A Comparison of High-Dimensional Sample Covariance and Correlation Matrices of a Heavy-Tailed Time Series

Johannes Heiny

University of Aarhus

Joint work with Thomas Mikosch (Copenhagen).

Aarhus, August 15 2017

Setup

- **Data matrix** $\boldsymbol{X} = \boldsymbol{X}_n$: $p \times n$ matrix of iid entries.

$$\boldsymbol{X} = (X_{it})_{i=1,\dots,p;t=1,\dots,n}$$

- **Sample covariance matrix** $\mathbf{S} = \boldsymbol{X}\boldsymbol{X}'$
- **Ordered eigenvalues** of $\mathbf{S}$

$$\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(p)}$$

- **Applications:**
  - Principal Component Analysis
  - Linear Regression, ...

Assumptions:

1. iid, centered entries $X_{it} \stackrel{\mathrm{d}}{=} X$
2. **Moment condition:** $\mathbb{E}[X^4] < \infty$
3. **Growth regime:** $p/n \to \gamma \in (0, \infty)$

Known Results

Assumptions:

1. iid, centered entries $X_{it} \stackrel{\mathrm{d}}{=} X$
2. **Moment condition:** $\mathbb{E}[X^4] < \infty$
3. **Growth regime:** $p/n \to \gamma \in (0, \infty)$

Behavior of $\lambda_{(1)}$ under finite fourth moment

Then Bai, Yin, Krishnaiah (1988)

$$\frac{1}{n}\lambda_{(1)} \to (1 + \sqrt{\gamma})^2 \, \mathbb{E}[X^2] \quad \text{a.s.}$$

If $X \sim N(0,1)$, Johnstone (2001) showed that

$$n^{2/3} \frac{(\sqrt{\gamma})^{1/3}}{(1+\sqrt{\gamma})^{4/3}} \Big(\frac{\lambda_{(1)}}{n} - \big(1 + \sqrt{\tfrac{p}{n}}\big)^2 \Big) \stackrel{\mathrm{d}}{\to} \text{Tracy–Widom}.$$

Our Problem

New assumptions:

1. X_{it} iid, centered if the mean exists
2. **Moment condition:** $\mathbb{E}[X^4] = \infty$
3. **General growth regime:** $p = n^\beta \ell(n) \to \infty$ for $\beta \in [0, \infty)$.

Goal: Understand precise **asymptotic behavior of $\lambda_{(1)}$.**

Infinite Fourth Moment

- **Regular variation** with index $\alpha \in (0,4)$:

$$\mathbb{P}(|X| > x) = x^{-\alpha} L(x),$$

where L is a slowly varying function.
This implies $\mathbb{E}[X^4] = \infty$.

Infinite Fourth Moment

- **Regular variation** with index $\alpha \in (0,4)$:

$$\mathbb{P}(|X| > x) = x^{-\alpha} L(x),$$

where L is a slowly varying function.
This implies $\mathbb{E}[X^4] = \infty$.

- **Normalizing sequence** (a_{np}^2) such that

$$np\,\mathbb{P}(X^2 > a_{np}^2 x) \to x^{-\alpha/2}, \quad \text{as } n \to \infty \text{ for } x > 0.$$

Then $a_{np} = (np)^{1/\alpha} \ell(np)$ for a slowly varying function ℓ.

More Notation

- **Diagonal entries of S**

$$D_i = \mathbf{S}_{ii} = \sum_{t=1}^{n} X_{it}^2$$

and their **order statistics**

$$D_{(1)} \geq D_{(2)} \geq \cdots \geq D_{(p)}$$

- **Order statistics** of $X_{it}^2, i = 1, \ldots, p; \ t = 1, \ldots, n$

$$X_{(1)}^2 \geq X_{(2)}^2 \geq \ldots \geq X_{(np)}^2$$

Main Result

Theorem (Heiny and Mikosch, 2016)

$\boldsymbol{X}$ with iid regularly varying entries $\alpha \in (0,4)$ and $p_n = n^\beta \ell(n)$ with $\beta \in [0,1]$.

1. If $\beta \in [0,1]$, then

$$a_{np}^{-2} \max_{i=1,\ldots,p} \left|\lambda_{(i)} - D_{(i)}\right| \stackrel{\mathbb{P}}{\to} 0\,.$$

2. If $\beta \in ((\alpha/2-1)_+, 1]$, then

$$a_{np}^{-2} \max_{i=1,\ldots,p} \left|\lambda_{(i)} - X_{(i)}^2\right| \stackrel{\mathbb{P}}{\to} 0\,.$$

Comparison

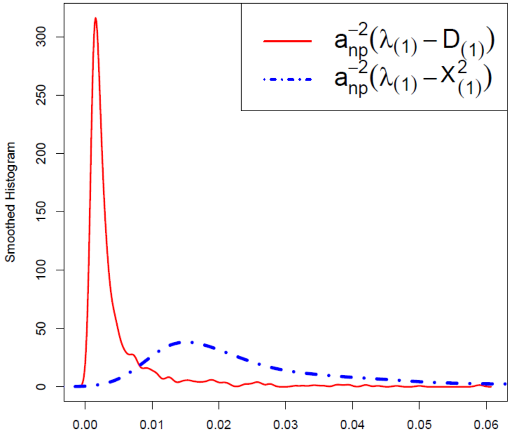


Figure: Smoothed histogram based on 20000 simulations of the approximation error for the normalized eigenvalue $a_{np}^{-2}\lambda_{(1)}$ for entries X_{it} with $\alpha = 1.6$, $\beta = 1$, $n = 1000$ and $p = 200$.

Reduction to Diagonal

Diagonal

$\boldsymbol{X}$ with iid regularly varying entries $\alpha \in (0,4)$ and $p_n = n^\beta \ell(n)$ with $\beta \in [0,1]$. We have

$$a_{np}^{-2} \|\mathbf{S} - \operatorname{diag}(\mathbf{S})\|_2 \stackrel{\mathbb{P}}{\to} 0\,,$$

where $\|\cdot\|_2$ denotes the spectral norm.

$$\mathbf{S}_{ij} = \sum_{t=1}^{n} X_{it} X_{jt}.$$

Idea of the Proof

- **Weyl's inequality**

$$\max_{i=1,\ldots,p} \big|\lambda_{(i)}(\mathbf{A}+\mathbf{B}) - \lambda_{(i)}(\mathbf{A})\big| \le \|\mathbf{B}\|_2 \,.$$

- Choose $\mathbf{A}+\mathbf{B} = \mathbf{S}$ and $\mathbf{A} = \mathrm{diag}(\mathbf{S})$ to obtain

$$a_{np}^{-2} \max_{i=1,\ldots,p} \big|\lambda_{(i)} - \lambda_{(i)}(\mathrm{diag}(\mathbf{S}))\big| \stackrel{\mathbb{P}}{\to} 0\,, \quad n \to \infty\,.$$

- **Note:** Limit theory for $(\lambda_{(i)})$ reduced to $(D_{(i)})$.

Point process convergence

$$N_n = \sum_{i=1}^{p} \delta_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{\mathrm{d}} \sum_{i=1}^{\infty} \delta_{\Gamma_i^{-2/\alpha}} = N$$

The limit is a PRM on $(0,\infty)$ with mean measure $\mu(x,\infty) = x^{-\alpha/2}, x > 0$, and

$$\Gamma_i = E_1 + \cdots + E_i\,, \quad (E_i) \text{ iid standard exponential.}$$

- **Limiting distribution:** For $k \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}(a_{np}^{-2}\lambda_{(k)} \leq x) = \lim_{n\to\infty} \mathbb{P}(N_n(x,\infty) < k) = \mathbb{P}(N(x,\infty) < k)$$
$$= \sum_{s=0}^{k-1} \frac{\left(x^{-\alpha/2}\right)^s}{s!} \, \mathrm{e}^{-x^{-\alpha/2}}, \quad x > 0\,.$$

Point Process of Normalized Eigenvalues

- **Limiting distribution:** For $k \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}(a_{np}^{-2}\lambda_{(k)} \leq x) = \lim_{n\to\infty} \mathbb{P}(N_n(x,\infty) < k) = \mathbb{P}(N(x,\infty) < k)$$
$$= \sum_{s=0}^{k-1} \frac{\left(x^{-\alpha/2}\right)^s}{s!} \mathrm{e}^{-x^{-\alpha/2}}, \quad x > 0\,.$$

- **Largest eigenvalue**

$$\frac{\lambda_{(1)}}{a_{np}^2} \xrightarrow{\mathrm{d}} \Gamma_1^{-\alpha/2}\,,$$

where the limit has a *Fréchet distribution* with parameter $\alpha/2$.
Soshnikov (2006), Auffinger et al. (2009), Auffinger and Tang (2016), Davis et al. $(2014, 2016^2)$, JH and Mikosch (2016)

Point Process of Normalized Eigenvalues

- **Mapping theorem:** For fixed k,

$$a_{np}^{-2}(\lambda_{(1)},\ldots,\lambda_{(k)}) \stackrel{\mathrm{d}}{\to} (\Gamma_1^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha})\,.$$

- We also have

$$\Big(\frac{\lambda_{(2)}}{\lambda_{(1)}},\ldots,\frac{\lambda_{(k)}}{\lambda_{(k-1)}}\Big) \stackrel{\mathrm{d}}{\to} \Big(\Big(\frac{\Gamma_1}{\Gamma_2}\Big)^{2/\alpha},\ldots,\Big(\frac{\Gamma_{k-1}}{\Gamma_k}\Big)^{2/\alpha}\Big).$$

Eigenvectors

- $\mathbf{v}_k$ unit eigenvector of $\mathbf{S}$ associated to $\lambda_{(k)}$
- Unit eigenvectors of $\mathrm{diag}(\mathbf{S})$ are canonical basisvectors $\mathbf{e}_j$.

Eigenvectors

$\boldsymbol{X}$ with iid regularly varying entries with index $\alpha \in (0,4)$ and $p_n = n^\beta \ell(n)$ with $\beta \in [0,1]$. Then for any fixed $k \geq 1$,

$$\|\mathbf{v}_k - \mathbf{e}_{L_k}\|_{\ell_2} \xrightarrow{\mathbb{P}} 0\,, \quad n \to \infty\,.$$

Localization vs. Delocalization

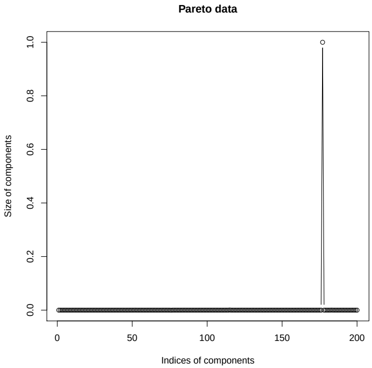


Figure: $X \sim$ Pareto(0.8)

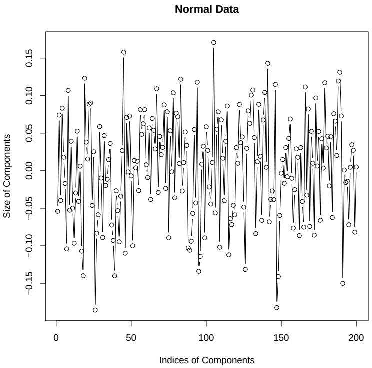


Figure: $X \sim N(0,1)$

Components of eigenvector $\mathbf{v}_1$. $p = 200$, $n = 1000$.

High-dimensional sample correlation matrices

Sample Correlation Matrix

Assumptions:

1. iid, centered entries $X_{it} \stackrel{\mathrm{d}}{=} X$
2. **Growth regime:** $\lim_{n\to\infty} \frac{p}{n} = \gamma \in (0,1]$

Assumptions:

1. iid, centered entries $X_{it} \stackrel{\mathrm{d}}{=} X$
2. **Growth regime:** $\lim_{n\to\infty} \frac{p}{n} = \gamma \in (0,1]$

- **Sample correlation matrix $\mathbf{R}$** with entries

$$R_{ij} = \frac{\frac{1}{n}\sum_{t=1}^{n} X_{it}X_{jt}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} X_{it}^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n} X_{jt}^2}}, \quad i,j = 1,\ldots,p$$

- $\mathbf{R} = \boldsymbol{Y}\boldsymbol{Y}'$, where

$$\boldsymbol{Y} = (Y_{ij})_{p\times n} = \left(\frac{X_{ij}}{\sqrt{\sum_{t=1}^{n} X_{it}^2}}\right)_{p\times n}$$

Objective

- **Eigenvalues** of $\mathbf{R}$

$$\mu_{(1)} \geq \cdots \geq \mu_{(p)}$$

- **Problem:** Asymptotic behavior of $\mu_{(1)}$ and $\mu_{(p)}$

A Comparison

With $\mathbf{F} = (\mathrm{diag}(\mathbf{S}))^{-1}$, we have

$$\mathbf{R} = \mathbf{F}^{1/2}\,\mathbf{S}\,\mathbf{F}^{1/2}\,.$$

- **Weyl's inequality**:

$$\begin{aligned}\max_{i=1,\dots,p} |\mu_{(i)} - n^{-1}\lambda_{(i)}| &\le \|\mathbf{S}\,\mathbf{F} - n^{-1}\mathbf{S}\|_2 \\ &\le n^{-1}\|\mathbf{S}\|_2\|n\mathbf{F} - \mathbf{I}\|_2 \\ &= n^{-1}\lambda_{(1)} \max_{i=1,\dots,p}\left|\frac{n}{D_i} - 1\right|.\end{aligned}$$

A Comparison

With $\mathbf{F} = (\mathrm{diag}(\mathbf{S}))^{-1}$, we have

$$\mathbf{R} = \mathbf{F}^{1/2}\,\mathbf{S}\,\mathbf{F}^{1/2}\,.$$

- **Weyl's inequality**:

$$\begin{aligned}\max_{i=1,\dots,p} |\mu_{(i)} - n^{-1}\lambda_{(i)}| &\le \|\mathbf{S}\,\mathbf{F} - n^{-1}\mathbf{S}\|_2 \\ &\le n^{-1}\|\mathbf{S}\|_2 \|n\mathbf{F} - \mathbf{I}\|_2 \\ &= n^{-1}\lambda_{(1)} \max_{i=1,\dots,p} \Big|\frac{n}{D_i} - 1\Big|\,.\end{aligned}$$

- Conclusion: If $\mathbb{E}[X^4] < \infty$,

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.}$$

Jiang (2004), Xiao and Zhou (2010)

A Way To Study Eigenvalues

- **Empirical spectral distribution** of $p \times p$ matrix $\mathbf{A}$ with real eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})$:

$$F_{\mathbf{A}}(x) = \frac{1}{p} \sum_{i=1}^{p} \mathbb{1}_{\{\lambda_i(\mathbf{A}) \le x\}}, \qquad x \in \mathbb{R}\,.$$

- **Limiting spectral distribution**:
  Weak convergence of $(F_{\mathbf{A}_n})$ to distribution function F a.s.

Marčenko–Pastur law F_γ has density

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}\,, & \text{if } x \in [a,b], \\ 0\,, & \text{otherwise,} \end{cases}$$

where $a = (1-\sqrt{\gamma})^2$ and $b = (1+\sqrt{\gamma})^2$.

Limiting Spectral Distribution

Marčenko–Pastur Theorem

If $\mathbb{E}[X^2] = 1$, then $(F_{n^{-1}\mathbf{S}})$ converges weakly to F_γ.

Limiting Spectral Distribution

Marčenko–Pastur Theorem

If $\mathbb{E}[X^2] = 1$, then $(F_{n^{-1}\mathbf{S}})$ converges weakly to F_γ.

Heiny and Mikosch (2017)

Under the domain of attraction type-condition for the Gaussian law,

$$\mathbb{E}\big[Y_{11}Y_{12}\big] = o(n^{-2}) \quad \text{and} \quad \mathbb{E}\big[Y_{11}^4\big] = o(n^{-1}),$$

the sequence $(F_{\mathbf{R}})$ converges weakly to F_γ.

Here $Y_{ij} = \frac{X_{ij}}{\sqrt{\sum_{t=1}^n X_{it}^2}}$.

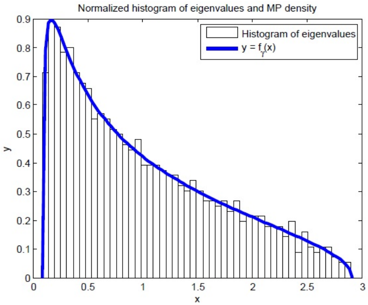
Simulation Study

- **Regular variation** with index $\alpha > 0$
- This implies $\mathbb{E}[|X|^{\alpha+\varepsilon}] = \infty$ for any $\varepsilon > 0$.

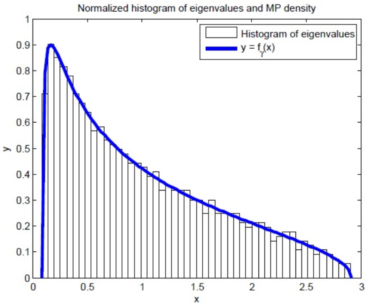
Simulation Study

- **Regular variation** with index $\alpha > 0$
- This implies $\mathbb{E}[|X|^{\alpha+\varepsilon}] = \infty$ for any $\varepsilon > 0$.
- Procedure:
  1. Simulate $\boldsymbol{X}$
  2. Plot histograms of $(\mu_{(i)})$ and $(\lambda_{(i)}/n)$
  3. Compare with Marčenko–Pastur density

$\alpha = 6$



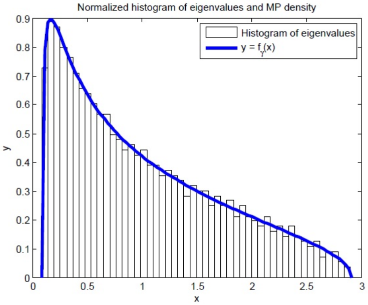
(a) Sample correlation



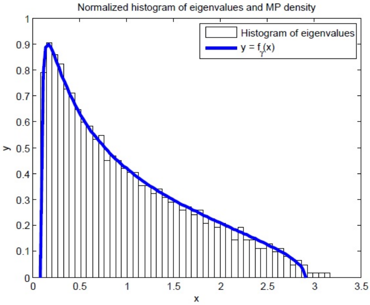
(b) Sample covariance

$$\alpha = 6, n = 2000, p = 1000$$

$\alpha = 3.99$



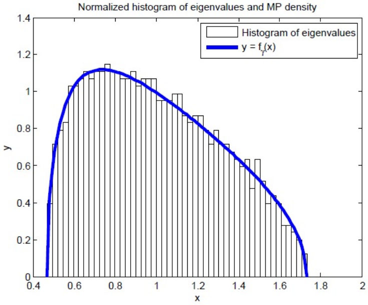
(a) Sample correlation



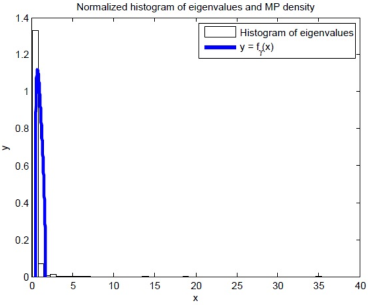
(b) Sample covariance

$$\alpha = 3.99, n = 2000, p = 1000$$

$\alpha = 2.1$



(a) Sample correlation



(b) Sample covariance

$$\alpha = 2.1, n = 10000, p = 1000$$

Assume X is symmetric.

Limiting spectral distribution

Under the domain of attraction type-condition for the Gaussian law,

$$\lim_{n\to\infty} n\,\mathbb{E}[Y_{11}^4] = 0\,,$$

the sequence $(F_{\mathbf{R}})$ converges weakly to F_γ.

Assume X is symmetric.

Limiting spectral distribution

Under the domain of attraction type-condition for the Gaussian law,

$$\lim_{n\to\infty} n\,\mathbb{E}[Y_{11}^4] = 0\,,$$

the sequence $(F_{\mathbf{R}})$ converges weakly to F_γ.

Limits of extreme eigenvalues

If

$$\lim_{n\to\infty} (\log n)^5\, n\,\mathbb{E}[Y_{11}^4] = 0\,,$$

we have

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.}$$

Summary

Sample covariance matrix

Normalization

$\mathbb{E}[X^4]$

Sample correlation matrix

Self-normalization

$\mathbb{E}[X^2]$

- Same results if $\mathbb{E}[X^4] < \infty$.
- Non-iid case and application to financial time series on demand.

Thank you!

Materials of this Talk

[1] Heiny, J., and Mikosch, T. Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. *Stochastic Process. Appl.* (2016), 29. [pdf]

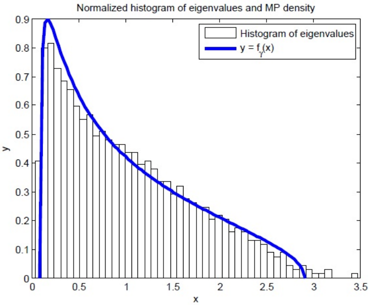
[2] Davis, R. A., Heiny, J., Mikosch, T., and Xie, X. Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes 19*, 3 (2016), 517–547. [pdf]

[3] Heiny, J., and Mikosch, T. Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices under infinite fourth moment. *Submitted for publication.*

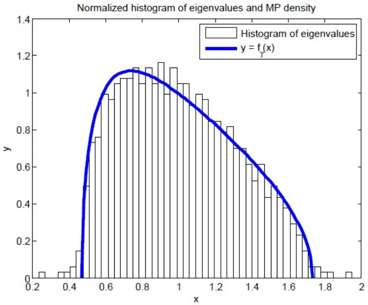
[4] Heiny, J., and Mikosch, T. Limit theory for the singular values of the sample autocovariance matrix function of multivariate time series. *In preparation.*

[5] Heiny, J., Mikosch, T., and Xie, X. Asymptotic theory for high-dimensional stochastic volatility matrices. *In preparation.*

$\alpha < 2$



(a) Sample correlation: $\alpha = 1.5$



(b) Sample correlation: $\alpha = 1.8$

$n = 2000, p = 1000$

$n = 10000, p = 1000$

(Z_{it}): iid field of regularly varying random variables.

- **Stochastic volatility model**:

$$\boldsymbol{X} = \left(Z_{it}\,\sigma_{it}^{(n)}\right)_{p\times n}$$

$\sigma^{(n)}$-field: distribution changes with n, possible long range dependence

Heavy Tails and Dependence

(Z_{it}): iid field of regularly varying random variables.

- **Stochastic volatility model**:

$$\boldsymbol{X} = \left(Z_{it}\,\sigma_{it}^{(n)}\right)_{p\times n}$$

$\sigma^{(n)}$-field: distribution changes with n, possible long range dependence

- **Generate deterministic covariance structure $\mathbf{A}$**:

$$\boldsymbol{X} = \mathbf{A}^{1/2}\mathbf{Z}$$

Heavy Tails and Dependence

(Z_{it}): iid field of regularly varying random variables.

- **Dependence among rows and columns**:

$$X_{it} = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} h_{kl} Z_{i-k,t-l}$$

with some constants h_{kl}. Davis et al. (2016)

Known Results

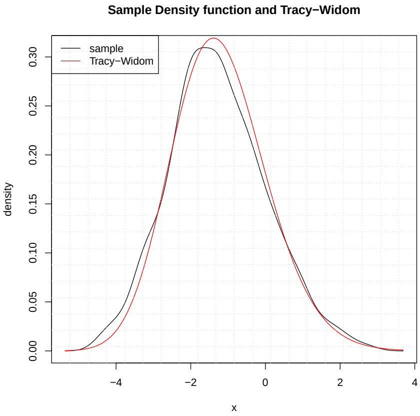


Figure: $\mathbb{P}(X = \sqrt{3}) = \mathbb{P}(X = -\sqrt{3}) = 1/6$, $\mathbb{P}(X = 0) = 2/3$.
Note: The first 4 moments of X match those of the standard normal distribution. $p = 200$, $n = 1000$, 2000 simulations. Tao and Vu (2010)

Application: S&P 500 Index

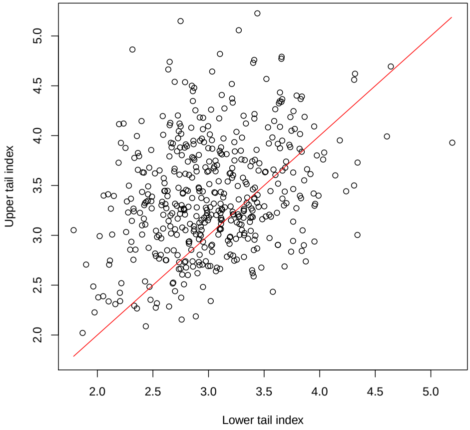


Figure: Estimated tail indices of log-returns of 478 time series in the S&P 500 index.

Application: S&P 500 Index

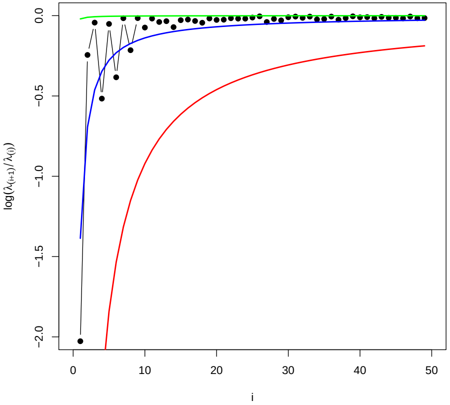


Figure: Logarithms of the ratios $\lambda_{(i+1)}/\lambda_{(i)}$ for the S&P 500 series after rank transform. Quantiles at level 1, 50 and 99% of $\log((\Gamma_i/\Gamma_{i+1})^2)$.

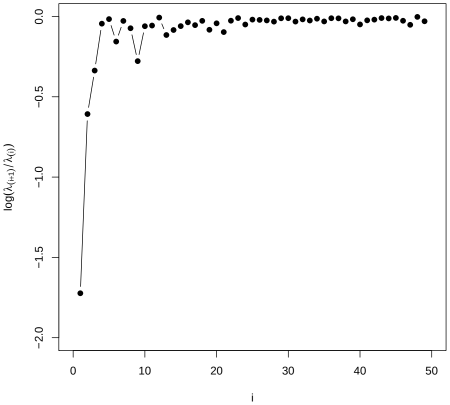


Figure: Logarithms of the ratios $\lambda_{(i+1)}/\lambda_{(i)}$ for the S&P 500 log-return data.

Condition (C_q)

- *There exists a sequence $q = q_n \to \infty$ such that for some integer sequence $k = k_n$ with $k/\log n \to \infty$ we have $(k^3 q)/n \to 0$, and the moment inequality*

$$\mathbb{E}[Y_1^{2m_1} \cdots Y_r^{2m_r}] \le \frac{q_n}{n} \mathbb{E}[Y_1^{2m_1} \cdots Y_{r-1}^{2m_{r-1}} Y_r^{2m_r - 2}] \qquad (C_q)$$

holds for $1 \le r \le k-1$ and any positive integers $m_1, \ldots, m_r$ satisfying $m_1 + \cdots + m_r = k$.

- Giné et al. (1997):

$$\mathbb{E}[Y_1^{2m_1} \cdots Y_r^{2m_r}] = \frac{1}{(k-1)!} \int_0^\infty \lambda^{k-1} (\mathbb{E}[\mathrm{e}^{-\lambda X^2}])^{n-r} \prod_{j=1}^r \mathbb{E}[X^{2m_j}\, \mathrm{e}^{-\lambda X^2}]\, \mathrm{d}\lambda$$