_1)\Big| \\ = \;& a_{np}^{-4}\max_{i=1,\ldots,p}\Big|\sum_{s=s_0}^{s_1}\lambda_{(i)}(\mathbf{A}_n(s)\mathbf{A}_n(s)')-Z_{(i),np}^4C_{c,\theta}(s_0,s_1)\Big|+o_P(1)\,. \end{aligned}$$

In particular, for $s_1\ge s_0$ we get the weak convergence of the point processes towards a PRM:

$$\begin{aligned} & \sum_{i=1}^{p}\varepsilon_{a_{np}^{-4}\big(\lambda_i\big(\sum_{s=s_0}^{s_0}\mathbf{A}_n(s)\mathbf{A}_n(s)'\big),\ldots,\lambda_i\big(\sum_{s=s_0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'\big)\big)} \\ \stackrel{d}{\to}\; & \sum_{i=1}^{\infty}\varepsilon_{\Gamma_i^{-4/\alpha}\big(C_{c,\theta}(s_0,s_0),\ldots,C_{c,\theta}(s_0,s_1)\big)}\,. \end{aligned}$$

Example 5.20. In Figure 5.10 we calculate the largest eigenvalues $\lambda_{(1)}\big(\sum_{s=0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'\big)$ for $s_1=0,\ldots,5$ as well as the sums of the largest eigenvalues $\sum_{s=0}^{s_1}\lambda_{(1)}\big(\mathbf{A}_n(s)\mathbf{A}_n(s)'\big)$ the log-return series from the S&P 500 index described in Section 5.1.2. The data are not rank-transformed. We notice that the two values are surprisingly close across the values $s_0=0,\ldots,5$. This phenomenon could be explained by the structure of the eigenvalues in Example 5.19. Also note that the largest eigenvalue $\mathbf{A}_n(0)\mathbf{A}_n(0)'$ makes a major contribution to the values in Figure 5.10; the contribution of the squares $\mathbf{A}_n(s)\mathbf{A}_n(s)'$, $s=1,\ldots,5$, to the largest eigenvalue of the sum of squares is less substantial.

5.A Auxiliary results

Let (Z_i) be iid copies of Z whose distribution satisfies

$$\mathbb{P}(Z>x)\sim p_+\frac{L(x)}{x^\alpha}\quad\text{and}\quad\mathbb{P}(Z\le -x)\sim p_-\frac{L(x)}{x^\alpha}\,,\quad x\to\infty\,,$$

for some tail index $\alpha>0$, where $p_+,p_-\ge 0$ with $p_++p_-=1$ and L is a slowly varying function. If $\mathbb{E}[|Z|]<\infty$ also assume $\mathbb{E}[Z]=0$. The product Z_1Z_2 is regular varying with the same index α and $\mathbb{P}(|Z_1Z_2|>x)=x^{-\alpha}L_1(x)$, where L_1 is slowly varying function different from L; see Embrechts and Goldie [57]. Write

$$S_n=Z_1+\cdots+Z_n\,,\quad n\ge 1,$$

and consider a sequence (a_n) such that $\mathbb{P}(|Z|>a_n)\sim n^{-1}$.

5.A.1 Large deviation results

The following theorem can be found in Nagaev [92] and Cline and Hsing [35] for $\alpha > 2$ and $\alpha \le 2$, respectively; see also Denisov et al. [48].

Theorem 5.21. *Under the assumptions on the iid sequence (Z_t) given above the following relation holds*

$$\sup_{x \ge c_n} \left| \frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(|Z| > x)} - p_+ \right| \to 0\,,$$

where (c_n) is any sequence satisfying $c_n/a_n \to \infty$ for $\alpha \le 2$ and $c_n \ge \sqrt{(\alpha-2)n\log n}$ for $\alpha > 2$.

5.A.2 A point process convergence result

Assume that the conditions at the beginning of Appendix 5.A hold. Consider a sequence of iid copies $(S_n^{(t)})_{t=1,2,\ldots}$ of S_n and the sequence of point processes

$$N_n = \sum_{t=1}^{p} \varepsilon_{a_{np}^{-1} S_n^{(t)}}\,, \quad n = 1, 2, \ldots\,,$$

for an integer sequence $p = p_n \to \infty$. We assume that the state space of the point processes N_n is $\overline{\mathbb{R}}_0 = [\mathbb{R} \cup \{\pm\infty\}] \backslash \{0\}$.

Lemma 5.22. *Assume $\alpha \in (0,2)$ and the conditions of Appendix 5.A on the iid sequence (Z_t) and the normalizing sequence (a_n). Then the limit relation $N_n \xrightarrow{\text{d}} N$ holds in the space of point measures on $\overline{\mathbb{R}}_0$ equipped with the vague topology (see [96, 97]) for a Poisson random measure N with state space $\overline{\mathbb{R}}_0$ and intensity measure $\mu_\alpha(dx) = \alpha|x|^{-\alpha-1}(p_+\mathbf{1}_{\{x>0\}} + p_-\mathbf{1}_{\{x<0\}})dx$.*

Proof. According to Resnick [96], Proposition 3.21, we need to show that $p\,\mathbb{P}(a_{np}^{-1}S_n \in \cdot) \xrightarrow{\text{v}} \mu_\alpha$, where $\xrightarrow{\text{v}}$ denotes vague convergence of Radon measures on $\overline{\mathbb{R}}_0$. Observe that we have $a_{np}/a_n \to \infty$ as $n \to \infty$. This fact and $\alpha \in (0,2)$ allow one to apply Theorem 5.21:

$$\frac{\mathbb{P}(S_n > x a_{np})}{n\,\mathbb{P}(|Z| > a_{np})} \to p_+ x^{-\alpha} \quad \text{and} \quad \frac{\mathbb{P}(S_n \le -x a_{np})}{n\,\mathbb{P}(|Z| > a_{np})} \to p_-\, x^{-\alpha}\,, \quad x > 0\,.$$

On the other hand, $n\,\mathbb{P}(|Z| > a_{np}) \sim p^{-1}$ as $n \to \infty$. This proves the lemma. □

Bibliography

[1] Abramowitz, M., and Stegun, I. *Handbook of Mathematical Functions.* Dover, New York, 1972.

[2] Alexander, C., and Chibumba, A. Multivariate orthogonal factor GARCH. Tech. rep., University of Sussex, 1996. Discussion Papers in Mathematics.

[3] Alsmeyer, G., and Mentemeier, S. Tail behaviour of stationary solutions of random difference equations: the case of regular matrices. *Journal of Difference Equations and Applications 18*, 8 (2012), 1305–1332.

[4] Andersen, T., Davis, R. A., Kreiss, J., and Mikosch, T. *Handbook of Financial Time Series.* Springer, Heidelberg, 2009.

[5] Anderson, T. W. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics 34*, 1 (1963), 122–148.

[6] Asmussen, S., and Glynn, P. *Stochastic Simulation : Algorithms and Analysis.* Stochastic modelling and applied probability. Springer, New York, 2007.

[7] Auffinger, A., Ben Arous, G., and Péché, S. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Annals of Institute Henri Poincaré Probability and Statistics 45*, 3 (2009), 589–610.

[8] Back, K. Asset pricing for general processes. *Journal of Mathematical Economics 20*, 4 (1991), 371 – 395.

[9] Bai, Z., and Silverstein, J. W. *Spectral Analysis of Large Dimensional Random Matrices*, second ed. Springer Series in Statistics. Springer, New York, 2010.

[10] Bai, Z. D., Silverstein, J. W., and Yin, Y. Q. A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *Journal of Multivariate Analysis 26*, 2 (1988), 166–168.

[11] Barndorff-Nielsen, O., and Shephard, N. Modelling by Lévy processes for financial econometrics. In *Lévy Processes. Theory and Applications*, O. Barndorff-Nielsen, T. Mikosch, and S. Resnick, Eds. Birkhäuser, Boston, 2001, ch. V, pp. 283–318.

[12] Barndorff-Nielsen, O. E., and Shephard, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*, 2 (2002), 253–280.

[13] Basrak, B., Davis, R. A., and Mikosch, T. A characterization of multivariate regular variation. *Annals of Applied Probability 12*, 3 (2002), 908–920.

[14] Basrak, B., Davis, R. A., and Mikosch, T. Regular variation of GARCH processes. *Stochastic Processes and their Applications 99*, 1 (2002), 95–115.

[15] Basrak, B., and Segers, J. Regularly varying multivariate time series. *Stochastic Processes and their applications 119*, 4 (2009), 1055–1080.

[16] BELINSCHI, S., DEMBO, A., AND GUIONNET, A. Spectral measure of heavy tailed band and covariance random matrices. *Communications of Mathematical Physics 289*, 3 (2009), 1023–1055.

[17] BELITSKIĬ, G. R., AND LYUBICH, Y. I. *Matrix Norms and their Applications*, vol. 36 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1988.

[18] BEN AROUS, G., AND GUIONNET, A. The spectrum of heavy tailed random matrices. *Communications of Mathematical Physics 278*, 3 (2008), 715–751.

[19] BENAYCH-GEORGES, F., AND PÉCHÉ, S. Localization and delocalization for heavy tailed band matrices. *Annals of Institute Henri Poincaré Probability and Statistics 50*, 4 (2014), 1385–1403.

[20] BHATIA, R. *Matrix Analysis*, vol. 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[21] BINGHAM, N. H., GOLDIE, C. M., AND TEUGELS, J. L. *Regular Variation*, vol. 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.

[22] BLANCHET, J., AND GLYNN, P. Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Annals of Applied Probability 18*, 4 (Aug. 2008), 1351–1378.

[23] BLANCHET, J., AND LIU, J. Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *Journal of Applied Probability 47*, 2 (June 2010), 301–322.

[24] BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 3 (1986), 307–327.

[25] BOLLERSLEV, T. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *Review of Economics and Statistics 72*, 3 (Aug. 1990), 498–505.

[26] BOUGEROL, P., AND PICARD, N. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics 52*, 1-2 (1992), 115–127.

[27] BREIMAN, L. On some limit theorems similar to the arc-sin law. *Thoery of Probability & Its Applications 10*, 2 (1965), 323–331.

[28] BURACZEWSKI, D., DAMEK, E., GUIVARC'H, Y., AND MENTEMEIER, S. On multidimensional Mandelbrot cascades. *Journal of Difference Equations and Applications 20*, 11 (2014), 1523–1567.

[29] BURACZEWSKI, D., DAMEK, E., AND MIKOSCH, T. *Stochastic Models with Power-Law Tails: The Equation $X = AX + B$*. Springer, New York, 2016.

[30] CAMPBELL, J., LO, A., AND MACKINLAY, A. *The Econometrics of Financial Markets*. Princeton University Press, 1996.

[31] CHALABI, Y. `https://cran.r-project.org/web/packages/fGarch/index.html`, 2016.

[32] CHOVER, J., NEY, P., AND WAINGER, S. Functions of probability measures. *Journal d'Analyse Mathématique 26* (1973), 255–302.

[33] CLARK, P. K. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society 41*, 1 (Jan. 1973), 135–155.

[34] CLINE, D. Convolution tails, product tails and domains of attraction. *Probability Th. Rel. Fields 72* (1986), 529–557.

[35] CLINE, D. B. H., AND HSING, T. Large deviation probabilities for sums of random variables with heavy or subexponential tails. *Technical report. Statistics Dept., Texas A&M University.* (1998).

[36] COLLAMORE, J., DIAO, G., AND VIDYASHANKAR, A. Rare event simulation for processes generated via stochastic fixed point equations. *Annals of Applied Probability 24*, 5 (Oct. 2014), 2143–2175.

[37] COLLAMORE, J., AND MENTEMEIER, S. Large excursions and conditioned laws for recursive sequences generated by random matrices. *arXiv preprint arXiv:1608.05175* (2016).

[38] DAVIS, R. A., HEINY, J., MIKOSCH, T., AND XIE, X. Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes 19*, 3 (2016), 517–547.

[39] DAVIS, R. A., AND HSING, T. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability 23*, 2 (1995), 879–917.

[40] DAVIS, R. A., AND MIKOSCH, T. The sample autocorrelations of financial time series models. In *Nonlinear and Nonstationary Signal Processing*, W. Fitzgerald, R. Smith, A. Walden, and P. Young, Eds. Cambridge University Press, 1999, pp. 247–274.

[41] DAVIS, R. A., AND MIKOSCH, T. Point process convergence of stochastic volatility processes with application to sample autocorrelation. *Journal of Applied Probability 38A* (Feb. 2001), 93–104.

[42] DAVIS, R. A., AND MIKOSCH, T. The extremogram: A correlogram for extreme events. *Bernoulli 15*, 4 (2009), 977–1009.

[43] DAVIS, R. A., AND MIKOSCH, T. Probabilistic properties of stochastic volatility models. In *Handbook of Financial Time Series*, T. Andersen, R. Davis, J. Kreiß, and T. Mikosch, Eds. Springer, Heidelberg, 2009.

[44] DAVIS, R. A., MIKOSCH, T., AND PFAFFEL, O. Asymptotic theory for the sample covariance matrix of a heavy-tailed multivariate time series. *Stochastic Processes and their Applications 126*, 3 (2016), 767–799.

[45] Davis, R. A., Pfaffel, O., and Stelzer, R. Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. *Stochastic Processes and their Applications 124*, 1 (2014), 18–50.

[46] Davis, R. A., and Resnick, S. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability 13*, 1 (1985), 179–195.

[47] Davis, R. A., and Resnick, S. Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics 14*, 2 (1986), 533–558,.

[48] Denisov, D., Dieker, A. B., and Shneer, V. Large deviations for random walks under subexponentiality: the big-jump domain. *Annals of Probability 36*, 5 (2008), 1946–1991.

[49] Doukhan, P. Mixing: Properties and examples. Tech. rep., Université de Paris-Sud. Département de Mathématique, 1994.

[50] Drees, H. Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli 9*, 1 (2003), 617–657.

[51] Drees, H. Some aspects of extreme value statistics under serial dependence. *Extremes 11*, 1 (2008), 35–53.

[52] Drees, H., and en, H. R. Limit Theorems for Empirical Processes of Cluster Functionals. *Annals of Statistics 38*, 4 (2010), 2145–2186.

[53] Drees, H., Haan, L. d., and Resnick, S. How to make a hill plot. *Annals of Statistics 28*, 1 (Feb. 2000), 254–274.

[54] Eberlein, E. Applications of generalized hyperbolic Lévy motions to finance. In *Lévy Processes. Theory and Applications.*, O. Barndorff-Nielsen, T. Mikosch, and S. Resnick, Eds. Birkhäuser, Boston, 2001, ch. V, pp. 319–336.

[55] El Karoui, N. On the largest eigenvalue of Wishart matrices with identity covariance when n,p and p/n tend to infinity. *Available at* http://arxiv.org/abs/math/0309355 (2003).

[56] Embrechts, P., and Goldie, C. On convolution tails. *Stochastic Processes and their Applications 13*, 3 (1982), 263–278.

[57] Embrechts, P., and Goldie, C. M. On closure and factorization properties of subexponential and related distributions. *Journal of Australian Mathematical Society Series A 29*, 2 (1980), 243–256.

[58] Embrechts, P., Klüppelberg, C., and Mikosch, T. *Modelling Extreme Events for Insurance and Finance.* Springer-Verlag, Berlin Heidelberg, 1997.

[59] Embrechts, P., and Veraverbeke, N. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics 1*, 1 (1982), 55–72.

[60] Engle, R., and Ng, V. Measuring and testing the impact of news on volatility. *Journal of Finance 48*, 5 (1993), 1749–1778.

[61] Feller, W. *An Introduction to Probability Theory and its Applications. Vol. II.* John Wiley & Sons, New York, 1966.

[62] Geman, S. A limit theorem for the norm of random matrices. *Annals of Probability 8*, 2 (1980), 252–261.

[63] Glosten, L., Jagannathan, R., and Runkle, D. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 5 (1993), 1779–1801.

[64] Goldie, C. Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability 1*, 1 (1991), 126–166.

[65] Guivarc'h, Y., and Le Page, E. Spectral gap properties for linear random walks and pareto's asymptotics for affine stochastic recursions. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 52*, 2 (2016), 503–574.

[66] Gul, F. A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society 59*, 3 (May 1991), 667–686.

[67] Haan, L. d., and Ferreira, A. *Extreme Value Theory: An Introduction.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[68] Heiny, J., and Mikosch, T. Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. Tech. rep., University of Copenhagen, 2016.

[69] Heiny, J., Mikosch, T., and Davis, R. Limit theory for the singular values of the sample autocovariance matrix function of multivariate time series. *Work in progress* (2015).

[70] Hennion, H. Limit theorems for products of positive random matrices. *Annals of Probability 25*, 4 (Oct. 1997), 1545–1587.

[71] Hill, B., et al. A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*, 5 (1975), 1163–1174.

[72] Hoga, Y. Testing for changes in (extreme) VaR. *The Econometrics Journal 20*, 1 (Feb. 2017), 23–51.

[73] Horn, R. A., and Johnson, C. R. *Matrix Analysis*, second ed. Cambridge University Press, Cambridge, 2013.

[74] Hult, H., and Lindskog, F. Extremal behavior of regularly varying stochastic processes. *Stochastic Processes and their Applications 115*, 2 (2005), 249–274.

[75] Hult, H., and Lindskog, F. Regular variation for measures on metric space. *Publ. de l'Inst. Mathematical 80*, 94 (2006), 121–140.

[76] Jacod, J. Limit of random measures associated with the increments of a brownian semimartingale. *Journal of Financial Econometrics*, 021 (2017), 1–44.

[77] Jansen, D., and de Vries, C. On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics 73*, 1 (1991), 18–24.

[78] Janssen, A., and Drees, H. A stochastic volatility model with flexible extremal dependence structure. *Bernoulli 22*, 3 (2016), 1448–1490.

[79] Janssen, A., Mikosch, T., Mohsen, R., and Xiaolei, X. The eigenvalues of the sample covariance matrix of a multivariate heavy-tailed stochastic volatility model. *To appear in Bernoulli* (2016).

[80] Jeantheau, T. Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*, 01 (1998), 70–86.

[81] Jessen, A. H., and Mikosch, T. Regularly varying functions. *Publications of the Institute of Mathematics 80*, 94 (2006), 171–192.

[82] Johansson, K. Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Communications of Mathematical Physics 215*, 3 (2001), 683–705.

[83] Johnstone, I. M. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics 29*, 2 (2001), 295–327.

[84] Kesten, H. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica 131*, 1 (1973), 207–248.

[85] Kulik, R., and Soulier, P. Heavy tailed time series with extremal independence. *Extremes 18*, 2 (2015), 273–299.

[86] Lam, C., and Yao, Q. Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics 40*, 2 (2012), 694–726.

[87] Lanne, M., and Pentti, S. Modelling conditional skewness in stock returns. *European Journal of Finance 13*, 8 (2007), 691–704.

[88] Mikosch, T. Modelling dependence and tails of financial time series. *Extreme Values in Finance, Telecommunications, and the Environment* (2003), 185–286.

[89] Mikosch, T., and Rezapour, M. Stochastic volatility models with possible extremal clustering. *Bernoulli 19*, 5A (2013), 1688–1713.

[90] Mikosch, T., and Wintenberger, O. A large deviations approach to limit theory for heavy-tailed time series. *Probability Theory Related Fields 166*, 1-2 (2016), 233–269.

[91] Muirhead, R. J. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[92] Nagaev, S. V. Large deviations of sums of independent random variables. *Annals of Probability 7*, 5 (Oct. 1979), 745–789.

[93] Ney, P., and Nummelin, E. Markov additive processes I. eigenvalue properties and limit theorems. *Annals of Probability 15*, 2 (1987), 561–592.

[94] Petrov, V. V. *Limit Theorems of Probability Theory*, vol. 4 of *Oxford Studies in Probability.* The Clarendon Press, Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[95] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Guhr, T., and Stanley, H. Random matrix approach to cross correlations in financial data. *Physics Review E 65*, 6 (2002).

[96] Resnick, S. *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag New York, 1987.

[97] Resnick, S. *Heavy-Tail Phenomena: Probabilistic and Statistical Modelling.* Springer Science & Business Media, 2007.

[98] Rootzén, H. Extreme value theory for moving average processes. *Annals of Probability 14*, 2 (1986), 612–652.

[99] Routledge, B., and Zin, S. Generalized disappointment aversion and asset prices. *Journal of Finance 65*, 4 (2010), 1303–1332.

[100] Sentana, E. Quadratic arch models. *Review of Economic Studies 62*, 4 (1995), 639–661.

[101] Shephard, N., and Andersen, T. G. Stochastic volatility: Origins and overview. In *Handbook of Financial Time Series.* Springer, Berlin, Heidelberg, 2009, pp. 233–254.

[102] Shores, T. S. *Applied Linear Algebra and Matrix Analysis.* Undergraduate Texts in Mathematics. Springer, New York, 2007.

[103] Siegmund, D. Importance sampling in the monte carlo study of sequential tests. *The Annals of Statistics 4*, 4 (1976), 673–684.

[104] Silvennoinen, A., and Teräsvirta, T. Multivariate GARCH models. In *Handbook of Financial Time Series.* Springer, Heidelberg, 2009, pp. 201–229.

[105] Soshnikov, A. Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. *Electronic Communications in Probability 9* (2004), 82–91.

[106] Soshnikov, A. Poisson statistics for the largest eigenvalues in random matrix ensembles. In *Mathematical physics of quantum mechanics*, vol. 690 of *Lecture Notes in Physics.* Springer, Berlin, 2006, pp. 351–364.

[107] Stărică, C. Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance 6*, 5 (1999), 515–553.

[108] Tao, T., and Vu, V. Random matrices: universality of local eigenvalue statistics up to the edge. *Communications of Mathematical Physics 298*, 2 (2010), 549–572.

[109] Tao, T., and Vu, V. Random covariance matrices: universality of local statistics of eigenvalues. *Annals of Probability 40*, 3 (2012), 1285–1315.

[110] Taylor, J. Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices. In *Time Series Analysis: Theory & Practice*. John Wiley & Sons, New York-London-Sydney, 1982, ch. 1, pp. 203–226.

[111] Taylor, S. *Modelling Financial Time Series*. John Wiley & Sons, New York-London-Sydney, 1986.

[112] Tracy, C. A., and Widom, H. Distribution functions for largest eigenvalues and their applications. In *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)* (2002), Higher Ed. Press, Beijing, pp. 587–596.

[113] van der Weide, R. GO-GARCH: a multivariate generalized orthogonal garch model. *Journal of Applied Econometrics 17*, 5 (2002), 549–564.

[114] Vanneste, J. Estimating generalized lyapunov exponents for products of random matrices. *Physical Review E 81*, 3 (2010), 036701.

[115] von Bahr, B., and Esseen, C.-G. Inequalities for the r-th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *The Annals of Mathematical Statistics 36*, 1 (1965), 299–303.

[116] Vrontos, I., Dellaportas, P., and Politis, D. A full-factor multivariate GARCH model. *Econometrics Journal 6*, 2 (2003), 312–334.