Convergence of the largest eigenvalues of a sample covariance matrix for heavy-tailed multivariate time series[1]

Thomas Mikosch

University of Copenhagen

Joint work with Richard A. Davis (Columbia) and Oliver Pfaffel (Munich Re)

[1] Banff, April 30, 2014

1. 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 such that the first few retain most of the variation of the data.**

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

2. The Setup

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

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

- **The $p \times p$ sample covariance matrix (normalized) is given by**

$$\mathbf{X}\mathbf{X}' = \left(\sum_{t=1}^{n} X_{it}X_{jt}\right)_{i,j=1,\dots,p}$$

- **Objective: study the ordered eigenvalues**

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

of the $p \times p$ sample covariance matrix $\mathbf{X}\mathbf{X}'$.

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

$$\frac{1}{n}\mathbf{X}\mathbf{X}' = \Big(\frac{1}{n}\sum_{t=1}^{n} X_{it}X_{jt}\Big)_{i,j=1,\ldots,p} \xrightarrow{\text{a.s.}} I_p$$

3. Known results for the largest eigenvalue

- **Assume the entries of X are iid standard normal.**
- **For $n \longrightarrow \infty$ and fixed p, Anderson (1963) proved that**

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

- **Johnstone (2001) showed that for $p, n \longrightarrow \infty$ such that**

  $p/n \longrightarrow \gamma \in (0, \infty)$

$$\frac{\sqrt{n}+\sqrt{p}}{\left(1/\sqrt{n}+1/\sqrt{p}\right)^{1/3}}\left(\frac{\lambda_{(1)}}{(\sqrt{n}+\sqrt{p})^2} - 1\right) \xrightarrow{d} \text{Tracy-Widom distr.}$$

4. Our objective

- **The assumption of Gaussianity in Johnstone's result can be relaxed to a moment condition; cf. 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 led us to consider heavy-tailed random matrices with dependent entries.

5. Model setup

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

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

- **Regularly varying iid noise (Z_{it}) with index $\alpha \in (0,4)$, (infinite 4th moment) i.e. there exists $a_n = n^{1/\alpha}\ell(n)$ such that**

$$n\,P(|Z| > a_n\,x) \to x^{-\alpha}\,, \quad n \to \infty\,, \quad x > 0\,.$$

and a tail balance condition holds.

- **Summability condition on $H = (h_{kl})$.**

- Growth conditions on $p = p_n \to \infty$.

  $\alpha \in (0,1): p = O(n^\beta)$ for any $\beta > 0$.

  $\alpha \in [1,2): p = O(n^\beta)$ for any $\beta < (\alpha - 1)^{-1}$.

  $\alpha \in [2,4): p = O(n^\beta)$ for any $\beta < (2 - 0.5\alpha)(\alpha - 1)^{-1}$.

  (excludes case $p/n \to \gamma > 0$)

- A crucial role play the quantities

$$D_i = D_i^{(n)} = \sum_{t=1}^{n} Z_{it}^2 \,, \quad i = 1, \ldots, p \,,$$

  and their order statistics $D_{(1)} \geq \cdots \geq D_{(p)}$.

- as well as the operator $M = HH'$, where

$$M_{ij} = \sum_{l=0}^{\infty} h_{il} h_{jl} \,, \quad i, j = 0, 1, \ldots \,,$$

  and its eigenvalues $v_1 \geq v_2 \geq \cdots$.

6. Main result

- **Let (λ_i) be the eigenvalues of $\mathbf{XX}'$ for $\alpha \in (0,2)$ and of $\mathbf{XX}' - E\mathbf{XX}'$ for $\alpha \in (2,4)$, $\lambda_{(1)} \ge \cdots \ge \lambda_{(p)}$ the ordered eigenvalues, $k^2 = o(p)$ an integer sequence,**
- **Then**

$$a_{np}^{-2} \max_{i=1,\ldots,p} \left|\lambda_{(i)} - \delta_{(i)}\right| \xrightarrow{P} 0\,, \quad n \to \infty\,,$$

- **where $\delta_{(1)} \ge \cdots \ge \delta_{(p)}$ are the ordered values of the set $\{D_{(i)} v_j, i = 1,\ldots,k; j = 1,2,\ldots\}$ for $\alpha \in (0,2)$ and of the set $\{(D_{\ell_i} - ED) v_j, i = 1,\ldots,k; j = 1,2,\ldots\}$, $|D_{\ell_1} - ED| \ge \cdots \ge |D_{\ell_p} - ED|$ for $\alpha \in (2,4)$.**

7. Point process convergence

- It follows from Nagaev-type large deviations that for $\alpha \in (0,2)$,
$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2} D_i} \xrightarrow{d} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}},$$
where $\Gamma_i = E_1 + \cdots + E_i$, $i \geq 1$, for iid standard exponentials (E_i).

- The continuous mapping theorem implies (r is the rank of M)
$$\sum_{j=1}^{r} \sum_{i=1}^{p} \varepsilon_{a_{np}^{-2} D_i v_j} \xrightarrow{d} \sum_{j=1}^{r} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha} v_j}.$$

- Hence, by the Main Result,
$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2} \lambda_i} \xrightarrow{d} \sum_{j=1}^{r} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha} v_j}.$$

- The corresponding result holds for $\alpha \in (2,4)$ for $(D_i - ED)$.

- According to Resnick (1987),

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2} D_i} \xrightarrow{d} N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}$$

holds for $\alpha \in (0, 2)$ if and only if A. Nagaev (1968), S. Nagaev (1979)

$$p\, P(D_1 > a_{np}^2 x) \to x^{-\alpha} = \mu(x, \infty]\,, \quad x > 0\,,$$

- For $\alpha \in (2, 4)$,

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(D_i - ED)} \xrightarrow{d} N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}$$

holds if and only if

$$p\, P(D - ED > a_{np}^2 x) \to x^{-\alpha} = \mu(x, \infty]$$

$$p\, P(D - ED \le -a_{np}^2 x) \to 0 = \mu[-\infty, -x]\,, \quad x > 0\,,$$

- μ is the intensity measure of the Poisson process N.

8. The largest eigenvalues

- **The point process convergence implies for fixed m**

$$a_{np}^{-2}\left(\lambda_{(1)},\ldots,\lambda_{(m)}\right) \xrightarrow{d} \left(d_{(1)},\ldots,d_{(m)}\right),$$

where $d_{(1)} \geq \cdots \geq d_{(m)}$ are the m largest values of $\{\Gamma_i^{-2/\alpha} v_j, i = 1, 2, \ldots, j = 1, \ldots, r\}$.

- **For independent entries this was shown by** Soshnikov (2006) **for $\alpha < 2$ and by** Auffinger, Ben Arous, Péché (2009) **for $\alpha \in (2,4)$.**
- **For matrices with linear structure in each of the iid rows this was shown by** Davis, Pfaffel, Stelzer (2013) **for $\alpha \in (0,4)$.**

- **Example.** $X_{it} = Z_{i,t} - Z_{i,t-1} - 2(Z_{i-1,t} - Z_{i-1,t-1})$. Then $v_1 = 8$ and $v_2 = 2$ so that

$$a_{np}^{-2}\big(\lambda_{(1)}, \lambda_{(2)}\big) \xrightarrow{d} \big(8\Gamma_1^{-2/\alpha}, 2\Gamma_1^{-2/\alpha} \vee 8\Gamma_2^{-2/\alpha}\big)$$

- **Example:** The separable case $h_{kl} = \theta_k c_l$. Then $M = \sum_{l=0}^{\infty} c_l^2 (\theta_i \theta_j)_{i,j\ge 0}$ has rank $r = 1$ and $v_1 = \sum_{l=0}^{\infty} c_l^2 \sum_{k=0}^{\infty} \theta_k^2$. The limit point process is Poisson as in the iid case.

$$a_{np}^{-2}\big(\lambda_{(1)}, \ldots, \lambda_{(m)}\big) \xrightarrow{d} v_1\big(\Gamma_1^{-2/\alpha}, \ldots, \Gamma_m^{-2/\alpha}\big)\,,$$

$$\frac{\lambda_{(1)}}{\lambda_{(1)} + \cdots + \lambda_{(m)}} \xrightarrow{d} \frac{\Gamma_1^{-2/\alpha}}{\Gamma_1^{-2/\alpha} + \cdots + \Gamma_m^{-2/\alpha}}\,.$$

9. The trace

- Point process convergence and the continuous mapping theorem also imply that for $\alpha \in (0,2)$[2]

$$a_{np}^{-2}\Big(\lambda_{(1)}, \sum_{i=1}^{p} \lambda_i\Big) \xrightarrow{d} \Big(v_1 \Gamma_1^{-2/\alpha}, \sum_{j=1}^{r} v_j \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}\Big).$$

- Here $\Gamma_1^{-2/\alpha}$ has a Fréchet $\Phi_{\alpha/2}$ distribution and $\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}$ has an $\alpha/2$-stable distribution.
- In particular,

$$\frac{\lambda_{(1)}}{\mathrm{trace}(\mathbf{X}\mathbf{X}')} \xrightarrow{d} \frac{v_1}{v_1 + \cdots + v_r} \, \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}}.$$

[2] The corresponding result for $\alpha \in (2,4)$ holds but requires compensation for the sums.

10. Consistency

- **For $\alpha \in (2, 4)$, $\lambda_{(1)}/n$ is the largest eigenvalue of $n^{-1}(\mathbf{X}_n\mathbf{X}_n' - E\mathbf{X}_n\mathbf{X}_n')$.**

- **One might expect that $(\lambda_{(1)} - \delta_{(1)})/n \xrightarrow{P} 0$ but this is true if and only if $a_{np}^2/n \to 0$ (satisfied if $p = o(n^\gamma)$ for small γ because we also have $p = O(n^\beta)$ for $\beta < (2 - 0.5\alpha)(\alpha - 1)^{-1}$).**

11. Elements of proof

Special case: $X_{it} = \theta_0 Z_{i,t} + \theta_1 Z_{i-1,t}$ **and** $\alpha \in (0,2)$.

$$\begin{aligned}\sum_{t=1}^{n} X_{i,t}^2 &= \sum_{t=1}^{n} \left(\theta_0^2 Z_{i,t}^2 + \theta_1^2 Z_{i-1,t}^2\right) + 2\theta_0\theta_1 \sum_{t=1}^{n} Z_{i,t} Z_{i-1,t} \\ &= \theta_0^2 D_i + \theta_1^2 D_{i-1} + o_P(a_n^2)\,.\end{aligned}$$

Here we used that Z^2 has tail index $\alpha/2$, while $Z_1 Z_2$ has tail index α. Similarly,

$$\begin{aligned}\sum_{t=1}^{n} X_{i,t} X_{i+1,t} &= \theta_0\theta_1 \sum_{t=1}^{n} Z_{i,t}^2 + o_P(a_n^2) \\ &= \theta_0\theta_1 D_i + o_P(a_n^2)\,.\end{aligned}$$

This leads to the approximation

$$\begin{pmatrix} \sum_{t=1}^{n} X_{i,t}^2 & \sum_{t=1}^{n} X_{i,t} X_{i+1,t} \\ \sum_{t=1}^{n} X_{i,t} X_{i+1,t} & \sum_{t=1}^{n} X_{i+1,t}^2 \end{pmatrix}$$

$$\approx \begin{pmatrix} \theta_0^2 & \theta_0\theta_1 \\ \theta_0\theta_1 & \theta_1^2 \end{pmatrix} D_i + \begin{pmatrix} \theta_1^2 & 0 \\ 0 & 0 \end{pmatrix} D_{i-1} + \begin{pmatrix} 0 & 0 \\ 0 & \theta_0^2 \end{pmatrix} D_{i+1}.$$

The sample covariance matrix can be approximated by

$$\left\| \mathrm{XX}' - \sum_{i=1}^{p} D_i M_i \right\|_2 = o_P(a_{np}^2)\,,$$

where

$$M_1 = \begin{pmatrix} \theta_0^2 & \theta_0\theta_1 & 0 & \dots & 0 \\ \theta_0\theta_1 & \theta_1^2 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, M_2 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \theta_0^2 & \theta_0\theta_1 & \dots & 0 \\ 0 & \theta_0\theta_1 & \theta_1^2 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \dots$$

- Denote the order statistics of the D_i's by $D_{(1)} \geq \cdots \geq D_{(p)}$ and $D_{L_i} = D_{(i)}$.
- Then

$$a_{np}^{-2} \Big\| XX' - \sum_{i=1}^{p} D_{L_i} M_{L_i} \Big\|_2 \xrightarrow{P} 0\,.$$

- For $k = k_n \to \infty$ slowly,

$$a_{np}^{-2} \Big\| XX' - \sum_{i=1}^{k} D_{L_i} M_{L_i} \Big\|_2 \xrightarrow{P} 0\,.$$

- Since (D_i) is iid, $(L_1, \ldots, L_p)$ is a random permutation of $(1, \ldots, p)$, hence the event

$$A_k = \{|L_i - L_j| > 1, i \neq j = 1, \ldots, k\}$$

has probability close to one provided $k = o(p^2)$.

- **On the set A_k, the matrix $\sum_{i=1}^{k} D_{L_i} M_{L_i}$ is block-diagonal with non-zero eigenvalues $D_{L_i}(\theta_0^2 + \theta_1^2)$, $i = 1, \dots, k$.**

- **Here we used that M_{L_i} has rank 1 with non-zero eigenvalue equal to $\theta_0^2 + \theta_1^2$.**

- **By Weyl's inequality,**

$$a_{np}^{-2} \max_{i=1,\dots,k} \left| \lambda_{(i)} - D_{L_i}(\theta_0^2 + \theta_1^2) \right| \le a_{np}^{-2} \left\| \mathrm{XX}' - \sum_{i=1}^{k} D_{L_i} M_{L_i} \right\|_2 \xrightarrow{P} 0\,.$$

- **Extension to general structure: Use truncation.**

12. Open problems

- **Centering in the case $\alpha \in (2,4)$.**
- **Order of magnitude of $p = p_n \to \infty$.**
- **Minima, lower order statistics, eigenvectors, determinant,. . . .**
- **Non-linear structure of X_{it}, where the tail of the squares of the noise does not dominate.**