Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails[1]

Richard A. Davis, Columbia University

Oliver Pfaffel, Technical University of Munich
Robert Stelzer, University of Ulm

January 14–17, 2013

Workshop on Heavy-Tailed Distributions and Extreme Value Theory
Kolkata, India

[1] Research supported by the Institute for Advanced Study, TUM-IAS

Motivation

- Large dimensional data sets appear in many quantitative fields like finance, environmental sciences, wireless communications, fMRI, and genetics.
- Structure in this data can often be analyzed via sample covariances.
- PCA is used to transform data to a new set of variables, the principal components, ordered s.t. the first few retain most of the variation of the data.

This suggests the need for an eigenvalue decomposition of the sample covariance matrix.

Game Plan

- The Setup
- Background
- The case $\alpha \in (0, 2)$ for linear time series
  - Elements of the proof I (basics)
  - Elements of the proof II
- Extensions
  - Random coefficient models
  - Hidden Markov model
  - Nonlinear models–stochastic volatility and GARCH(1,1)
- The case $\alpha \in (2, 4)$

The Setup

- Data matrix: A $p \times n$ matrix X consisting of n observations of a p-dimensional time series, i.e.,

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix}.$$

- Sample covariance matrix: the $p \times p$ sample covariance matrix (normalized) is given by

$$XX^T = n\hat{\Gamma}(0) = \left[\sum_{t=1}^{n} X_{it} X_{jt} \right]_{i,j=1}^{p}.$$

- Objective: study the ordered eigenvalues

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

of the $p \times p$ sample covariance matrix XX^T.

The Setup-continued

Data matrix and sample covariance matrix:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix} \text{ and } XX^T = n\hat{\Gamma}(0)$$

- Note that if the rows are independent and identically distributed ergodic time series (with mean 0 and variance 1), then for p fixed,

$$\hat{\Gamma}(0) \xrightarrow{P} I_p \,.$$

- Relation to PCA: $\lambda_{(1)}$ is the empirical variance of the first principal component, $\lambda_{(2)}$ of the second, and so on.

Known results for the largest eigenvalue

- Assume the entries of X are iid Gaussian (with mean zero and variance one)
- For $n \to \infty$ and fixed p, Anderson [1963] proved that

$$\sqrt{\frac{n}{2}}\left(\frac{\lambda_{(1)}}{n} - 1\right) \stackrel{d}{\to} \mathrm{N}(0,1)\,.$$

- Johnstone [2001] showed that for $p, n \to \infty$ s.t. $p/n \to \gamma \in (0, \infty)$

$$\frac{\sqrt{n}+\sqrt{p}}{\sqrt[3]{\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{p}}}}\left(\frac{\lambda_{(1)}}{\left(\sqrt{n}+\sqrt{p}\right)^2} - 1\right) \stackrel{d}{\to} \text{Tracy-Widom distribution}$$

Our objective

- The assumption of Gaussianity in Johnstone's result can be relaxed to a *moment condition* (c.f. Four Moment Theorem by Tao and Vu [2011]; and work by Erdös, Johansson, Péché, Schlein, Soshnikov, Yau and others).
- BUT: in applications one often has neither independent observations, nor Gaussianity or even the existence of sufficient moments.

This lead us to consider heavy-tailed random matrices with dependent entries.

Setting for our results

- Suppose $X = (X_{it})_{i,t}$, $i = 1, \dots, p$, $t = 1, \dots, n$, with

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j} \, ,$$

where $(Z_{i,t})$ is iid with regularly varying tails of index $\alpha \in (0,2)$ (infinite variance), i.e.,

$$n \, P(|Z_{11}| > a_n x) \to x^{-\alpha} \text{ as } n \to \infty, \text{ for } x > 0,$$

and $\sum_{j=-\infty}^{\infty} |c_j|^\delta < \infty$ for some $\delta < \min\{1, \alpha\}$. ($a_n = L(n) n^{1/\alpha}$)

- For $\alpha \in (5/3, 2)$ assume the existence of the following tail balancing limits

$$\lim_{x\to\infty} \frac{P(Z_{11} > x)}{P(|Z_{11}| > x)} \quad \text{and} \quad \lim_{x\to\infty} \frac{P(Z_{11} \le -x)}{P(|Z_{11}| > x)} \, .$$

Theorem (The case $\alpha \in (0, 2)$)

Suppose $p_n, n \to \infty$ such that

$$\limsup_{n\to\infty} \frac{p_n}{n^\beta} < \infty$$

for some $\beta > 0$ satisfying

1. $\beta < \infty$ *if* $\alpha \in (0, 1]$, *and*
2. $\beta < \max\left\{\frac{2-\alpha}{\alpha-1}, \frac{1}{2}\right\}$ *if* $\alpha \in (1, 2)$.

Then, we have the point process convergence,

$$N_p := \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2},$$

where $\Gamma_i = E_1 + \ldots + E_i$ is the cumulative sum of iid standard (i.e., mean one) exponentially distributed rv's.

Condition on β: Growth on p_n is more restrictive as the tail becomes lighter.

The largest eigenvalues

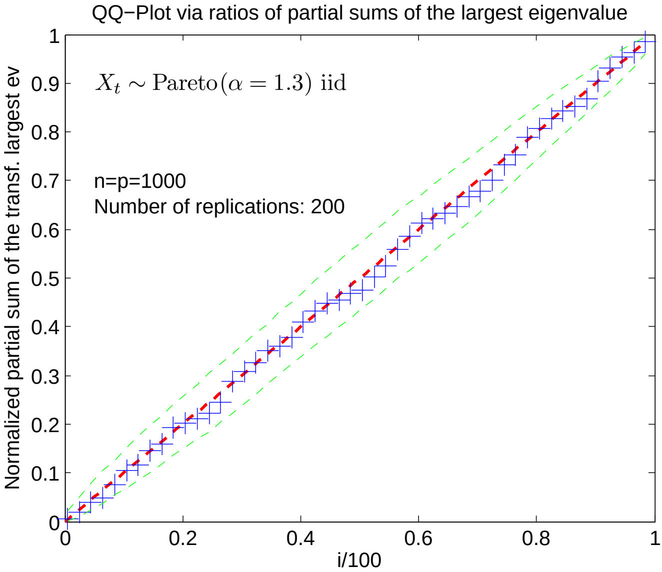
- The theorem implies the joint convergence of the k-largest eigenvalues

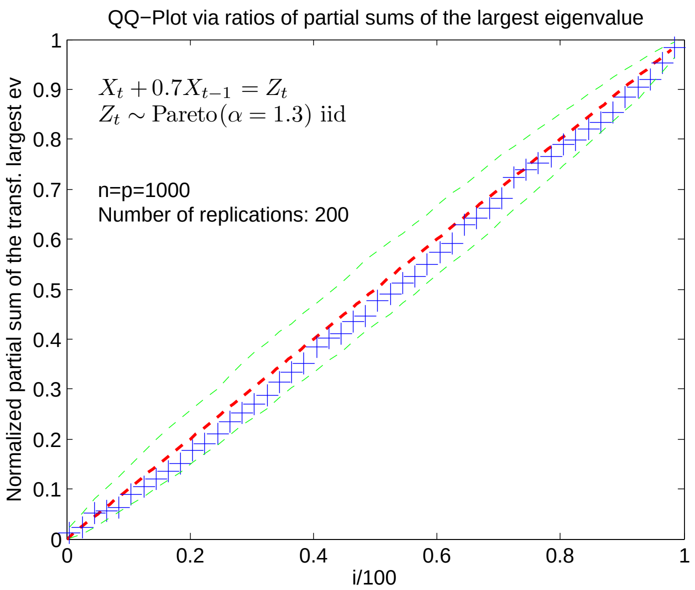
$$a_{np}^{-2}\left(\lambda_{(1)},\ldots,\lambda_{(k)}\right)\xrightarrow{d}\left(\Gamma_1^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha}\right)\left(\sum_{j=-\infty}^{\infty}c_j^2\right). \qquad (1)$$

- For independent entries this was shown by Soshnikov [2006] for $\alpha < 2$, and by Auffinger, Ben Arous and Péché [2009] for $2 \le \alpha < 4$.
- Since Γ_1 is standard exponential, (1) shows, for n, p large, that

$$\left(\frac{\lambda_{(1)}}{a_{np}^2\sum_{j=-\infty}^{\infty}c_j^2}\right)^{-\alpha/2} \text{ is approx. standard exponential}$$

How well is this approximation for finite n and p?


QQ−Plot via ratios of partial sums of the largest eigenvalue
$X_t \sim \text{Pareto}(\alpha = 1.3)$ iid
n=p=1000
Number of replications: 200
Normalized partial sum of the transf. largest ev
i/100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


QQ−Plot via ratios of partial sums of the largest eigenvalue
$X_t + 0.7X_{t-1} = Z_t$
$Z_t \sim \text{Pareto}(\alpha = 1.3)$ iid
n=p=1000
Number of replications: 200
Normalized partial sum of the transf. largest ev
i/100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Elements of the proof I (the basics)

- By definition of $X_{it} = \sum_j c_j Z_{i,t-j}$ we have

$$\sum_{t=1}^{n} X_{it}^2 = \sum_j c_j^2 \sum_{t=1}^{n} \underbrace{Z_{i,t-j}^2}_{\text{tail index } \alpha/2} + 2 \sum_j \sum_{k>j} c_j c_k \sum_{t=1}^{n} \underbrace{Z_{i,t-j} Z_{i,t-k}}_{\text{tail index } \alpha}$$

$$= \sum_j c_j^2 \sum_{t=1}^{n} Z_{i,t}^2 + o_p(a_n^2)$$

- Classical EVT plus large deviations:

$$\max_{1 \le i \le p} a_{np}^{-2} \sum_{t=1}^{n} Z_{it}^2 \xrightarrow{d} \Gamma_1^{-2/\alpha} \quad \text{as } n, p \to \infty.$$

- If $0 < \alpha < 2$ and $\frac{p}{n^\beta} \to \gamma \in [0, \infty)$ for some $\beta > 0$, then as $n, p \to \infty$

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\sum_{t=1}^{n} X_{it}^2)} \sim \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\sum_{t=1}^{n} Z_{it}^2) \sum_{j=-\infty}^{\infty} c_j^2} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2}.$$

Elements of the proof II

- Important tool: $\|A\|_2 = \sqrt{\text{largest eigenvalue of } AA^T}$ (operator 2-norm).
- Define $D \in \mathbb{R}^{p \times p}$ by $D_{ii} = (XX^T)_{ii}$ and $D_{ij} = 0$ for $i \neq j$. Then

$$a_{np}^{-2} \left\| XX^T - D \right\|_2 \xrightarrow{P} 0 \text{ as } p, n \to \infty.$$

- By Weyl's inequality

$$a_{np}^{-2} \left| \lambda_{(1)} - \max_{1 \le i \le p} \sum_{t=1}^{n} X_{it}^2 \right| \le a_{np}^{-2} \left\| XX^T - D \right\|_2 \xrightarrow{P} 0 \text{ as } p, n \to \infty$$

and likewise for $\lambda_{(2)}, \lambda_{(3)}, \ldots$

Hence, we "only" have to derive the extremal behavior of the diagonal elements $(\sum_{t=1}^{n} X_{it}^2)_i$ of XX^T.

Random coefficient models

- So far the rows of X are assumed to be independent and identically distributed processes.
- How to relax the assumption of independence between rows?
- Problem: the number of linear processes is p, and $p \to \infty$.
  $\longrightarrow$ How to parametrize an infinite number of processes?
- One approach: Consider a random coefficient model

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j(\theta_i) Z_{i,t-j} ,$$

where θ_i is some random sequence specified below.

Random coefficient models

- $X = (X_{it})$ with
$$X_{it} = \sum_{j=-\infty}^{\infty} c_j(\theta_i) Z_{i,t-j}$$
  - where $c_j(\cdot)$ is a family of functions s.t.
$$sup_\theta |c_j(\theta)| \le \tilde{c}_j \text{ with } \sum_{j=-\infty}^{\infty} |\tilde{c}_j|^\delta < \infty, \delta < \min\{1, \alpha\}, \text{ and}$$
  - (θ_i) is a stationary ergodic sequence independent of $(Z_{it})_{i,t}$.
- Conditionally on (θ_i), each process $(X_{it})_t$ in the rows of X has a different set of coefficients $(c_j(\theta_i))_j$.
- Unconditionally, the row processes are identically distributed with regularly varying tail probabilities.
- Rows of X are dependent if the θ_i's are dependent.

Random coefficient models

Theorem (Random coefficient models)

Suppose $p_n, n \to \infty$ such that

$$\limsup_{n\to\infty} \frac{p_n}{n^\beta} < \infty$$

for some $\beta > 0$ satisfying

1. $\beta < \infty$ *if* $\alpha \in (0,1]$, *and*
2. $\beta < \max\left\{\frac{2-\alpha}{\alpha-1}, \frac{1}{2}\right\}$ *if* $\alpha \in (1,2)$.

Then, conditionally on (θ_i) as well as unconditionally,

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha}\left(E\left|\sum_{j=-\infty}^{\infty} c_j^2(\theta_1)\right|^{\alpha/2}\right)^{2/\alpha}} .$$

The largest eigenvalue

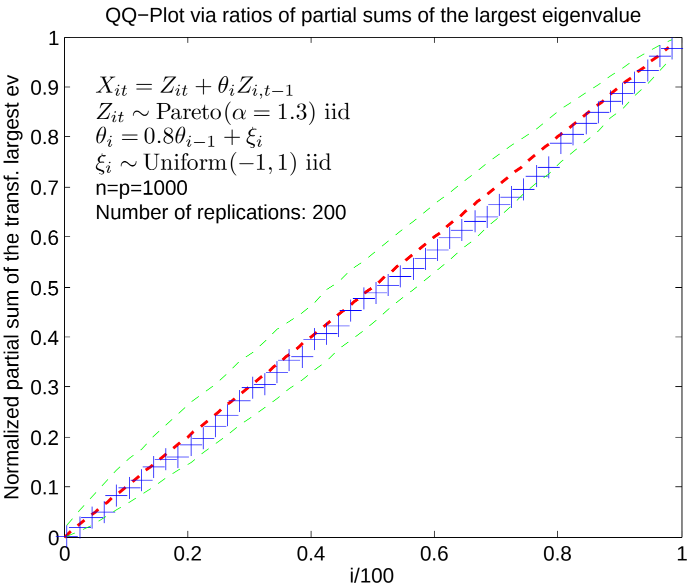
- In particular we have

$$\left(\frac{\lambda_{(1)}}{a_{np}^2}\right)^{-\alpha/2} \xrightarrow{d} \left(E\left|\sum_{j=-\infty}^{\infty} c_j^2(\theta_1)\right|^{\alpha/2}\right)^{-1} \Gamma_1$$

$$= \text{constant} \times \text{standard exponential}$$

as $p, n \to \infty$

- Also a good approximation for finite p and n?


QQ−Plot via ratios of partial sums of the largest eigenvalue
$X_{it} = Z_{it} + \theta_i Z_{i,t-1}$
$Z_{it} \sim \text{Pareto}(\alpha = 1.3)$ iid
$\theta_i = 0.8\theta_{i-1} + \xi_i$
$\xi_i \sim \text{Uniform}(-1, 1)$ iid
n=p=1000
Number of replications: 200
Normalized partial sum of the transf. largest ev
i/100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8

Hidden Markov models

Suppose that (θ_i) is either

- an irreducible Markov chain on a countable state space Θ, or
- positive Harris in the sense of Meyn and Tweedie [2009].
- If (θ_i) has a stationary probability distribution π then, conditionally on (θ_i) as well as unconditionally,

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha}\left(\int |\sum_j c_j^2(\theta)|^{\alpha/2}\pi(d\theta)\right)^{2/\alpha}} .$$

Stochastic volatility models—special case

Suppose the rows are independent copies of the SV process given by

$$X_t = \sigma_t Z_t$$

where (Z_t) is iid RV(α) and $(\ln \sigma_t^2)$ is a purely nondeterministic stationary Gaussian process (this can be weakened), independent of (Z_t).

Theorem *Suppose $p_n, n \to \infty$ such that*

$$\limsup_{n\to\infty} \frac{p_n}{n^\beta} < \infty\,, \quad \textit{for some } \beta > 0 \textit{ satisfying}$$

1. $\beta < \infty$ *if* $\alpha \in (0, 1)$*, and*
2. $\beta < \frac{2-\alpha}{\alpha-1}$ *if* $\alpha \in (1, 2)$*.*

Then, we have the point process convergence,

$$N_p := \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha}}\,.$$

Stochastic volatility models—special case

Point process convergence:

$$N_p := \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha}} .$$

Remarks:

- Proof uses a large deviation result of Davis and Hsing (1995); see also Mikosch and Wintenberger (2012).
- Likely that we can weaken the restriction on β
- Similar results hold for GARCH processes if X_t is RV(α) with $\alpha \in (0, 2)$.

The case $\alpha \in (2,4)$

As before, consider the linear time series

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j}\,,$$

where $(Z_{i,t})$ is iid with mean zero and RV(α) of with $\alpha \in (2,4)$ (i.e., finite variance), and $\sum_{j=-\infty}^{\infty} |c_j| < \infty$. ($a_n = L(n)n^{1/\alpha}$)

Now define the normalized sample covariance matrix by

$$S_n = XX^T - n\mathrm{Var}(X_{11})I_p$$

and let $\lambda_{(1)} \geq \lambda_{(2)} \geq \cdots \geq \lambda_{(p)}$ be the ordered eigenvalues of S_n.

Theorem (The case $\alpha \in (2,4)$)

Suppose $p_n, n \to \infty$ such that

$$\limsup_{n\to\infty} \frac{p_n}{n^\beta} < \infty$$

for some $\beta > 0$ satisfying

1. $\beta < \max\left\{\frac{4-\alpha}{4(\alpha-1)}, \frac{1}{3}\right\}$ *if $\alpha \in (2,3)$, and*
2. $\beta < \frac{4-\alpha}{3\alpha-1}$ *if $\alpha \in (3,4)$.*

Then, we have the point process convergence,

$$N_p := \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2} .$$

Point process convergence,

$$N_p := \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{(i)}} \xrightarrow{d} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2} .$$

Note that if we normalized the data to have variance 1, and consider the ordered eigenvalues $\lambda^*_{(i)}$ from the normalized covariance matrix (without subtracting the identity matrix), we have

$$na_{np}^{-2}(n^{-1}\lambda^*_{(1)} - 1) \xrightarrow{d} \Gamma_1^{-2/\alpha},$$

where the order na_{np}^{-2} is roughly $n^{1-2/\alpha-\beta 2/\alpha}$, which is smaller than the scaling in the Tracy-Widom result.

References

Richard A. Davis, Oliver Pfaffel and Robert Stelzer
Limit Theory for the largest eigenvalues of sample covariance matrices with heavy-tails.
arXiv:1108.5464v1

Ian M. Johnstone
On the Distribution of the Largest Eigenvalue in Principal Components Analysis.
Ann. Statist., 2001, 29, 295-327.

Alexander Soshnikov
Poisson Statistics for the Largest Eigenvalues in Random Matrix Ensembles.
Lect. Notes in Phys., 2006, 690, 351-364.

Antonio Auffinger, Gérard Ben Arous, and Sandrine Péché
Poisson convergence for the largest eigenvalues of heavy tailed random matrices.
Ann. Inst. H. Poincaré Probab. Statist., 2009, 45, 589-610.