# Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails

Richard A. Davis[a], Oliver Pfaffel[b,*], Robert Stelzer[c,]

[a]*Department of Statistics, Columbia University, New York, USA*
[b]*TUM Institute for Advanced Study & Department of Mathematics, Technische Universität München, Germany*
[c]*Institute of Mathematical Finance, Ulm University, Germany*

**Abstract**

We study the joint limit distribution of the $k$ largest eigenvalues of a $p \times p$ sample covariance matrix $XX^\mathsf{T}$ based on a large $p \times n$ matrix $X$. The rows of $X$ are given by independent copies of a linear process, $X_{it} = \sum_j c_j Z_{i,t-j}$, with regularly varying noise $(Z_{it})$ with tail index $\alpha \in (0,4)$. It is shown that a point process based on the eigenvalues of $XX^\mathsf{T}$ converges, as $n \to \infty$ and $p \to \infty$ at a suitable rate, in distribution to a Poisson point process with an intensity measure depending on $\alpha$ and $\sum c_j^2$. This result is extended to random coefficient models where the coefficients of the linear processes $(X_{it})$ are given by $c_j(\theta_i)$, for some ergodic sequence $(\theta_i)$, and thus vary in each row of $X$. As a by-product of our techniques we obtain a proof of the corresponding result for matrices with iid entries in cases where $p/n$ goes to zero or infinity and $\alpha \in (0,2)$.

*Keywords:* Random Matrix Theory, heavy-tailed distribution, random matrix with dependent entries, largest singular value, sample covariance matrix, largest eigenvalue, linear process, random coefficient model
*2010 MSC:* primary: 60B20, 62G32; secondary: 60G55, 62H25

## 1. Introduction

Recently there has been increasing interest in studying *large dimensional data sets* that arise in finance, wireless communications, genetics and other fields. Patterns in these data can often be summarized by the *sample covariance matrix*, as done in multivariate regression and dimension reduction via factor analysis. Therefore, our objective is to study the asymptotic behavior of the eigenvalues $\lambda_{(1)} \geq \ldots \geq \lambda_{(p)}$ of a $p \times p$ sample covariance matrix $XX^\mathsf{T}$, where the *data matrix* $X$ is obtained from $n$ observations of a high-dimensional stochastic process with values in $\mathbb{R}^p$. Classical results in this direction often assume that the entries of $X$ are independent and identically distributed (iid) or satisfy some moment conditions. For example, the Four Moment Theorem of Tao and Vu [39] shows that the asymptotic behaviour of the eigenvalues of $XX^\mathsf{T}$ is determined by the first four moments of the distribution of the iid matrix entries of $X$. Our goal is to weaken the moment conditions by allowing for heavy-tails, and the assumption of independent entries by allowing for dependence within the rows and columns. Potential applications arise in portfolio management in finance, where observations typically have heavy-tails and dependence.

*Corresponding author
*Email addresses:* rdavis@stat.columbia.edu (Richard A. Davis), o.pfaffel@gmx.de (Oliver Pfaffel), robert.stelzer@uni-ulm.de (Robert Stelzer)

Assuming that the data comes from a multivariate normal distribution, one is able to compute the joint distribution of the eigenvalues $(\lambda_{(1)}, \dots, \lambda_{(p)})$, see [26]. Under the additional assumption that the dimension $p$ is fixed while the sample size $n$ goes to infinity, Anderson [2] obtains a central limit like theorem for the largest eigenvalue. Clearly, it is not possible to derive the joint distribution in a general setting where the distribution of $X$ is not invariant with respect to orthogonal transformations. Furthermore, since in modern applications with large dimensional data sets, $p$ might be of similar or even larger order than $n$, it might be more suitable to assume that both $p$ and $n$ go to infinity, so Anderson's result may not be a good approximation in this setting. For example, considering a financial index like the S&P 500, the number of stocks is $p = 500$, whereas, if daily returns of the past 5 years are given, $n$ is only around 1300. In genetic studies, the number of investigated genes $p$ might easily exceed the number of participating individuals $n$ by several orders of magnitude. In this *large n, large p* framework results differ dramatically from the corresponding *fixed p, large n* results - with major consequences for the statistical analysis of large data sets [27].

Spectral properties of large dimensional random matrices is one of many topics that has become known under the banner *Random Matrix Theory (RMT)*. The original motivation for RMT comes from mathematical physics [20], [42], where large random matrices serve as a finite-dimensional approximation of infinite-dimensional operators. Its importance for statistics comes from the fact that RMT may be used to correct traditional tests or estimators which fail in the 'large $n$, large $p$' setting. For example, Bai et al. [4] gives corrections on some likelihood ratio tests that fail even for moderate $p$ (around 20), and El Karoui [21] consistently estimates the spectrum of a large dimensional covariance matrix using RMT. Thus statistical considerations will be our motivation for a random matrix model with heavy-tailed and dependent entries.

Before describing our results, we will give a brief overview of some of the key results from RMT for real-valued sample covariance matrices $XX^\top$. A more detailed account on RMT can be found, for instance, in the textbooks [1], [5], or [31]. Here $X$ is a real $p \times n$ random matrix, and $p$ and $n$ go to infinity simultaneously. Let us first assume that the entries of $X$ are iid with variance 1. Results on the *global behavior of the eigenvalues* of $XX^\top$ mostly concern the *spectral distribution*, that is the random probability measure of its eigenvalues $p^{-1} \sum_{i=1}^{p} \epsilon_{n^{-1}\lambda_{(i)}}$, where $\epsilon$ denotes the Dirac measure. The spectral distribution converges, as $n, p \to \infty$ with $p/n \to \gamma \in (0, 1]$, to a deterministic measure with density function

$$\frac{1}{2\pi x\gamma}\sqrt{(x_+ - x)(x - x_-)}\mathbf{1}_{(x_-,x_+)}(x), \quad x_\pm := (1 \pm \sqrt{\gamma})^2,$$

where $\mathbf{1}$ denotes the indicator function. This is the so called *Marčenko–Pastur law* [30], [41]. One obtains a different result if $XX^\top$ is perturbed via an affine transformation [30], [33]. Partially based on these results, [6, 7, 35, 43] treat the case where the rows of $X$ are given by independent copies of a linear process (in a Gaussian setting this is a special case of results in [24]). [7] allows also for non-linear time series models describing the entries. Apart from a few special cases, the limiting spectral distribution is not known in a closed form if the entries of $X$ are not independent.

Although the eigenvalues of $XX^\top$ offer various interesting local properties to be studied, we will only focus on the joint asymptotic behavior of the $k$ largest eigenvalues $(\lambda_{(1)}, \dots, \lambda_{(k)})$, $k \in \mathbb{N}$. This is motivated from a statistical point of view since the variances of the first $k$ principal components are given by the $k$ largest eigenvalues of the covariance matrix. Geman [23] shows, assuming that the entries of $X$ are iid and have finite fourth moments, that $n^{-1}\lambda_{(1)}$ converges to $x_+ = (1 + \sqrt{\gamma})^2$ almost surely if $p/n \to \gamma \in (0, \infty)$. Moreover, if the entries of $X$ are iid standard Gaussian, Johnstone [27] shows that

$$\frac{\sqrt{n}+\sqrt{p}}{\sqrt[3]{\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{p}}}}\left(\frac{\lambda_{(1)}}{\left(\sqrt{n}+\sqrt{p}\right)^2}-1\right)\xrightarrow{D}\xi,$$

where $\xi$ follows the *Tracy–Widom distribution* with $\beta = 1$. Soshnikov [37] extends this to more general symmetric non-Gaussian distributions if the matrix $X$ is nearly square, and obtains a similar result for the joint convergence of the $k$ largest eigenvalues. The Tracy–Widom distribution first appeared as the limit of the largest eigenvalue of a Gaussian Wigner matrix [40]. Péché [34] shows that the assumption of Gaussianity in Johnstone's result can be replaced by the assumption that the entries of $X$ have a symmetric distribution with sub-Gaussian tails, and she allows for $\gamma$ being zero or infinity.

There exist results on extreme eigenvalues of $XX^\top$ which include dependence within the rows or columns of $X$, but most of them are only valid if $X$ has complex-valued entries such that its real as well as its complex part have a non-zero variance. A notable exception, where the real-valued case is considered, is [13]. They assume that the rows of $X$ are normally distributed with a covariance matrix which has exactly one eigenvalue not equal to one. To the best of our knowledge there are no results in the literature for the extreme eigenvalues of a random matrix with (light-tailed) dependent entries given by a general time series model.

In contrast to the light tailed case described above, there exist only a handful of articles dealing with sample covariance matrices $XX^\top$ obtained from heavy-tailed observations. Almost all these results only apply to matrices $X$ with iid entries. Cizeau and Bochaud [17] seems to be the first systematic investigation of the spectrum of heavy-tailed random matrices and Ben Arous and Guionnet [9] rigorously proved the convergence of the spectrum, i.e. a Marčenko–Pastur type result. Belinschi et al. [8] compute the limiting spectral distribution of sample covariance matrices based on observations with infinite variance, which is also investigated in Bordenave et al. [15]. Under a special dependence assumption Bordenave et al. [14] obtain also a Marčenko–Pastur type of result. Regarding the $k$-largest eigenvalues, Soshnikov [38] gives the weak limit in case the underlying distribution of the matrix entries is Cauchy. Biroli et al. [12] argued, using heuristic arguments and numerical simulations, that Soshnikov's result extends to general distributions with regularly varying tails with index $0 < \alpha < 4$. A mathematically rigorous proof of this claim followed by Auffinger et al. [3]. To the best of our knowledge the first results on the limiting behaviour of the bulk of the spectrum in the heavy-tailed case, where the entries are dependent and modelled via time series models, are given in Banna and Merlevède [7] who still assume finite second moments though.

We extend the previous results on the extreme eigenvalues for $0 < \alpha < 4$ by allowing for dependent entries. More specifically, the rows of $X$ are given by independent copies of some linear process. Their respective coefficients can either all be equal (Section 2.1) or, more generally, conditionally on a latent process, vary in each row (Section 2.3). In the latter case the rows of $X$ are not necessarily independent. The limiting Poisson process of the eigenvalues of $XX^\top$ depends on the tail index $\alpha$ as well as the coefficients of the observed linear processes. As a by-product, we obtain an independent proof of Soshnikov's result for iid entries which also holds in cases where $\gamma \in \{0, \infty\}$.

The paper is organized as follows. The main results will be presented in Section 2 while the proofs will be given in Section 3. Results from the theory of point processes and regular variation are required through most of this paper. A detailed account on both topics can be found in a number of texts. We mainly adopt the setting, including notation and terminology, of Resnick [36].

## 2. Main results on heavy-tailed random matrices with dependent entries

### *2.1. A first result on the largest eigenvalue*

Let $(Z_{it})_{i,t}$ be an array of iid random variables with marginal distribution that is regularly varying with tail index $\alpha > 0$ and *normalizing sequence* $a_n$, i.e.,

$$\lim_{n\to\infty} nP(|Z_{it}| > a_n x) = x^{-\alpha}, \quad \text{for each } x > 0. \tag{1}$$

Equivalently, this means that $(|Z_{it}|)$ is in the maximum domain of attraction of a Fréchet distribution with parameter $\alpha > 0$. The sequence $a_n$ is then necessarily characterized by

$$a_n = n^{1/\alpha} L(n), \tag{2}$$

for some slowly varying function $L : \mathbb{R}_+ \to \mathbb{R}_+$, i.e., a function with the property that, for each $x > 0$, $\lim_{t\to\infty} L(tx)/L(t) = 1$. In certain cases we also assume that $Z_{11}$ satisfies the *tail balancing condition*, i.e., the existence of the limits

$$\lim_{x\to\infty} \frac{P(Z_{11} > x)}{P(|Z_{11}| > x)} = q \quad \text{and} \quad \lim_{x\to\infty} \frac{P(Z_{11} \le -x)}{P(|Z_{11}| > x)} = 1 - q \tag{3}$$

for some $0 \le q \le 1$. For each $p, n \in \mathbb{N}$, let $X = (X_{it})$ be the $p \times n$ data matrix, where, for each $i$,

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j} \tag{4}$$

is a stationary linear times series. To guarantee that the series in (4) converges almost surely, we assume that

$$\sum_{j=-\infty}^{\infty} |c_j|^{\delta} < \infty \quad \text{for some } \delta < \min\{\alpha, 1\}. \tag{5}$$

Thus in our model the rows of $X$ are given by iid copies of a linear process. We denote by $\lambda_1, \dots, \lambda_p \ge 0$ the eigenvalues of the $p \times p$ sample covariance matrix $XX^\mathsf{T}$. They are studied via the induced point process

$$N_n = \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\lambda_i - n\mu_{X,\alpha})}, \tag{6}$$

where

$$\mu_{X,\alpha} = \begin{cases} 0 & \text{for } 0 < \alpha < 2, \\ E\left(Z_{11}^2 \mathbf{1}_{\{Z_{11}^2 \le a_{np}^2\}}\right) \sum_j c_j^2 & \text{for } \alpha = 2 \text{ and } EZ_{11}^2 = \infty, \\ E\left(Z_{11}^2\right) \sum_j c_j^2 & \text{else.} \end{cases} \tag{7}$$

Since we are only interested in the largest eigenvalues, we consider $N_n$ as a point process on $(0, \infty)$ and only count eigenvalues $\lambda_i$ which are positive. Observe that the centralization term $n\mu_{X,\alpha}$ is equal to the mean of the diagonal elements of $XX^\mathsf{T}$ if the observations have a finite variance. In case the observations have an infinite variance, we do not have to center, except when $\alpha = 2$ and $EZ_{11}^2 = \infty$, where we use a truncated version of the mean. In the latter case $\mu_{X,\alpha}$ also depends on $p$ and $n$.

We will always assume that $p = p_n$ is an integer-valued sequence in $n$ that goes to infinity as $n \to \infty$ in order to obtain results in the 'large $n$, large $p$' setting. In the following we suppress the dependence of $p$ on $n$ so as to simplify the notation wherever this does not cause any ambiguity. In [3, 38] the iid case is considered, i.e., $X_{it} = Z_{it}$, assuming that the condition (1) holds for $0 < \alpha < 4$. They show, if $p, n \to \infty$ with

$$\lim_{n\to\infty} \frac{p_n}{n} = \gamma \in (0, \infty), \tag{8}$$

that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_i} \xrightarrow[n\to\infty]{D} N, \tag{9}$$

where $N$ is a Poisson process with intensity measure $\hat{\nu}((x, \infty]) = x^{-\alpha/2}$. Our next theorem extends this result by considering the case where $X$ has dependent entries. More precisely, the rows of $X$ are given by independent copies of a linear process. It will turn out that the intensity measure of the limiting Poisson process depends on the sum of the squared coefficients of the underlying linear process. In contrast to [3], we necessarily have to center the eigenvalues $\lambda_i$ by $n\mu_{X,\alpha}$ when $\alpha \geq 2$, since in that case we consider a regime where $p \approx n^\beta$ with $\beta < 1$ instead of (8).

**Theorem 1.** *Define the matrix $X = (X_{it})$ as in equations (1), (4) and (5) with $\alpha \in (0, 4)$. Suppose $p_n \to \infty$ and $n \to \infty$ such that*

$$\limsup_{n\to\infty} \frac{p_n}{n^\beta} < \infty \tag{10}$$

*for some $\beta > 0$ satisfying*

*(i)* $\beta < \infty$ *if* $0 < \alpha \leq 1$, *and*
*(ii)* $\beta < \max\left\{\frac{2-\alpha}{\alpha-1}, \frac{1}{2}\right\}$ *if* $1 < \alpha < 2$.
*(iii)* $\beta < \max\left\{\frac{1}{3}, \frac{4-\alpha}{4(\alpha-1)}\right\}$ *if* $2 \leq \alpha < 3$, *or*
*(iv)* $\beta < \frac{4-\alpha}{3\alpha-4}$ *if* $3 \leq \alpha < 4$.

*Further assume, in the case $\alpha \in (5/3, 4)$, that $Z_{11}$ has mean zero and satisfies the tail balancing condition (3). Then the point process $N_n$, as defined in (6), converges in distribution to a Poisson point process $N$ with intensity measure $\nu$ which is given by*

$$\nu((x, \infty]) = x^{-\alpha/2} \left| \sum_{j=-\infty}^{\infty} c_j^2 \right|^{\alpha/2}, \quad x > 0.$$

Theorem 1 weakens the assumption of independent entries made so far in the literature on heavy-tailed random matrices at the expense of assumption (10), which is more restrictive than the usual assumption (8) if $\alpha \in [1.5, 4)$. It seems that this more restrictive assumption is necessary for our proof to work, but it is not clear whether it is necessary for the results to hold. However, if $\alpha \in (0, 1.5)$, our assumption (10) is more general than (8). This is important for statistical applications, because $p$ and $n$ are usually fixed and there is no functional relationship between the two of them.

If we restrict ourselves to the iid case, then Theorem 2 shows that the point process convergence result also holds in many cases where the limit $\gamma$ from condition (8) is zero or infinity, for example, by assuming that $p$ is regularly varying in $n$. This slightly extends the known theory for the iid case.

**Theorem 2.** *Assume that* $X_{it} = Z_{it}$ *and equation* (1) *is satisfied with* $\alpha \in (0, 2)$. *Further, let either*

(*i*) $p_n = n^\kappa l(n)$ *for some* $\kappa \in [0, \infty)$, *where* $l$ *is a slowly varying function which converges to infinity if* $\kappa = 0$, *and is bounded away from zero if* $\kappa = 1$, *or*
(*ii*) $p_n \sim C \exp(cn^\kappa)$ *for some* $\kappa, c, C > 0$.

*Then* $N_n$ *converges in distribution to a Poisson point process with intensity measure* $\hat{\nu}((x, \infty]) = x^{-\alpha/2}$.

It is well known [36] that a Poisson process has an explicit representation as a transformation of a homogeneous Poisson process. In our case, the limiting Poisson process $N$ with intensity measure $\nu$ from Theorem 1 can be written as

$$N \stackrel{D}{=} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2}, \tag{11}$$

where $\Gamma_i = \sum_{k=1}^{i} E_k$ is the successive sum of iid exponential random variables $E_k$ with mean one. The points of $N$ are labeled in decreasing order so that, by the continuous mapping theorem, we can easily deduce the weak limit of the (centered) $k$ largest eigenvalues of $XX^\top$.

**Corollary 1.** *Denote by* $\lambda_{(1)} \geq \ldots \geq \lambda_{(p)}$ *the upper order statistics of the eigenvalues of* $XX^\top - n\mu_{X,\alpha} I_p$. *Under the assumptions of Theorem 1 we have, for each fixed integer* $k \geq 1$, *that the k-largest eigenvalues jointly converge,*

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

*In particular, for each* $x > 0$,

$$P\left(\frac{\lambda_{(k)}}{a_{np}^2} \leq x\right) \xrightarrow[n\to\infty]{} P(N(x, \infty) \leq k - 1) = e^{-x^{-\alpha/2}} \sum_{m=0}^{k-1} \frac{x^{-m\alpha/2}}{m!} \left(\sum_j c_j^2\right)^{m\alpha/2}.$$

*This implies for the largest eigenvalue* $\lambda_{(1)}$ *of* $XX^\top - n\mu_{X,\alpha} I_p$ *that*

$$\frac{\lambda_{(1)}}{a_{np}^2 \sum_j c_j^2} \xrightarrow[n\to\infty]{D} V,$$

*where* $V$ *has a Fréchet distribution with parameter* $\alpha/2$, *i.e.*, $P(V \leq x) = e^{-x^{-\alpha/2}}$.

In a nutshell, the results in this section give the asymptotic behavior of the $k$ largest eigenvalues of a sample covariance matrix $XX^\top$ when the rows of $X$ are given by iid copies of some linear process with infinite variance. Our results will be generalized further in Section 2.3, where, conditionally on a latent process, the rows of $X$ will be independent but not identically distributed.

### *2.2. Examples and discussion*

Theorem 1 holds for any linear process which has regularly varying noise with infinite variance as long as condition (5) is satisfied. Since the coefficients of a causal ARMA process decay exponentially, (5) is trivially satisfied in this case. As two simple examples, consider an MA(1) process $X_{it} = Z_{it} + \theta Z_{i,t-1}$, which satisfies $\sum_j c_j^2 = 1 + \theta^2$; and a causal AR(1) process $X_{it} - \phi X_{i,t-1} = Z_{it}$, $|\phi| < 1$, where $\sum_j c_j^2 = (1-\phi^2)^{-1}$. Yet another example of a linear process fitting in our framework is a fractionally integrated ARMA$(p,d,q)$ processes with $d < 0$ and regularly varying noise with index $\alpha \in [1,4)$, see, e.g., [16] for further details. In this case $|c_j| \le C j^{d-1}$ is summable and therefore condition (5) is satisfied for $\alpha \ge 1$.

Regarding the normalization in (6), the sequence $a_n$ is chosen such that the individual entries of the matrix $Z := (Z_{it})_{i,t}$ satisfy (1). Replacing the iid sequence in the rows of $Z$ with a linear process to obtain the matrix $X$ changes the tail behavior of its entries. Indeed, the result stated in Davis and Resnick [19, eq. (2.7)] shows, under the assumption (3) and $EZ_{11} = 0$ if $\alpha > 1$, that

$$nP\left(\Big|\sum_j c_j Z_{1,t-j}\Big| > a_n x\right) \underset{n\to\infty}{\longrightarrow} x^{-\alpha}\sum_j |c_j|^\alpha.$$

In view of (1) this suggests the normalization $\tilde{X}_{it} = \left(\sum_j |c_j|^\alpha\right)^{-1/\alpha} X_{it}$. Denote by $\tilde{\lambda}_1,\dots,\tilde{\lambda}_p$ the eigenvalues of $\tilde{X}\tilde{X}^\intercal$, where $\tilde{X} = (\tilde{X}_{it})_{i,t}$, and let $\mu_{\tilde{X},\alpha} = E\tilde{X}_{11}^2 = \left(\sum_j |c_j|^\alpha\right)^{-2/\alpha}\mu_{X,\alpha}$. Since this is just a multiplication by a constant, we immediately obtain, by Theorem 1, that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\lambda_i - n\mu_{\tilde{X},\alpha})} \xrightarrow[n\to\infty]{D} \tilde{N},$$

where $\tilde{N}$ is a Poisson process with intensity measure $\tilde{\nu}$ given by

$$\tilde{\nu}((x,\infty]) = x^{-\alpha/2}\frac{\left|\sum_j c_j^2\right|^{\alpha/2}}{\sum_j |c_j|^\alpha}. \tag{12}$$

Thus $\left|\sum_j c_j^2\right|^{\alpha/2}(\sum_j |c_j|^\alpha)^{-1}$ quantifies the effect of the dependence on the point process of the eigenvalues when the tail behavior of each marginal $X_{it}$ is equivalent to the iid case.

Assume for a moment that the dimension $p$ is fixed for any $n$, and that $0 < \alpha < 2$. Then it follows easily from [19, Theorem 4.1] and arguments of our paper that $a_n^{-2}\lambda_{(1)} \to \sum_j c_j^2 \max_{1\le i\le p} S_i$ in distribution as $n \to \infty$, where $(S_i)$ are independent positive stable with index $\alpha/2$. If $p$ is large, one would intuitively expect that $\max_{1\le i\le p} S_i \approx p^{2/\alpha}\Gamma_1^{-2/\alpha}$, where $\Gamma_1$ is exponentially distributed with mean 1. Corollary 1 not only makes this intuition precise but also gives the correct normalization $a_{np}^{-2}$. The distribution of the maximum of $p$ independent stables is not known analytically, hence 'large n, large p' in fact gives a simpler solution than the traditional 'fixed p, large n' setting.

### *2.3. Extension to random coefficient models*

So far we have assumed that our observed process has independent components, each of which are modelled by the same linear process. From now on we will allow for a different set of coefficients in each row. To this end, let $(\theta_i)_{i\in\mathbb{N}}$ be a sequence of random variables *independent of* $(Z_{it})$ with values in some space $\Theta$. Assume that there is a family of measurable functions $(c_j : \Theta \to \mathbb{R})_{j\in\mathbb{N}}$ such that

$$\sup_{\theta\in\Theta} |c_j(\theta)| \le \widetilde{c}_j, \quad \text{for some deterministic } \widetilde{c}_j \text{ satisfying condition (5).} \tag{13}$$

Our observed processes have the form

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j(\theta_i) Z_{i,t-j} \tag{14}$$

where $(Z_{it})$ is given as in (1) with $\alpha \in (0, 4)$. Thus, conditionally on the latent process $(\theta_i)$, the rows of $X$ are independent linear processes with different coefficients. Unconditionally, the rows of $X$ are dependent if the sequence $(\theta_i)$ is dependent. Theorem 3 below covers three classes among which $(\theta_i)$ may be chosen: stationary ergodic; stationary but not necessarily ergodic; and ergodic in the Markov chain sense but not necessarily stationary. In the following we say that a sequence of point processes $\mathscr{M}_n$ converges, conditionally on a sigma-algebra $\mathcal{H}$, in distribution to a point process $\mathscr{M}$, if the conditional Laplace functionals converge almost surely, i.e., if there exists a measurable set $B$ with $P(B) = 1$ such that for all $\omega \in B$ and all nonnegative continuous functions $f$ with compact support,

$$E\left(e^{-\mathcal{M}_n(f)}\middle|\mathcal{H}\right)(\omega) \underset{n\to\infty}{\longrightarrow} E\left(e^{-\mathcal{M}(f)}\middle|\mathcal{H}\right)(\omega) \quad \text{as } n \to \infty. \tag{15}$$

**Theorem 3.** *Define $X = (X_{it})$ with $X_{it}$ as given in (14). Suppose that (13) is satisfied, and $p, n \to \infty$ such that (10) holds under the same conditions as in Theorem 1. Further assume, in case $\alpha \in (5/3, 4)$, that $Z_{11}$ has mean zero and satisfies the tail balancing condition (3).*

*(i) If $(\theta_i)$ is a stationary ergodic sequence, then, both conditionally on $(\theta_i)$ as well as unconditionally, we have that*

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\lambda_i - n\mu_{X,\alpha})} \overset{D}{\underset{n\to\infty}{\longrightarrow}} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha}\left\|\sum_j c_j^2(\theta_1)\right\|_{\frac{\alpha}{2}}}, \tag{16}$$

*with the constant $\left\|\sum_j c_j^2(\theta_1)\right\|_{\frac{\alpha}{2}} = \left(E\left|\sum_j c_j^2(\theta_1)\right|^{\alpha/2}\right)^{2/\alpha}$, and $(\Gamma_i)$ as in (11).*

*(ii) If $(\theta_i)$ is stationary but not necessarily ergodic, then we have, conditionally on $(\theta_i)$, that*

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\lambda_i - n\mu_{X,\alpha})} \overset{D}{\underset{n\to\infty}{\longrightarrow}} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} Y^{2/\alpha}},$$

*with $Y = E\left(|\sum_j c_j^2(\theta_1)|^{\alpha/2}|\mathcal{G}\right)$, where $\mathcal{G}$ is the invariant $\sigma$-field generated by $(\theta_i)$. In particular, $Y$ is independent of $(\Gamma_i)$.*

*(iii) Suppose $(\theta_i)$ is either an irreducible Markov chain on a countable state space $\Theta$ or a positive Harris chain in the sense of Meyn and Tweedie [32]. If $(\theta_i)$ has a stationary probability distribution $\pi$ then, conditionally on $(\theta_i)$ as well as unconditionally, (16) holds with*

$$\left\|\sum_j c_j^2(\theta_1)\right\|_{\frac{\alpha}{2}} = \left(\int_\Theta \left|\sum_j c_j^2(\theta)\right|^{\alpha/2} \pi(d\theta)\right)^{2/\alpha}.$$

One can view the assumptions (i) and (ii) of Theorem 3 in a Bayesian framework in which the parameters of the observed process are drawn from an unknown prior distribution. As an example, let $(\theta_i)$ be a stationary ergodic AR(1) process $\theta_i = \phi\theta_{i-1} + \xi_i$, where $|\phi| \neq 1$ and $(\xi_i)$ is a sequence of bounded iid random variables, and set $X_{it} = Z_{it} + \theta_i Z_{i,t-1}$. Then, by Theorem 3 (i), we would expect, for $n$ and $p$ large enough, that

$$a_{np}^{-2}(\lambda_{(1)} - n\mu_{X,\alpha}) \approx \Gamma_1^{-\alpha/2}\left(E|1+\theta_1|^{\alpha/2}\right)^{2/\alpha}.$$

Models of this kind are referred to as *random coefficient models* and often used in times series analysis, see, e.g., [29] for an overview. In the setting of Theorem 3 (iii) one might think of a *Hidden Markov Model* where the latent Markov process $(\theta_i)$ evolves along the rows of $X$, each state $\theta_i$ defining another univariate linear model.

## 3. Proofs and auxiliary results

The first step is to show that the matrix $XX^\mathsf{T}$ is well approximated by its diagonal, see Section 3.2. In the second step we then derive the extremes of the diagonal of $XX^\mathsf{T}$ in Section 3.3. Both steps together yield the proofs of Theorem 1 and Theorem 2 in Section 3.4. The proof of Theorem 3 follows then by an extension of the previous methods in Section 3.5. In the following we make frequent use of a large deviation result which is presented in the upcoming section.

### 3.1. A large deviation result and its consequences

The next theorem gives the joint large deviations of the sum and the maximum of iid nonnegative random variables with infinite variance. It suffices to deal with the case where $0 < \alpha < 2$ since later on we mostly consider squared random variables that have tail index $\alpha/2$ with $0 < \alpha/2 < 2$.

**Proposition 3.1.** *Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences of nonnegative numbers with $x_n \to \infty$ such that $x_n/y_n \to \gamma \in (0,\infty]$. Suppose $(Y_t)_{t\in\mathbb{N}}$ is an iid sequence of nonnegative random variables with tail index $\alpha \in (0,2)$ and normalizing sequence $b_n$. If $1 \le \alpha < 2$, we assume that $b_n x_n/n^{1+\delta} \to \infty$ for some $\delta > 0$. Then*

$$\lim_{n\to\infty} \frac{P\left(\sum_{t=1}^n Y_t > b_n x_n, \max_{1\le t\le n} Y_t > b_n y_n\right)}{nP(Y_1 > b_n \max\{x_n, y_n\})} = 1. \tag{17}$$

*Proof.* Let us first assume that $0 < \alpha < 1$. Using standard arguments from the theory of regularly varying functions, see e.g. [36], it can be easily seen that for any positive sequence $z_n \to \infty$ we have

$$\lim_{n\to\infty} \frac{P(\max_{1\le t\le n} Y_t > b_n z_n)}{nP(Y_1 > b_n z_n)} = 1. \tag{18}$$

Obviously the limit in (17) is greater or equal than one because $\sum_{t=1}^n Y_t \ge \max_{1\le t\le n} Y_t$. Thus it is only left to prove that it is also smaller. Denote by $Y_{(1)} \ge \ldots \ge Y_{(n)}$ the upper order statistics of $Y_1, \ldots, Y_n$. By decomposing $\sum_t Y_t$ into the sum of $\max_t Y_t$ and lower order terms we see that, for any $\theta \in (0,1)$,

$$\frac{P\left(\sum_{t=1}^n Y_t > b_n x_n, \max_{1\le t\le n} Y_t > b_n y_n\right)}{nP(Y_1 > b_n \max\{x_n, y_n\})} \le \frac{P(\max_{1\le t\le n} Y_t > b_n \max\{\theta x_n, y_n\})}{nP(Y_1 > b_n \max\{x_n, y_n\})} + \frac{P\left(\sum_{t=2}^n Y_{(t)} > b_n x_n(1-\theta)\right)}{nP(Y_1 > b_n \max\{x_n, y_n\})}.$$

By an application of [36, Proposition 0.8 (iii)] one can show similarly as in the proof of (18) that

$$\lim_{\theta\to 1}\lim_{n\to\infty} \frac{P(\max_{1\le t\le n} Y_t > b_n \max\{\theta x_n, y_n\})}{nP(Y_1 > b_n \max\{x_n, y_n\})} = 1.$$

Hence, it is only left to show that the second summand vanishes as $n \to \infty$. To this end we partition the underlying probability space into $\{Y_{(2)} \le \epsilon b_n x_n\} \cup \{Y_{(2)} > \epsilon b_n x_n\}$, $\epsilon > 0$, to obtain

$$\frac{P\left(\sum_{t=2}^n Y_{(t)} > b_n x_n(1-\theta)\right)}{nP(Y_1 > b_n \max\{x_n, y_n\})} \le \frac{P\left(\sum_{t=2}^n Y_{(t)} \mathbf{1}_{\{Y_{(2)}\le \epsilon b_n x_n\}} > b_n x_n(1-\theta)\right)}{nP(Y_1 > b_n \max\{x_n, y_n\})} + \frac{P\left(Y_{(2)} > \epsilon b_n x_n\right)}{nP(Y_1 > b_n \max\{x_n, y_n\})} = \Sigma_1 + \Sigma_2.$$

Denote by $M_n = \max_{1\le t\le n} Y_t$ and $z_n = \max\{x_n, y_n\}$. Then easy combinatorics and (18) yield

$$\begin{aligned}\Sigma_2 =&\frac{1-P\left(Y_{(2)}\le \epsilon b_n x_n\right)}{nP(Y_1>b_n z_n)}\\ =&\frac{1-P\left(M_n\le \epsilon b_n x_n\right)}{nP(Y_1>b_n z_n)}-\frac{nP\left(M_{n-1}\le \epsilon b_n x_n\right)P(Y_1>\epsilon b_n x_n)}{nP(Y_1>b_n z_n)}\\ =&\frac{P\left(M_n>\epsilon b_n x_n\right)}{nP(Y_1>\epsilon b_n x_n)}\frac{P(Y_1>\epsilon b_n x_n)}{P(Y_1>b_n z_n)}-\frac{P(Y_1>\epsilon b_n x_n)}{P(Y_1>b_n z_n)}P\left(M_{n-1}\le \epsilon b_n x_n\right)\\ \sim&\frac{P(Y_1>\epsilon b_n x_n)}{P(Y_1>b_n z_n)}\left(1-P\left(M_{n-1}\le \epsilon b_n x_n\right)\right)\underset{n\to\infty}{\longrightarrow}0.\end{aligned}$$

The convergence to zero follows from $P\left(M_{n-1}\le \epsilon b_n x_n\right)\to 1$ and, by [36, Proposition 0.8 (iii)], $\frac{P(Y_1>\epsilon b_n x_n)}{P(Y_1>b_n z_n)}\to \epsilon^{-\alpha}\max\{1,\gamma^{-\alpha}\}$. Thus it is only left to show that $\Sigma_1$ goes to zero. By Markov's inequality and Karamata's Theorem [36, Theorem 0.6] we have that

$$\begin{aligned}\Sigma_1 \le&\frac{P\left(\sum_{t=1}^n Y_t \mathbf{1}_{\{Y_t\le \epsilon b_n x_n\}}>b_n x_n(1-\theta)\right)}{nP(Y_1>b_n\max\{x_n,y_n\})}\le \frac{1}{b_n x_n(1-\theta)}\frac{E(Y_1\mathbf{1}_{\{Y_1\le \epsilon b_n x_n\}})}{P(Y_1>b_n z_n)}\\ \sim&\frac{1}{(1-\theta)}\frac{\alpha}{1-\alpha}\frac{\epsilon P(Y_1>\epsilon b_n x_n)}{P(Y_1>b_n z_n)}\underset{n\to\infty}{\longrightarrow}\frac{1}{(1-\theta)}\frac{\alpha}{1-\alpha}\epsilon^{1-\alpha}\max\{1,\gamma^{-\alpha}\},\end{aligned}$$

which converges to zero as $\epsilon$ goes to zero, since $\alpha<1$. Thus for $0<\alpha<1$ the proof is complete. If $1\le\alpha<2$, only $\Sigma_1$ has to be treated differently. The truncated mean $\mu_n = E(Y_1\mathbf{1}_{\{Y_1\le \epsilon b_n x_n\}})$ either converges to a constant or is a slowly varying function. In either case, we have that $b_n x_n/(n\mu_n) = b_n x_n n^{-1-\delta}\, n^{\delta}/\mu_n\to\infty$ by assumption. Thus, a mean-correction argument and Karamata's Theorem imply

$$\begin{aligned}\limsup_{n\to\infty}\Sigma_1 \le&\limsup_{n\to\infty}\frac{P\left(\sum_{t=1}^n Y_t\mathbf{1}_{\{Y_t\le \epsilon b_n x_n\}}-n\mu_n>b_n x_n(1-\theta)-n\mu_n\right)}{nP(Y_1>b_n z_n)}\\ \le&\frac{1}{(1-\theta)^2}\limsup_{n\to\infty}\frac{1}{b_n^2 x_n^2}\frac{\mathrm{Var}(Y_1\mathbf{1}_{\{Y_1\le \epsilon b_n x_n\}})}{P(Y_1>b_n z_n)}\le\frac{1}{(1-\theta)^2}\frac{\alpha/2}{1-\alpha/2}\epsilon^{2-\alpha}\max\{1,\gamma^{-\alpha}\}\underset{\epsilon\to 0}{\longrightarrow}0,\end{aligned}$$

since $\alpha<2$. This completes the proof. □

We finish this section with a few consequences of Proposition 3.1. Note that (1) implies

$$pnP(Z_{11}^2>a_{np}^2 x)\underset{n\to\infty}{\longrightarrow}x^{-\alpha/2}\quad\text{for each } x>0. \tag{19}$$

Choosing $Y_t = Z_{1t}^2$, $b_n = a_n^2$, $x_n = x a_{np}^2/a_n^2$ and $y_n = y a_{np}^2/a_n^2$, we have from Proposition 3.1 and (19), for $\alpha\in(0,2)$, that

$$pP\left(\sum_{t=1}^n Z_{1t}^2>a_{np}^2 x, \max_{1\le t\le n} Z_{1t}^2>a_{np}^2 y\right)\underset{n\to\infty}{\longrightarrow}\max\{x,y\}^{-\alpha/2}\quad\text{for each } x,y>0.$$

Therefore, by [36, Proposition 3.21], we obtain the point process convergence

$$\sum_{i=1}^p \epsilon_{a_{np}^{-2}(\sum_{t=1}^n Z_{it}^2,\max_{1\le t\le n} Z_{it}^2)}\overset{D}{\underset{n\to\infty}{\longrightarrow}}\sum_{i=1}^\infty \epsilon_{\Gamma_i^{-2/\alpha}(1,1)}, \tag{20}$$

with $(\Gamma_i)$ as in (11). Note that $\Gamma_i^{-2/\alpha}(1,1)$ is the point with both coordinates being $\Gamma_i^{-2/\alpha}$, i.e. the $-2/\alpha$th power of the random variable $\Gamma_i$. For another application of Proposition 3.1, set $Y_t = |Z_{1t}|$, $b_n = a_n$, $x_n = xa_{np}/a_n$ and $y_n = ya_{np}/a_n$. Under the additional assumption

$$\liminf_{n\to\infty} \frac{p}{n} \in (0,\infty]$$

we have $b_n x_n/n^{1+\gamma} \to \infty$ for some $\gamma < (2-\alpha)/\alpha$, thus, for $\alpha \in (0,2)$,

$$pP\left(\sum_{t=1}^n |Z_{1t}| > a_{np}x, \max_{1\le t\le n} |Z_{1t}| > a_{np}y\right) \underset{n\to\infty}{\longrightarrow} \max\{x,y\}^{-\alpha} \quad \text{for each } x,y>0.$$

Therefore we obtain as before

$$\sum_{i=1}^p \epsilon_{a_{np}^{-1}(\sum_{t=1}^n |Z_{it}|, \max_{1\le t\le n}|Z_{it}|)} \xrightarrow[n\to\infty]{D} \sum_{i=1}^\infty \epsilon_{\Gamma_i^{-1/\alpha}(1,1)}. \tag{21}$$

The result of the following proposition is also a consequence of Proposition 3.1.

**Proposition 3.2.** *Let $(Z_{it})$ be as in (1) with $0 < \alpha < 2$. Suppose that (10) is satisfied for some $0 < \beta < \infty$. Then*

$$a_{np}^{-2} \max_{1\le i<j\le p} \sum_{t=1}^n |Z_{it}Z_{jt}| \xrightarrow[n\to\infty]{P} 0.$$

*Proof.* By [22], the iid random variables $Y_t = |Z_{1t}Z_{2t}|$ are regularly varying with tail index $\alpha$ with some normalizing sequence $b_n$. Thus, there exists a slowly varying $L_1$ such that $P(Y_1 > x) = x^{-\alpha}L_1(x)$. Using (2) this implies

$$p^2nP(Y_1 > a_{np}^2\epsilon) = n^{-1}\epsilon^{-\alpha}\, L(np)^{-2\alpha}\, L_1\left((np)^{2/\alpha}L(np)^2\epsilon\right).$$

By Potter's bound, see, e.g., [36, Proposition 0.8 (ii)], for any slowly varying function $\tilde{L}$ and any $\delta > 0$ there exist $c_1, c_2 > 0$ such that $c_1 n^{-\delta} < \tilde{L}(n) < c_2 n^{\delta}$ for $n$ large enough. An application of this bound together with assumption (10) shows that

$$p^2nP(Y_1 > a_{np}^2\epsilon) \underset{n\to\infty}{\longrightarrow} 0. \tag{22}$$

Hence, using Proposition 3.1 with $x_n = a_{np}^2/b_n\epsilon$ and $y_n = 0$ yields

$$P\left(\max_{1\le i<j\le p} \sum_{t=1}^n |Z_{it}Z_{jt}| > a_{np}^2\epsilon\right) \le p^2P\left(\sum_{t=1}^n Y_t > a_{np}^2\epsilon\right) \underset{n\to\infty}{\longrightarrow} 0,$$

since $b_n x_n/n^{1+\gamma} = a_{np}^2/n^{1+\gamma} \to \infty$ for $\alpha < 2$ and some $\gamma < (2-\alpha)/\alpha$. □

*3.2. Convergence in Operator Norm*

Denote by $D = \mathrm{diag}(XX^\top)$ the diagonal of the matrix $XX^\top$, i.e., $D_{ii} = (XX^\top)_{ii}$ and $D_{ij} = 0$ for $i \ne j$. In this section we show that $a_{np}^{-2}(XX^\top - D)$ converges in probability to 0 in operator norm. This implies that the off-diagonal elements of $a_{np}^{-2}XX^\top$ do not contribute to the limiting eigenvalue spectrum. Recall that, for a real $p\times n$ matrix $A$, the operator 2-norm $\|A\|_2$ is the square root of the largest eigenvalue of $AA^\top$, and the infinity-norm is given by $\|A\|_\infty = \max_{1\le i\le p}\sum_{t=1}^n |A_{it}|$.

In the upcoming Proposition 3.3 we only deal with the case where $0 < \alpha < 2$. Note that Proposition 3.3 holds under a much more general setting than assumed in Theorem 1 by allowing for an arbitrary dependence structure within the rows of $X$.

**Proposition 3.3.** *Let $X = (X_{it})_{i,t}$ be a $p \times n$ random matrix whose entries are identically distributed with tail index $\alpha \in (0, 2)$ and normalizing sequence $(a_n)$. Assume that the rows of $X$ are independent. Suppose that (10) holds for some $\beta > 0$. If $1 < \alpha < 2$, assume additionally that $\beta < \frac{2-\alpha}{\alpha-1}$. Then we have*

$$a_{np}^{-2} \left\| XX^{\top} - D \right\|_2 \xrightarrow[n\to\infty]{P} 0. \tag{23}$$

*Proof.* Since $\left\|XX^{\top} - D\right\|_2 \le \left\|XX^{\top} - D\right\|_\infty$, it is enough to show that for every $\epsilon \in (0, 1)$,

$$P\left( \max_{i=1,\dots,p} \sum_{\substack{j=1\\ j\neq i}}^{p} \left| \sum_{t=1}^{n} X_{it}X_{jt} \right| > a_{np}^2 \epsilon \right) \le pP\left( \sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}| > a_{np}^2\epsilon \right) \xrightarrow[n\to\infty]{} 0.$$

By partitioning the underlying probability space into $\{\max_{j,t} |X_{1t}X_{jt}| \le a_{np}^2\}$ and its complement, we obtain that

$$\begin{aligned} pP\left( \sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}| > a_{np}^2\epsilon \right) \le & pP\left( \sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}| \mathbf{1}_{\{|X_{1t}X_{jt}| \le a_{np}^2\}} > a_{np}^2\epsilon \right) \\ & + pP\left( \max_{2\le j\le p} \max_{1\le t\le n} |X_{1t}X_{jt}| > a_{np}^2\epsilon \right) = \mathrm{I} + \mathrm{II}. \end{aligned}$$

The same argument used for (22) shows that $\mathrm{II} \le p^2 n P(|X_{11}X_{21}| > a_{np}^2) \xrightarrow[n\to\infty]{} 0$ by independence of the rows of $X$. To deal with term I we first assume that $\alpha > 1$ and choose some $\gamma \in (\alpha, 2)$. Hölder's inequality shows that

$$\left( \sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}| \right)^{\gamma} \le \left( \sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}|^{\gamma} \right) (np)^{\gamma-1},$$

and therefore

$$\mathrm{I} \le pP\left( \sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}|^{\gamma} \mathbf{1}_{\{|X_{1t}X_{jt}| \le a_{np}^2\}} > \frac{a_{np}^{2\gamma}}{(np)^{\gamma-1}} \epsilon \right).$$

Note that $|X_{1t}X_{jt}|^{\gamma}$ has regularly varying tails with index $\alpha/\gamma < 1$. Hence we can apply Markov's Inequality and Karamata's Theorem to infer that

$$\mathrm{I} \le c_1 \frac{p^2 n (np)^{\gamma-1}}{a_{np}^{2\gamma}} E\left( |X_{11}X_{21}|^{\gamma} \mathbf{1}_{\{|X_{11}X_{21}| \le a_{np}^2\}} \right) \sim c_2 p^2 n (np)^{\gamma-1} P(|X_{11}X_{21}| > a_{np}^2). \tag{24}$$

Therefore, the proof of Proposition 3.2 shows that the term in (24) goes to zero if $(np)^{\gamma-1}/n$ does. In view of assumption (10) this is true for $\beta < (2-\gamma)/(\gamma-1)$. Since we can choose $\gamma$ arbitrary close to $\alpha$ it suffices that $\beta < (2-\alpha)/(\alpha-1)$. If $\alpha < 1$ we do not need Hölder's inequality since the above argument can be applied with $\gamma = 1$, thus it suffices that (10) holds for some $\beta < \infty$. For the remaining case $\alpha = 1$, observe that, for any given $\beta < \infty$, we choose $\gamma$ arbitrarily close to 1 so that $(np)^{\gamma-1}/n \to 0$. □

The next proposition improves the previous Proposition 3.3 for $5/3 < \alpha < 2$ by relaxing the condition $\beta < \frac{2-\alpha}{\alpha-1}$ at the expense of the additional assumption that the rows of $X$ are realizations of a linear process. Furthermore, Proposition 3.4 also covers the case where $2 \le \alpha < 4$.

**Proposition 3.4.** *The assumptions of Theorem 1 imply* (23).

*Proof.* Considering the result from Proposition 3.3, we only have to deal with the case $\alpha \in (5/3, 4)$ here. In this proof, $c$ denotes a positive constant that may vary from expression to expression. Define

$$Z_{it}^L = Z_{it}\mathbf{1}_{\{|Z_{it}|\le a_{np}\}}, \quad X_{it}^L = \sum_k c_k Z_{i,t-k}^L,$$
$$Z_{it}^U = Z_{it}\mathbf{1}_{\{|Z_{it}|> a_{np}\}}, \quad X_{it}^U = \sum_k c_k Z_{i,t-k}^U.$$

Using $\left\|XX^\top - D\right\|_2 \le \left\|XX^\top - D\right\|_\infty$ as before we have

$$\begin{aligned} P\left(\left\|XX^\top - D\right\|_2 > a_{np}^2\epsilon\right) \le & pP\left(\sum_{j=2}^p\left|\sum_{t=1}^n X_{1t}X_{jt}\right| > a_{np}^2\epsilon\right) \\ \le & pP\left(\sum_{j=2}^p\left|\sum_{t=1}^n X_{1t}^L X_{jt}^L\right| > \frac{a_{np}^2}{4}\epsilon\right) + pP\left(\sum_{j=2}^p\left|\sum_{t=1}^n X_{1t}^L X_{jt}^U\right| > \frac{a_{np}^2}{4}\epsilon\right) \\ & + pP\left(\sum_{j=2}^p\left|\sum_{t=1}^n X_{1t}^U X_{jt}^L\right| > \frac{a_{np}^2}{4}\epsilon\right) + pP\left(\sum_{j=2}^p\left|\sum_{t=1}^n X_{1t}^U X_{jt}^U\right| > \frac{a_{np}^2}{4}\epsilon\right) \\ = & \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}. \end{aligned}$$

We will show that each of theses terms converges to zero. To this end, note that $E|Z_{11}^L|$ converges to a constant, and, by Karamata's Theorem,

$$E|Z_{11}^U| \sim c a_{np} P(|Z_{11}| > a_{np}) \sim c a_{np}(np)^{-1}, \quad n \to \infty.$$

Therefore, by Markov's inequality, we have

$$\mathrm{II} \le \frac{4p}{a_{np}^2\epsilon}\sum_{j=2}^p\sum_{t=1}^n\sum_{k,l}|c_k c_l|\, E|Z_{1,t-k}^L|\, E|Z_{j,t-l}^U| \sim c\left(\sum_k |c_k|\right)^2 \frac{p^2 n}{a_{np}^2} a_{np}(np)^{-1} = c\frac{p}{a_{np}},$$

and, by (2), we obtain that this is equal to $c\, L(np)^{-1} p^{1-1/\alpha} n^{-1/\alpha} \to 0$ as $n \to \infty$. By symmetry, III can be handled the same way. It is easy to see that term IV is of even lower order, namely

$$c\frac{p^2 n}{a_{np}^2}(a_{np}(np)^{-1})^2 = cn^{-1} \to 0.$$

Thus it is only left to show that I converges to zero. To this end, we use Karamata's Theorem to obtain

$$E\left[(Z_{11}^L)^2\right] = E\left[Z_{11}^2\mathbf{1}_{\{|Z_{11}|\le a_{np}\}}\right] \sim c a_{np}^2 P(|Z_{11}| > a_{np}) \sim c a_{np}^2 (np)^{-1}.$$

Since $Z_{11}$ satisfies the tail balancing condition (3), and $EZ_{11} = 0$, we can apply Karamata's Theorem to the positive and the negative tail of $Z_{11}^L$, thus, for $q \notin \{0, \frac{1}{2}, 1\}$,

$$\begin{aligned} \xi_n := & E[Z_{11}^L] = E[Z_{11}\mathbf{1}_{\{|Z_{11}|\le a_{np}\}}] = -E[Z_{11}\mathbf{1}_{\{|Z_{11}|> a_{np}\}}] = -E[Z_{11}\mathbf{1}_{\{Z_{11}> a_{np}\}}] + E[-Z_{11}\mathbf{1}_{\{-Z_{11}> a_{np}\}}] \\ \sim & -q\frac{\alpha}{\alpha-1}a_{np}P(|Z_{11}| > a_{np}) + (1-q)\frac{\alpha}{\alpha-1}a_{np}P(|Z_{11}| > a_{np}) \sim (1-2q)\frac{\alpha}{\alpha-1}a_{np}(np)^{-1}. \end{aligned}$$

Clearly, for any $0 \le q \le 1$, one therefore has

$$\frac{np\xi_n}{a_{np}} \to (1-2q)\frac{\alpha}{\alpha-1}.$$

As a consequence we obtain for $\mu_n = E(X_{11}^L X_{21}^L) = (EX_{11}^L)^2 = \xi_n^2$ that

$$\frac{\mu_n pn}{a_{np}^2} = (np)^{-1}\left(\frac{np\xi_n}{a_{np}}\right)^2\left(\sum_k c_k\right)^2 \to 0.$$

Therefore we obtain for summand I that

$$\mathrm{I} = pP\left(\sum_{j=2}^{p}\left|\sum_{t=1}^{n} X_{1t}^L X_{jt}^L\right| > \frac{a_{np}^2}{4}\epsilon\right) \le p^2 P\left(\left|\sum_{t=1}^{n} X_{1t}^L X_{2t}^L\right| > \frac{a_{np}^2}{4p}\epsilon\right) \sim p^2 P\left(\left|\sum_{t=1}^{n} X_{1t}^L X_{2t}^L - n\mu_n\right| > \frac{a_{np}^2}{4p}\epsilon\right).$$

Therefore, it is only left to show that

$$p^2 P\left(\left|\sum_{t=1}^{n} X_{1t}^L X_{2t}^L - n\mu_n\right| > \frac{a_{np}^2}{4p}\epsilon\right) \to 0, \tag{25}$$

with

$$\mu_n = (EX_{11}^L)^2 = \left(\sum_k c_k\right)^2 (EZ_{11}^L)^2 = O\left(\frac{a_{np}^2}{(np)^2}\right).$$

Now we have to treat the cases $\alpha < 2$, $2 \le \alpha < 3$, and $3 \le \alpha < 4$ separately.

Let $\alpha < 2$. By Proposition 3.3 it suffices to show that, for $\alpha \in (5/3, 2)$, the assumption

$$\lim_{n\to\infty} \frac{p}{\sqrt{n}} = 0 \tag{26}$$

implies convergence in operator norm in the sense of (23). Since we correct by the mean, Markov's inequality yields

$$\begin{aligned} p^2 P\left(\left|\sum_{t=1}^{n} X_{1t}^L X_{2t}^L - n\mu_n\right| > \frac{a_{np}^2}{4p}\epsilon\right) \le& \frac{16p^4}{a_{np}^4\epsilon^2}\mathrm{Var}\left(\sum_{t=1}^{n} X_{1t}^L X_{2t}^L\right) \\ =& \frac{16p^4}{a_{np}^4\epsilon^2}\sum_{t,t'=1}^{n}\sum_{k,k',l,l'} c_k c_{k'} c_l c_{l'} \mathrm{Cov}\left(Z_{1,t-k}^L Z_{2,t-l}^L, Z_{1,t'-k'}^L Z_{2,t'-l'}^L\right). \end{aligned} \tag{27}$$

Due to the independence of the $Z$'s, the covariance in the last expression is non-zero iff $t-k = t'-k'$ or $t-l = t'-l'$. This gives us three distinct cases we deal with separately. First, assume that both $t-k = t'-k'$ and $t-l = t'-l'$. Then the covariance in (27) is equal to $\mathrm{Var}(Z_{11}^L Z_{2,1}^L)$ and so bounded by

$$E[(Z_{11}^L)^2 (Z_{21}^L)^2] = (E(Z_{11}^L)^2)^2 \sim (ca_{np}^2(np)^{-1})^2 \sim ca_{np}^4(np)^{-2}.$$

Second, let $t-k = t'-k'$ but $t-l \ne t'-l'$. Then the covariance becomes

$$\begin{aligned} \mathrm{Cov}(Z_{1,t-k}^L Z_{2,t-l}^L, Z_{1,t'-k'}^L Z_{2,t'-l'}^L) =& E((Z_{1,t-k}^L)^2 Z_{2,t-l}^L Z_{2,t'-l'}^L) - \xi_n^4 \\ =& E((Z_{1,t-k}^L)^2)\xi_n^2 - \xi_n^4 \\ \sim& ca_{np}^2(np)^{-1}(\pm c\, a_{np}(np)^{-1})^2 - (\pm c\, a_{np}(np)^{-1})^4 \\ \sim& ca_{np}^4(np)^{-3} - ca_{np}^4(np)^{-4} \sim ca_{np}^4(np)^{-3}, \end{aligned}$$

which is of lower order than in the case considered before. By symmetry, the third case, where $t-l=t'-l'$ but $t-k\neq t'-k'$, can be dealt with in exactly the same way. In all cases $t'$ can be assumed to be fixed, thus we can bound (27) by

$$c\frac{p^4}{a_{np}^4}\left(\sum_k |c_k|\right)^4\sum_{t=1}^n a_{np}^4(np)^{-2}=c\frac{p^2}{n}\to 0,\quad n\to\infty.$$

This completes the proof in case $\alpha<2$.

If $\alpha>2$, the covariance in (27) converges to a constant. If $\alpha=2$ with $EZ_{11}^2=\infty$, then it is a slowly varying function. In either case (27) is of order

$$O\left(\frac{p^4}{a_{np}^4}n\,s(np)\right)\leq O\left(n^{\beta(4-4/\alpha)}n^{1-4/\alpha}s(np)\right)\to 0,$$

since $\beta<\frac{4-\alpha}{4(\alpha-1)}$, where $s(\cdot)$ is some slowly varying function. For a more general result we distinguish the sub-cases $\alpha\in(2,3)$ and $\alpha\in[3,4)$ in the following.

Let us now assume that $\alpha\in(2,3)$. By Markov's inequality applied to (25) we have

$$\begin{aligned}
&p^2P\left(\left|\sum_{t=1}^n X_{1t}^LX_{2t}^L-n\mu_n\right|>\frac{a_{np}^2}{4p}\epsilon\right)\\
\leq&\frac{64}{\epsilon^3}\frac{p^5}{a_{np}^6}\sum_{t_1,t_2,t_3=1}^n E\left(\prod_{i=1}^3\left(X_{1,t_i}^LX_{2,t_i}^L-\mu_n\right)\right)\\
=&\frac{64}{\epsilon^3}\frac{p^5}{a_{np}^6}\sum_{t_1,t_2,t_3=1}^n\sum_{k_1,k_2,k_3}\sum_{l_1,l_2,l_3}\prod_{j=1}^3(c_{k_j}c_{l_j})E\left(\prod_{i=1}^3\left(Z_{1,t_i-k_i}^LZ_{2,t_i-l_i}^L-\xi_n^2\right)\right),
\end{aligned}\tag{28}$$

where

$$\xi_n^2=\frac{\mu_n}{(\sum_k c_k)^2}=(EZ_{11}^L)^2=O\left(\frac{a_{np}^2}{(np)^2}\right).\tag{29}$$

To determine the order of the expectation in (28) we have to distinguish various cases. In the following we say that two index pairs $(a,b)$ and $(c,d)$ overlap if $a=c$ or $b=d$. If there exists a $j=1,2,3$ such that the index pair $(t_j-k_j,t_j-l_j)$ does not overlap with both the other two, then, due to independence, we are able to factor out the corresponding term and obtain

$$E\left(\prod_{i=1}^3\left(Z_{1,t_i-k_i}^LZ_{2,t_i-l_i}^L-\xi_n^2\right)\right)=E\left(\prod_{i\neq j}\left(Z_{1,t_i-k_i}^LZ_{2,t_i-l_i}^L-\xi_n^2\right)\right)E\left(Z_{1,t_j-k_j}^LZ_{2,t_j-l_j}^L-\xi_n^2\right)=0,$$

since $\xi_n^2=(EZ_{11}^L)^2=E\left(Z_{1,t_j-k_j}^LZ_{2,t_j-l_j}^L\right)$. Thus, in any non-trivial case, each index pair does overlap with (at least) one of the other two. Therefore we have at least two equalities of the form $t_i-k_i=t_{(i+1)\bmod 3}-k_{(i+1)\bmod 3}$ or $t_i-l_i=t_{(i+1)\bmod 3}-l_{(i+1)\bmod 3}$ for $i=1,2,3$. Hence $t_2$ and $t_3$ are immediately determined by some linear combination of $t=t_1$ and the $k_i's$ or $l_i's$. Therefore the triple sum $\sum_{t_1,t_2,t_3=1}^n$ is, if we only count terms where the covariance is non-zero, in fact a simple sum $\sum_{t=1}^n$ and so only has a contribution of order $n$. Now we have to determine the order of the products $E\left(\prod_{i=1}^3 Z_{1,t_i-k_i}^LZ_{2,t_i-l_i}^L\right)$. If we only have a single power then, by (29), this gives us

$$E\left(Z_{1,t_i-k_i}^LZ_{2,t_i-l_i}^L\right)=\xi_n^2=o(1).$$

Since $\alpha > 2$, powers of order two converge to a constant,

$$E\left(\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i}\right)^2\right) \to \mathrm{Var}(Z_{11})^2.$$

An application of Karamata's theorem yields that

$$E\left(\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i}\right)^3\right) \sim a^6_{np}(np)^{-2}.$$

Using the above facts, it is easy to see that

$$E\left(\prod_{i=1}^{3} Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i}\right) = O\left(a^6_{np}(np)^{-2}\right). \tag{30}$$

Thus we have, using (29) and (30), for the expectation in (28) that

$$\begin{aligned} E\left(\prod_{i=1}^{3}\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i} - \xi_n^2\right)\right) &= \sum_{k=0}^{3}(-1)^k \sum_{J\subseteq\{1,2,3\},|J|=k} E\left(\prod_{i\in\{1,2,3\}\setminus J} Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i}\right)\xi_n^{2|J|} \\ &= O\left(a^6_{np}(np)^{-2} - \frac{a^2_{np}}{(np)^2} - \frac{a^6_{np}}{(np)^6}\right) \\ &= O\left(a^6_{np}(np)^{-2}\right). \end{aligned}$$

The last calculation shows that the expectation in (28) is equal to $E\left(\prod_{i=1}^3 Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i}\right)$ plus lower order terms, and that the leading term is of order $a^6_{np}(np)^{-2}$. With this observation we can finally conclude for (28) that

$$\begin{aligned} &\frac{64}{\epsilon^3}\frac{p^5}{a^6_{np}}\sum_{t_1,t_2,t_3=1}^{n}\sum_{k_1,k_2,k_3}\sum_{l_1,l_2,l_3}\prod_{j=1}^{3}(c_{k_j}c_{l_j})E\left(\prod_{i=1}^{3}\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i} - \xi_n^2\right)\right) \\ &= O\left(\frac{64}{\epsilon^3}\frac{p^5}{a^6_{np}} n a^6_{np}(np)^{-2}\right) = \frac{64}{\epsilon^3}O\left(\frac{p^3}{n}\right) \to 0, \end{aligned}$$

which goes to zero by assumption. This completes the proof for $\alpha \in [2,3)$.

The method to deal with $\alpha \in [3,4)$ is similar to the one before and thus only described briefly. We use Markov's inequality with power four to obtain that the term in (25) is bounded by

$$\frac{256}{\epsilon^4}\frac{p^6}{a^8_{np}}\sum_{t_1,t_2,t_3,t_4=1}^{n}\sum_{k_1,k_2,k_3,k_4}\sum_{l_1,l_2,l_3,l_4}\prod_{j=1}^{4}(c_{k_j}c_{l_j})E\left(\prod_{i=1}^{4}\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i} - \xi_n^2\right)\right). \tag{31}$$

Observe that the expectation in (31) is only non-zero if either

($i$) all index pairs $\{(t_i - k_i, t_i - l_i)\}_{i=1,2,3,4}$ overlap, or

($ii$) there exist exactly two sets of overlapping index pairs, such that no index pair from one set overlaps with an index pair from the other set. We call these two sets disjoint.

Case (i) is similar to the previous case, so that one can see that

$$E\left(\prod_{i=1}^{4}\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i} - \xi_n^2\right)\right) = O\left((E((Z^L_{11})^4))^2\right) = O\left(\frac{a^8_{np}}{(np)^2}\right),$$

and that the contribution of $\sum_{t_1,t_2,t_3,t_4=1}^n$ is of order $n$. Therefore, in this case, the term in (31) is of the order

$$\frac{256}{\epsilon^4}\frac{p^6}{a_{np}^8}O\left(n\frac{a_{np}^8}{(np)^2}\right)=\frac{256}{\epsilon^4}O\left(\frac{p^4}{n}\right)\to 0.$$

Thus, we only have to determine the contribution in case (ii). Since the two sets of overlapping index pairs are disjoint, we obtain that

$$E\left(\prod_{i=1}^4\left(Z^L_{1,t_i-k_i}Z^L_{2,t_i-l_i}-\xi_n^2\right)\right)=E\left(\left(Z^L_{11}Z^L_{21}-\xi_n^2\right)^2\right)^2.$$

Since $\alpha>2$ this converges to a constant. In contrast to case (i), the contribution of $\sum_{t_1,t_2,t_3,t_4=1}^n$ is of order $n^2$. This is due to the fact that the two sets of overlapping index pairs are disjoint, hence only two out of the four indices $t_1,\dots,t_4$ are given by linear combinations of the other two and the $k's$ and $l's$. Therefore (31) is of the order

$$\frac{256}{\epsilon^4}\frac{p^6}{a_{np}^8}O\left(n^2\right)\to 0.$$

The convergence to zero is justified by

$$\frac{p^6}{a_{np}^8}n^2=n^{2-8/\alpha}p^{6-8/\alpha}L(np)^{-8}\le O\left(n^{2-8/\alpha+\beta(6-8/\alpha)}L(n^{\beta+1})^{-8}\right)\to 0,$$

since $\beta<\frac{4-\alpha}{3\alpha-4}$. This completes the proof of Proposition 3.4. □

### 3.3. *Extremes on the diagonal*

In this section we analyze the extremes of the diagonal entries of $XX^T$, which are partial sums of squares of linear processes. To this end, we start with two auxiliary results. While Lemma 3.1 is only valid for $\alpha<2$, Lemma 3.2 covers the case where $2\le\alpha<4$. Subsequently, these two lemmas help us to establish a general limit theorem for the diagonal entries of $XX^T$ for $0<\alpha<4$ in Proposition 3.5, the major result of this section.

**Lemma 3.1.** *Let $(Z_t)$ be an iid sequence such that $nP(|Z_1|>a_nx)\to x^{-\alpha}$ with $\alpha\in(0,2)$. For any sequence $(c_j)$ satisfying (5) we have, if $p$ and $n$ go to infinity, that*

$$pP\left(\sum_{t=1}^n\sum_{j=-\infty}^{\infty}c_j^2Z_{t-j}^2>a_{np}^2x\right)\to\left(\sum_{j=-\infty}^{\infty}c_j^2\right)^{\frac{\alpha}{2}}x^{-\alpha/2}$$

*Proof.* Fix some $x>0$. Observe that Proposition 3.1 and (19) imply for $n\to\infty$ that $pP(\sum_{t=1}^n Z_t^2>a_{np}^2x)\to x^{-\alpha/2}$. We begin by showing the claim for a linear process of finite order. For any $\eta>0$ we have

$$P\left(\left|\sum_{j=-m}^m c_j^2\sum_{t=1}^n Z_t^2-\sum_{t=1}^n\sum_{j=-m}^m c_j^2Z_{t-j}^2\right|>a_{np}^2\eta\right)\le P\left(\sum_{j=-m}^m c_j^2\sum_{t=1-j}^{j}Z_t^2>a_{np}^2\eta\right)\underset{n\to\infty}{\longrightarrow}0.$$

Consequently,

$$\lim_{n\to\infty}pP\left(\sum_{t=1}^n\sum_{j=-m}^m c_j^2Z_{t-j}^2>a_{np}^2x\right)=x^{-\alpha/2}\left(\sum_{j=-m}^m c_j^2\right)^{\frac{\alpha}{2}}. \tag{32}$$

This and the positivity of the summands implies

$$\liminf_{n\to\infty} pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \geq x^{-\alpha/2}\left(\sum_{j=-\infty}^{\infty} c_j^2\right)^{\frac{\alpha}{2}}. \tag{33}$$

Thus it is only left to show that the limsup is bounded by the right hand side of (33). Using Markov's inequality yields

$$pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \leq \sum_{j=-\infty}^{\infty} pnP\left(c_j^2 Z_1^2 > a_{np}^2 x\right) + \sum_{j=-\infty}^{\infty} c_j^2 \frac{pn}{a_{np}^2 x} E\left(Z_1^2 \mathbf{1}_{\{c_j^2 Z_1^2 \leq a_{np}^2 x\}}\right).$$

Since $E\left(Z_1^2 \mathbf{1}_{\{Z_1^2 \leq \cdot\}}\right)$ is a regularly varying function with index $\alpha/2 - 1$ we obtain, by Potter's bound, Karamata's Theorem and (5), that, for some constant $C_1 > 0$,

$$\begin{aligned} c_j^2 \frac{pn}{a_{np}^2 x} E\left(Z_1^2 \mathbf{1}_{\{c_j^2 Z_1^2 \leq a_{np}^2 x\}}\right) &= \frac{c_j^2}{x} \frac{E\left(Z_1^2 \mathbf{1}_{\{c_j^2 Z_1^2 \leq a_{np}^2 x\}}\right)}{E\left(Z_1^2 \mathbf{1}_{\{Z_1^2 \leq a_{np}^2 x\}}\right)} \frac{pn}{a_{np}^2} E\left(Z_1^2 \mathbf{1}_{\{Z_1^2 \leq a_{np}^2 x\}}\right) \\ &\leq C_1 \frac{c_j^2}{x} \left(c_j^{-2}\right)^{1-\alpha/2+(\alpha/2-\delta/2)} x^{1-\alpha/2} = C_1 x^{-\alpha/2} |c_j|^{\delta}. \end{aligned}$$

Likewise, $pnP(a_{np}^{-2} Z_1^2 > \cdot)$ is a regularly varying function with index $\alpha/2$, thus we obtain, by the same arguments as before, that

$$pnP\left(c_j^2 Z_1^2 > a_{np}^2 x\right) \leq C_2 x^{-\alpha/2} |c_j|^{\delta}.$$

With $C = C_1 + C_2$ this therefore implies

$$\limsup_{n\to\infty} pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \leq C \sum_{j=-\infty}^{\infty} |c_j|^{\delta} x^{-\alpha/2}. \tag{34}$$

Hence, by (32) and (34), we finally have, for some $\epsilon \in (0,1)$, that

$$\begin{aligned} \limsup_{n\to\infty} pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) &\leq \limsup_{n\to\infty} pP\left(\sum_{t=1}^{n}\sum_{j=-m}^{m} c_j^2 Z_{t-j}^2 > (1-2\epsilon) a_{np}^2 x\right) \\ &+ \limsup_{n\to\infty} pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty} c_j^2 Z_{t-j}^2 > \epsilon a_{np}^2 x\right) + \limsup_{n\to\infty} pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{-m-1} c_j^2 Z_{t-j}^2 > \epsilon a_{np}^2 x\right) \\ &\leq x^{-\alpha/2}\left((1-2\epsilon)^{-\alpha/2}\left(\sum_{j=-m}^{m} c_j^2\right)^{\frac{\alpha}{2}} + C\epsilon^{-\alpha/2}\sum_{j=m+1}^{\infty}|c_j|^{\delta} + C\epsilon^{-\alpha/2}\sum_{j=-\infty}^{-m-1}|c_j|^{\delta}\right). \end{aligned} \tag{35}$$

Assumption (5) shows that the last two terms in (35) vanish for $m \to \infty$. Letting $\epsilon \to 0$ thereafter completes the proof. □

For $2 \leq \alpha < 4$ and $m$-dependence, we state the upcoming lemma.

**Lemma 3.2.** *Assume that there exists an $m \in \mathbb{N}$ such that $c_j = 0$ if $|j| > m$. Then we have, for $2 \le \alpha < 4$ and $p, n$ going to infinity, that*

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\sum_{t=1}^{n} X_{it}^2 - n\mu_{X,\alpha})} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-m}^{m} c_j^2}. \tag{36}$$

*Proof.* Note that we replace $\mu_{X,\alpha}$ by $\mu_X$ in the following to simplify the notation. For any iid sequence $(Z_t)$ with tail index $2 < \alpha < 4$ we have that

$$pP\left(\sum_{t=1}^{n} Z_t^2 - n\mu_Z > a_{np}^2 x\right) \to x^{-\alpha/2} \tag{37}$$

where $\mu_Z = EZ_1^2$. Indeed, [25], and in greater generality also [18], show that, for any $x > 0$,

$$\frac{P\left(\sum_{t=1}^{n} Z_t^2 - n\mu_Z > a_{np}^2 x\right)}{nP\left(Z_1^2 - \mu_Z > a_{np}^2 x\right)} \to 1. \tag{38}$$

With $P\left(Z_1^2 - \mu_Z > a_{np}^2 x\right) \sim P\left(Z_1^2 > a_{np}^2 x\right) \sim p^{-1} x^{-\alpha/2}$, the result follows. Note that (38) also holds for $\alpha = 2$ if $EZ_{11}^2 < \infty$. In case $EZ_{11}^2 = \infty$ (which can only happen if $\alpha = 2$), one has to replace $\mu_Z$ by the sequence of truncated means $\mu_z^n = E(Z_{11}^2 \mathbf{1}_{\{Z_{11}^2 \le a_{np}^2\}})$. For notational simplicity, we exclude infinite variance case in the following. It is treated analogously to the finite variance case, except that everywhere $\mu_Z$ has to be replaced by $\mu_Z^n$, $\mu_X$ by $\mu_X^n = \sum_k c_k^2 \mu_Z^n$, and finally $\mu_{X,m}$ by $\mu_{X,m}^n = \sum_{|k| \le m} c_k^2 \mu_Z^n$. By the stationarity of the $Z$'s we have that

$$\begin{aligned} &P\left(\left|\sum_{j=-m}^{m} c_j^2 \sum_{t=1}^{n} (Z_{1,t}^2 - \mu_Z) - \sum_{t=1}^{n} \sum_{j=-m}^{m} c_j^2 (Z_{1,t-j}^2 - \mu_Z)\right| > a_{np}^2 \eta\right) \\ &\le P\left(\sum_{j=-m}^{m} c_j^2 \sum_{t=1-j}^{j} Z_{1,t}^2 > a_{np}^2 \eta\right) \to 0. \end{aligned}$$

Hence, using (37), this yields

$$pP\left(\sum_{t=1}^{n} \sum_{j=-m}^{m} c_j^2 (Z_{1,t-j}^2 - \mu_Z) > a_{np}^2 x\right) \to x^{-\alpha/2} \left|\sum_{j=-m}^{m} c_j^2\right|^{\alpha/2}. \tag{39}$$

This immediately implies that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\left(\sum_{t=1}^{n} \sum_{j=-m}^{m} c_j^2 (Z_{i,t-j}^2 - \mu_Z)\right)} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \left|\sum_{j=-m}^{m} c_j^2\right|^{\alpha/2}}.$$

Thus it is only left to show that, for any continuous $f : \mathbb{R}_+ \to \mathbb{R}_+$ with compact support,

$$\lim_{n\to\infty} P\left(\sum_{i=1}^{p} \left| f\left(a_{np}^{-2}\left(\sum_{t=1}^{n} X_{it}^2 - n\mu_X\right)\right) - f\left(a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=-m}^{m} c_j^2 (Z_{i,t-j}^2 - \mu_Z)\right)\right| > \eta\right) = 0.$$

For convenience, we define $f(x) = 0$ if $x \le 0$. Clearly, we have that

$$\left|\sum_{t=1}^{n} X_{it}^2 - n\mu_X - \sum_{t=1}^{n}\sum_{j=-m}^{m} c_j^2 (Z_{i,t-j}^2 - \mu_Z)\right| \le 2 \sum_{j=-m}^{m-1}\sum_{k=j+1}^{m} |c_j c_k| \left|\sum_{t=1}^{n} Z_{i,t-j} Z_{i,t-k}\right|.$$

Hence, it suffices to show that

$$a_{np}^{-2} \max_{1\le i\le p} \left|\sum_{t\in J_s} Z_{i,t-j} Z_{i,t-k}\right| \to 0$$

for each fixed $j \in \{-m, \dots, m-1\}$, $k \in \{j+1, \dots, m\}$ and $s \in \{0, \dots, k-j\}$, where $J_s := s + (k-j+1)\mathbb{N}_0$. Note that $(Z_{i,t-j} Z_{i,t-k})_{t\in J_s}$ is a sequence of iid random variables with mean zero. Therefore we have, by Markov's inequality,

$$\begin{aligned} P\left(\max_{1\le i\le p} \left|\sum_{t\in J_s} Z_{i,t-j} Z_{i,t-k}\right| > a_{np}^2 \eta\right) \le & pP\left(\left|\sum_{t\in J_s} Z_{1,t-j} Z_{1,t-k}\right| > a_{np}^2 \eta\right) \\ \le & \frac{p}{\eta^2 a_{np}^4} \sum_{t\in J_s} \mathrm{Var}(Z_{1,t-j} Z_{1,t-k}) \\ \le & \frac{pn}{\eta^2 a_{np}^4} (EZ_{11}^2)^2 \\ = & O\left(\frac{pn}{a_{np}^4}\right) = O\left((pn)^{1-4/\alpha} L(pn)^{-4}\right) \to 0 \end{aligned}$$

since $\alpha < 4$. □

Now we prove the major result of this section, that is, the point process convergence of the diagonal elements of the sample covariance $XX^\top$ (or its centered version). This indirectly characterizes the extremal behavior of the $k$-largest diagonal entries of $XX^\top$. Note that Proposition 3.5 holds for any $0 < \beta < \infty$ in (10) independently of $\alpha \in (0, 4)$.

**Proposition 3.5.** *Let $0 < \alpha < 4$ and suppose that (10) holds for some $\beta > 0$. Then we have that*

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\sum_{t=1}^{n} X_{it}^2 - n\mu_{X,\alpha})} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2} \tag{40}$$

*with $\mu_{X,\alpha}$ and $(\Gamma_i)$ as given in (7) and (11), respectively.*

*Proof.* For notational simplicity we assume without loss of generality that $X_{it} = \sum_{j=0}^{\infty} c_j Z_{i,t-j}$, and write $\mu_X = \mu_{X,\alpha}$. The extension to the non-causal case is obvious.

We begin with the case of $0 < \alpha < 2$. First we prove the claim for finite linear processes $X_{it,m} = \sum_{j=0}^{m} c_j Z_{i,t-j}$. From Lemma 3.1 we already have that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=0}^{m} c_j^2 Z_{i,t-j}^2} \xrightarrow[n\to\infty]{D} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=0}^{m} c_j^2}. \tag{41}$$

Thus it is only left to show that all terms involving cross products are negligible. By [28, Theorem 4.2] it suffices to show, for any $\eta > 0$, that

$$\lim_{n\to\infty} P\left(\sum_{i=1}^{p}\left|f\left(a_{np}^{-2}\sum_{t=1}^{n}X_{it,m}^2\right)-f\left(a_{np}^{-2}\sum_{t=1}^{n}\sum_{j=0}^{m}c_j^2Z_{i,t-j}^2\right)\right|>\eta\right)=0 \tag{42}$$

for any continuous function $f:\mathbb{R}_+\to\mathbb{R}_+$ with compact support $\mathrm{supp}(f)\subset[c,\infty]$ and $c>0$. Choose some $0<\gamma<c$ and let $K=[c-\gamma,\infty]$. On the set

$$A_n^{\gamma}=\left\{\max_{1\le i\le p}\left|\sum_{t=1}^{n}X_{it,m}^2-\sum_{t=1}^{n}\sum_{j=0}^{m}c_j^2Z_{i,t-j}^2\right|\le a_{np}^2\gamma\right\}$$

the following is true: if $a_{np}^{-2}\sum_{t=1}^{n}\sum_{j=0}^{m}c_j^2Z_{i,t-j}^2\notin K$, then the absolute difference in (42) is zero, else it is bounded by the modulus of continuity $\omega(\gamma)=\sup\{|f(x)-f(y)|:|x-y|\le\gamma\}$. Hence, the probability in (42) is bounded by

$$P\left(\omega(\gamma)\sum_{i=1}^{p}\epsilon_{a_{np}^{-2}\sum_{t=1}^{n}\sum_{j=0}^{m}c_j^2Z_{i,t-j}^2}(K)>\eta\right)+P\left(\left(A_n^{\gamma}\right)^c\right).$$

By (41), the first summand converges to

$$P\left(\omega(\gamma)\sum_{i=1}^{\infty}\epsilon_{\sum_{j=0}^{m}c_j^2\Gamma_i^{-2/\alpha}}(K)>\eta\right).$$

Since $\sum_{i=1}^{\infty}\epsilon_{\sum_{j=0}^{m}c_j^2\Gamma_i^{-2/\alpha}}(K)<\infty$ and $\omega(\gamma)\to0$ as $\gamma\to0$, this probability approaches zero as $\gamma$ tends to zero. To show that

$$P\left(\left(A_n^{\gamma}\right)^c\right)\le P\left(2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m}|c_jc_k|\max_{i=1:p}\sum_{t=1}^{n}|Z_{i,t-j}Z_{i,t-k}|>a_{np}^2\gamma\right)\underset{n\to\infty}{\longrightarrow}0 \tag{43}$$

we use the following observation for fixed $j\in\{0,\dots,m-1\}$ and $k\in\{j+1,\dots,m\}$: the product $Z_{i,t-j}Z_{i,t-k}$ has, because of independence, tail index $\alpha$, and $Z_{i,t-j}Z_{i,t-k}$ and $Z_{i,s-j}Z_{i,s-k}$ are independent if and only if $|s-t|\ne k-j$. Thus, we partition the natural numbers $\mathbb{N}$ into $k-j+1$ pairwise disjoint sets $s+(k-j+1)\mathbb{N}_0$, $s\in\{0,\dots,k-j\}$. Then we have, by Proposition 3.2 and the independence of the summands, that

$$a_{np}^{-2}\max_{1\le i\le p}\sum_{t\in s+(k-j+1)\mathbb{N}_0}|Z_{i,t-j}Z_{i,t-k}|\xrightarrow[n\to\infty]{P}0,$$

for each $s\in\{0,\dots,k-j\}$. Since $j,k$ only vary over finite sets this implies (43). Therefore we have shown (40) for a finite order moving average $X_{it,m}$.

Now we let $m$ go to infinity. Clearly, we have that

$$\sum_{i=1}^{\infty}\epsilon_{\Gamma_i^{-2/\alpha}\sum_{j=0}^{m}c_j^2}\xrightarrow[m\to\infty]{D}\sum_{i=1}^{\infty}\epsilon_{\Gamma_i^{-2/\alpha}\sum_{j=0}^{\infty}c_j^2}. \tag{44}$$

Thus, by [11, Theorem 3.2], it is only left to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty}P\left(\sum_{i=1}^{p}\left|f\left(a_{np}^{-2}\sum_{t=1}^{n}X_{it}^2\right)-f\left(a_{np}^{-2}\sum_{t=1}^{n}X_{it,m}^2\right)\right|>\eta\right)=0. \tag{45}$$

By repeating the previous arguments, it suffices to show

$$\limsup_{n\to\infty} P\left(a_{np}^{-2}\max_{1\le i\le p}\sum_{t=1}^{n}|X_{it}^2 - X_{it,m}^2| > \gamma\right) \le \limsup_{n\to\infty} pP\left(a_{np}^{-2}\sum_{t=1}^{n}|X_{1t}^2 - X_{1t,m}^2| > \gamma\right) \to 0,$$

as $m \to \infty$. Clearly, we have that

$$X_{1t}^2 - X_{1t,m}^2 = \sum_{j=m+1}^{\infty} c_j^2 Z_{1,t-j}^2 + 2\sum_{j=m+1}^{\infty}\sum_{k=0}^{m} c_j c_k Z_{1,t-j}Z_{1,t-k} + \sum_{j=m+1}^{\infty}\sum_{\substack{k=m+1\\k\neq j}}^{\infty} c_j c_k Z_{1,t-j}Z_{1,t-k}. \tag{46}$$

For the first summand on the right hand side of equation (46) we have, by Lemma 3.1, that

$$pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty} c_j^2 Z_{1,t-j}^2 > \eta a_{np}^2\right) \underset{n\to\infty}{\longrightarrow} \left(\sum_{j=m+1}^{\infty} c_j^2\right)^{\alpha/2} \eta^{-\alpha/2} \underset{m\to\infty}{\longrightarrow} 0.$$

Using Lemma 3.1 and the elementary inequality $2|ab| \le a^2 + b^2$, we obtain for the second term in equation (46) that

$$\begin{aligned} pP\left(2\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}\sum_{k=0}^{m}|c_j c_k Z_{1,t-j}Z_{1,t-k}| > \eta a_{np}^2\right) &\le pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}\sum_{k=0}^{m}|c_j c_k| Z_{1,t-j}^2 > \frac{\eta}{2}a_{np}^2\right) \\ + pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}\sum_{k=0}^{m}|c_j c_k| Z_{1,t-k}^2 > \frac{\eta}{2}a_{np}^2\right) &\sim 2\frac{\eta}{4}^{-\alpha/2}\left(\sum_{k=0}^{m}|c_k|\right)^{\alpha/2}\left(\sum_{j=m+1}^{\infty}|c_j|\right)^{\alpha/2}, \end{aligned}$$

and since $\sum_{j=0}^{\infty}|c_j| < \infty$, this term converges to zero as $m \to \infty$. The third term in equation (46) can be handled similarly. Thus the proof is complete for $0 < \alpha < 2$.

For $2 \le \alpha < 4$, Lemma 3.2 gives us the result for a finite moving average. Thus it is only left to show that to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\left(\sum_{i=1}^{p}\left|f\Big(a_{np}^{-2}\sum_{t=1}^{n}(X_{it}^2 - \mu_X)\Big) - f\Big(a_{np}^{-2}\sum_{t=1}^{n}(X_{it,m}^2 - \mu_{X,m})\Big)\right| > \gamma\right) = 0$$

for any continuous $f$ with compact support and $\gamma > 0$. By the arguments given before it suffices to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty} pP\left(\left|\sum_{t=1}^{n}(X_{1t}^2 - X_{1t,m}^2 - (\mu_X - \mu_{X,m}))\right| > a_{np}^2\gamma\right) = 0.$$

Clearly, we have that

$$\begin{aligned} &pP\left(\left|\sum_{t=1}^{n}(X_{1t}^2 - X_{1t,m}^2 - (\mu_X - \mu_{X,m}))\right| > a_{np}^2\gamma\right) \\ \le\, &pP\left(\left|\sum_{k=m+1}^{\infty} c_k^2\sum_{t=1}^{n}(Z_{1,t-k}^2 - \mu_Z)\right| > a_{np}^2\frac{\gamma}{3}\right) \\ &+ pP\left(2\left|\sum_{k=m+1}^{\infty}\sum_{l=0}^{m} c_k c_l\sum_{t=1}^{n} Z_{1,t-k}Z_{1,t-l}\right| > a_{np}^2\frac{\gamma}{3}\right) \\ &+ pP\left(2\left|\sum_{k=m+1}^{\infty}\sum_{l=k+1}^{\infty} c_k c_l\sum_{t=1}^{n} Z_{1,t-k}Z_{1,t-l}\right| > a_{np}^2\frac{\gamma}{3}\right) \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III} \end{aligned}$$

We will show in turn that I, II, III $\to$ 0. We begin with I. Clearly, there either exist a $t$ and a $k$ such that $|c_k Z_{1,t-k} > a_{np}|$, or $|c_k Z_{1,t-k} \le a_{np}|$ for all $t, k$. This simple fact and Chebyshev's inequality yield

$$\begin{aligned}
\mathrm{I} =& pP\left(\left|\sum_{k=m+1}^{\infty} c_k^2 \sum_{t=1}^{n}(Z_{1,t-k}^2 - \mu_Z)\right| > a_{np}^2 \frac{\gamma}{3}\right) \\
\le & \sum_{k=m+1}^{\infty} pnP(|c_k Z_{1,1-k} > a_{np}|) + \frac{3}{\gamma}\frac{p}{a_{np}^4}\mathrm{Var}\left(\sum_{k=m+1}^{\infty} c_k^2 \sum_{t=1}^{n} Z_{1,t-k}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}|\le a_{np}\}}\right) \\
&+ p\mathbf{1}_{\left\{\sum_{k=m+1}^{\infty} c_k^2 nE\left(Z_{11}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}|>a_{np}\}}\right) > a_{np}^2 \frac{\gamma}{3}\right\}} = \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3
\end{aligned}$$

For the first term we have by Karamata's theorem that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{I}_1 = \lim_{m\to\infty} \sum_{k=m+1}^{\infty} c_k^{\alpha} = 0.$$

Another application of Karamata's theorem shows that

$$E\left(Z_{11}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}|>a_{np}\}}\right) \sim |c_k|^{\alpha/2-1}\frac{a_{np}^2}{np},$$

therefore

$$\lim_{n\to\infty} \frac{p\mathbf{1}_{\left\{\sum_{k=m+1}^{\infty} c_k^{\alpha/2+1} > p\frac{\gamma}{3}\right\}}}{\mathrm{I}_3} = 1.$$

However, $p\mathbf{1}_{\left\{\sum_{k=m+1}^{\infty} c_k^{\alpha/2+1} > p\frac{\gamma}{3}\right\}} = 0$ for $n$ sufficiently large, since $p = p_n \to \infty$ and

$$\sum_{k=m+1}^{\infty} c_k^{\alpha/2+1} < \infty.$$

As a consequence, $\mathrm{I}_3 \to 0$. Regarding $\mathrm{I}_2$, observe that the covariance in

$$\mathrm{I}_2 = \frac{3}{\gamma}\frac{p}{a_{np}^4} \sum_{k=m+1}^{\infty}\sum_{k'=m+1}^{\infty} c_k^2 c_{k'}^2 \sum_{t=1}^{n}\sum_{t'=1}^{n} \mathrm{Cov}\left(Z_{1,t-k}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}|\le a_{np}\}}, Z_{1,t'-k'}^2 \mathbf{1}_{\{|c_k' Z_{1,t'-k'}|\le a_{np}\}}\right)$$

is zero if $t - k \ne t' - k'$. In the case of equality, $t - k = t' - k'$, we have that

$$\begin{aligned}
&\sum_{t=1}^{n}\sum_{t'=1}^{n} \mathrm{Cov}\left(Z_{1,t-k}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}|\le a_{np}\}}, Z_{1,t'-k'}^2 \mathbf{1}_{\{|c_k' Z_{1,t'-k'}|\le a_{np}\}}\right) \\
&= \sum_{t=1}^{n} \mathrm{Var}\left(Z_{1,t-k}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}|\le a_{np}\}} \mathbf{1}_{\{|c_k' Z_{1,t'-k'}|\le a_{np}\}}\right) \\
&\le nE\left(Z_{1,1-k}^4 \mathbf{1}_{\{|\min\{c_k, c_{k'}\} Z_{1,1-k}|\le a_{np}\}}\right)
\end{aligned}$$

Using Karamata's theorem and Potter's bound we obtain that there exists a $C > 0$ and an $\epsilon > 0$ such that

$$\frac{pn}{a_{np}^4} E\left(Z_{1,1-k}^4 \mathbf{1}_{\{|\min\{c_k, c_{k'}\} Z_{1,1-k}|\le a_{np}\}}\right) \le C \min\{c_k, c_{k'}\}^{\alpha/4-\epsilon-1}.$$

For $m$ sufficiently large the coefficients become smaller than one, thus

$$\min\{c_k, c_{k'}\}^{\alpha/4-\epsilon-1} \le c_k^{\alpha/4-\epsilon-1} c_{k'}^{\alpha/4-\epsilon-1}.$$

All in all we obtain

$$\lim_{m\to\infty}\limsup_{n\to\infty} \mathrm{I}_2 \le \frac{3C}{\gamma}\lim_{m\to\infty}\left(\sum_{k=m+1}^{\infty} c_k^{1+\alpha/4-\epsilon}\right)^2 = 0,$$

since $\sum_{k=0}^{\infty} c_k < \infty$. For the second term observe that it follows, using Chebyshev's inequality, $EZ_{11} = 0$ and the independence of the $Z$'s, that

$$\begin{aligned}
\mathrm{II} =& pP\left(2\left|\sum_{k=m+1}^{\infty}\sum_{l=0}^{m} c_k c_l \sum_{t=1}^{n} Z_{1,t-k}Z_{1,t-l}\right| > a_{np}^2\frac{\gamma}{3}\right)\\
\le& \frac{6}{\gamma}\frac{p}{a_{np}^4}\mathrm{Var}\left(\sum_{k=m+1}^{\infty}\sum_{l=0}^{m} c_k c_l \sum_{t=1}^{n} Z_{1,t-k}Z_{1,t-l}\right)\\
=& \frac{6}{\gamma}\frac{p}{a_{np}^4}\sum_{k,k'=m+1}^{\infty}\sum_{l,l'=0}^{m} c_k c_{k'} c_l c_{l'}\sum_{t,t'=1}^{n} E\left(Z_{1,t-k}Z_{1,t-l}Z_{1,t'-k'}Z_{1,t'-l'}\right)\\
\le& \frac{6}{\gamma}\frac{p}{a_{np}^4}\sum_{k,k'=m+1}^{\infty}\sum_{l,l'=0}^{m} c_k c_{k'} c_l c_{l'} nE\left(Z_{11}^2\right)^2\\
\le& O\left(\left(\sum_{k=m+1}^{\infty} c_k\right)^2\left(\sum_{l=0}^{m} c_l\right)^2\frac{pn}{a_{np}^4}\right)\underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}$$

since $2 < \alpha < 4$. The remaining term III can be dealt with similarly to the previous term II. Hence the proof is complete. □

### 3.4. *Proofs of Theorem 1 and Theorem 2*

In this section we use the foregoing results from Section 3.1, Section 3.2 and Section 3.3 to complete the proofs of Theorem 1 and Theorem 2.

*Proof of Theorem 1.* Denote by $S_k = (XX^\top)_{kk} = \sum_{t=1}^{n} X_{kt}^2$ the diagonal entries of $XX^\top$. Recall that $\lambda_{(1)} \ge \ldots \ge \lambda_{(p)}$ are the upper order statistics of the eigenvalues of $XX^\top - n\mu_{X,\alpha}I_p$ with $\mu_{X,\alpha}$ as given in (7). Similarly we denote by $S_{(1)} \ge \ldots \ge S_{(p)}$ the upper order statistics of $S_k - n\mu_{X,\alpha} = \sum_{t=1}^{n} X_{kt}^2 - n\mu_{X,\alpha}$. Weyl's Inequality, cf. [10, Corollary III.2.6], and Proposition 3.4 imply that

$$a_{np}^{-2}\max_{1\le k\le p}|\lambda_{(k)} - S_{(k)}| = a_{np}^{-2}\max_{1\le k\le p}|\lambda_k - S_k| \le a_{np}^{-2}\left\|XX^\top - D\right\|_2 \overset{P}{\underset{n\to\infty}{\longrightarrow}} 0, \tag{47}$$

where $D = \mathrm{diag}(XX^\top)$. From Proposition 3.5 we have

$$\widehat{N}_n = \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(S_i - n\mu_{X,\alpha})} = \sum_{i=1}^{p} \epsilon_{a_{np}^{-2}S_{(i)}} \overset{D}{\underset{n\to\infty}{\longrightarrow}} N. \tag{48}$$

Thus, by [28, Theorem 4.2], it suffices to show that

$$P(|\widehat{N}_n(f) - N_n(f)| > \eta) \le P\left(\sum_{i=1}^{p}\left|f\left(\frac{S_{(i)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(i)}}{a_{np}^2}\right)\right| > \eta\right)\underset{n\to\infty}{\longrightarrow} 0$$

for a nonnegative continuous function $f$ with compact support $\mathrm{supp}(f) \subset [c,\infty]$, for some $c > 0$. For convenience we set $f(x) = 0$ if $x \le 0$. Since $N((c/2,\infty]) < \infty$ almost surely, we can choose some $i \in \mathbb{N}$ large enough such that the probability $P(N((c/2,\infty]) \geq i) < \delta/2$. By (48), it follows that $P(a_{np}^{-2} S_{(i)} > c/2) = P(\widehat{N}_n((c/2,\infty]) \geq i) \to P(N((c/2,\infty]) \geq i)$ and thus, for $n$ large enough, $P(a_{np}^{-2} S_{(i)} > c/2) < \delta$. Consequently, by (47), it follows that $P(a_{np}^{-2}\lambda_{(i)} \geq c) < 2\delta$. Since $a_{np}^{-2} S_{(i)} \leq c/2$ and $a_{np}^{-2}\lambda_{(i)} < c$ imply that both $f(a_{np}^{-2} M_{(k)}) = 0$ and $f(a_{np}^{-2}\lambda_{(k)}) = 0$ for all $k \geq i$, we obtain

$$
\begin{aligned}
P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^2}\right)\right| > \eta\right) \leq & P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^2}\right)\right| > \eta, a_{np}^{-2} S_{(i)} > \frac{c}{2}\right) \\
& + P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^2}\right)\right| > \eta, a_{np}^{-2} S_{(i)} \leq \frac{c}{2}, a_{np}^{-2}\lambda_{(i)} \geq c\right) \\
& + P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^2}\right)\right| > \eta, a_{np}^{-2} S_{(i)} \leq \frac{c}{2}, a_{np}^{-2}\lambda_{(i)} < c\right) \\
\leq & 3\delta + P\left(\sum_{j=1}^{i-1}\left|f\left(\frac{S_{(j)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^2}\right)\right| > \eta\right),
\end{aligned}
$$

which becomes arbitrarily small due to equation (47) and the fact that $f$ is uniformly continuous. □

In the case when the entries of $X$ are iid and have tail index $\alpha < 2$, we can refine our techniques to weaken the assumptions on the growth of $p = p_n$, cf. Theorem 2.

*Proof of Theorem 2.* By assumption $X = (Z_{it})$. First we consider the case (i) and assume that $\kappa \geq 1$. We will show that, for any fixed positive integer $k$,

$$
\frac{\lambda_{(k)}}{S_{(k)}} \xrightarrow[n\to\infty]{P} 1. \tag{49}
$$

Equations (20) and (49) then imply

$$
\left|\frac{S_{(k)}}{a_{np}^2} - \frac{\lambda_{(k)}}{a_{np}^2}\right| = \left|1 - \frac{\lambda_{(k)}}{S_{(k)}}\right| \frac{S_{(k)}}{a_{np}^2} \xrightarrow[n\to\infty]{P} 0,
$$

and hence $N_n \to N$ as in the proof of Theorem 1. Define $M_i = \max_{1\leq t\leq n} X_{it}^2$ and denote by $M_{(1)} \geq \ldots \geq M_{(p)}$ the upper order statistics of $M_1, \ldots, M_p$. Observe that the continuous mapping theorem applied to (20) and (21) yields, for any fixed $k$,

$$
\frac{S_{(k)}}{M_{(k)}} \xrightarrow[n\to\infty]{P} 1, \quad \text{and} \quad \frac{\|X\|_\infty^2}{M_{(1)}} \xrightarrow[n\to\infty]{P} 1,
$$

because $\kappa \geq 1$. Now we start showing (49) by induction. For $k = 1$ we have, on the one hand, that

$$
\frac{\lambda_{(1)}}{S_{(1)}} = \frac{\left\|X_n X_n^\top\right\|_2}{S_{(1)}} \leq \frac{\|X_n\|_2^2}{S_{(1)}} \leq \frac{\|X_n\|_\infty^2}{S_{(1)}} = \frac{\|X_n\|_\infty^2}{M_{(1)}} \frac{M_{(1)}}{S_{(1)}} \xrightarrow[n\to\infty]{P} 1.
$$

Let us denote by $e_1, \ldots, e_p$ the standard Euclidean orthonormal basis in $\mathbb{R}^p$ and by $i_1$ the (random) index that satisfies $S_{i_1} = S_{(1)}$. Then we have, on the other hand, by the Minimax Principle [10, Corollary III.1.2], that

$$
\frac{\lambda_{(1)}}{S_{(1)}} = \frac{\max_{v\in\mathbb{R}^p}\left\langle v, XX^\top v\right\rangle}{S_{(1)}} \geq \frac{\left\langle e_{i_1}, XX^\top e_{i_1}\right\rangle}{S_{(1)}} = \frac{S_{i_1}}{S_{(1)}} = 1.
$$

This shows (49) for $k = 1$. To keep the notation simple, we describe the induction step only for $k = 2$. The arguments for the general case are exactly the same. Denote by $i_2$ the random index such that $S_{i_2} = S_{(2)}$. Let $X^{(2)}$ be the $(p-1)\times n$ matrix which is obtained from removing row $i_1$ from $X_n$ and denote by $\varrho_{(1)}$ the largest eigenvalue of $X^{(2)}(X^{(2)})^\top$. Since we have already shown the claim for the largest eigenvalue, it follows that $\varrho_{(1)}/S_{(2)} \to 1$ in probability. By the Cauchy Interlacing Theorem [10, Corollary III.1.5] this implies $\lambda_{(2)}/S_{(2)} \leq \varrho_{(1)}/S_{(2)} \to 1$. Another application of the Minimax Principle yields

$$\lambda_{(2)} = \max_{\substack{\mathcal{M}\subset\mathbb{R}^p\\ \dim(\mathcal{M})=2}} \min_{\substack{v\in\mathcal{M}\\ \|v\|=1}} v^\top XX^\top v \geq \min_{\substack{v\in\operatorname{span}\{e_{i_1},e_{i_2}\}\\ \|v\|=1}} v^\top XX^\top v$$
$$= \min_{\mu_1,\mu_2\in\mathbb{R}} (\mu_1^2+\mu_2^2)^{-1}\left(\mu_1^2 S_{(1)} + \mu_2^2 S_{(2)} + 2\mu_1\mu_2 (XX^\top)_{i_1 i_2}\right).$$

Since, by Proposition 3.2 and equation (20),

$$\frac{\left|\frac{2\mu_1\mu_2}{\mu_1^2+\mu_2^2}(XX^\top)_{i_1 i_2}\right|}{S_{(2)}} \leq \frac{a_{np}^{-2}\max_{1\leq i<j\leq p}\sum_{t=1}^n |Z_{it}Z_{jt}|}{a_{np}^{-2} S_{(2)}} \xrightarrow[n\to\infty]{P} 0.$$

uniformly in $\mu_1,\mu_2\in\mathbb{R}$, an application of the the continuous mapping theorem finally yields that $\lambda_{(2)}/S_{(2)} \geq 1 + o_P(1)$, where $o_P(1) \to 0$ in probability as $n\to\infty$. Thus the proof for $\kappa \geq 1$ is complete. Now let $\kappa\in(0,1)$. Since $X^\top X$ and $XX^\top$ have the same non-trivial eigenvalues, we consider the transpose $X^\top$ of $X$. This inverts the roles of $p$ and $n$. Therefore, using Potter's bounds and $1/\kappa > 1$, the result follows from the same arguments as before. Note that we are in a special case of Theorem 1 if $\kappa = 0$.

In case (ii) we have that $n \sim (1/c\log(p/C))^{1/\kappa}$ is a slowly varying function in $p$, thus an application of Theorem 2 (i) to $X^\top$ gives the result. □

### 3.5. *Proof of Theorem 3*

As we shall see, the proof of Theorem 3 will more or less follow the same lines of argument as given for Theorem 1. We focus on the setting of Theorem 3 (i) here and mention (ii) and (iii) later. The next result is a generalization of Proposition 3.5 allowing for random coefficients.

**Proposition 3.6.** *Define $X = (X_{it})$ with $X_{it}$ satisfying (13) and (14). Suppose (10) holds for some $\beta > 0$. If $(\theta_i)$ is a stationary ergodic sequence, then, conditionally on $(\theta_i)$ as well as unconditionally, we have*

$$\sum_{i=1}^p \epsilon_{a_{np}^{-2}(\sum_{t=1}^n X_{it}^2 - n\mu_{X,\alpha})} \xrightarrow[n\to\infty]{D} \sum_{i=1}^\infty \epsilon_{\Gamma_i^{-2/\alpha}\left(E\left|\sum_{j=-\infty}^{\infty} c_j^2(\theta_1)\right|^{\alpha/2}\right)^{2/\alpha}} \tag{50}$$

*with $\mu_{X,\alpha}$ and $(\Gamma_i)$ as given in (7) and (11).*

*Proof.* We prove the cases $0 < \alpha < 2$ and $2 \leq \alpha < 4$ separately.

Let $0 < \alpha < 2$. We first prove that, conditionally on $(\theta_i)$,

$$\sum_{i=1}^p \epsilon_{a_{np}^{-2}\sum_{t=1}^n \sum_j c_j^2(\theta_i) Z_{i,t-j}^2} \xrightarrow[n\to\infty]{D} \sum_{i=1}^\infty \epsilon_{\Gamma_i^{-2/\alpha}\left(E\left|\sum_j c_j^2(\theta_1)\right|^{\alpha/2}\right)^{2/\alpha}} \tag{51}$$

by showing a.s. convergence of the Laplace functionals. By arguments from the proof of [36, Proposition 3.17] it suffices to show (15) only for a countable subset of the space of all nonnegative continuous functions with compact support. Thus we fix one nonnegative continuous function $f$ with compact support $\operatorname{supp}(f) \subset [c, \infty]$, $c > 0$. Conditionally on the process $(\theta_m)$, the points of the point process are independent, and thus

$$
\begin{aligned}
E\left(e^{-\sum_{i=1}^p f(a_{np}^{-2}\sum_{t=1}^n \sum_j c_j^2(\theta_i) Z_{i,t-j}^2)} \Big| (\theta_m)\right) &= \prod_{i=1}^p \left(1 - \frac{1}{p}\int (1-e^{-f(x)}) pP\left(a_{np}^{-2}\sum_{t=1}^n \sum_j c_j^2(\theta_i) Z_{1,t-j}^2 \in dx \Big| \theta_i\right)\right) \\
&= \prod_{i=1}^p \left(1 - \frac{1}{p}B_{i,p}\right),
\end{aligned} \tag{52}
$$

where $B_{i,p} = \int (1-e^{-f(x)}) pP(a_{np}^{-2}\sum_{t=1}^n \sum_j c_j^2(\theta_i) Z_{1,t-j}^2 \in dx | \theta_i)$. First assume

$$
\frac{1}{p}\sum_{i=1}^p B_{i,p} \xrightarrow[n\to\infty]{a.s.} B := \int (1-e^{-f(x)})\nu(dx) \tag{53}
$$

with $\nu$ given by $\nu((x,\infty]) := x^{-\alpha/2} E\left|\sum_j c_j^2(\theta_1)\right|^{\alpha/2}$, and

$$
\frac{1}{p^2}\sum_{i=1}^p B_{i,p}^2 \xrightarrow[n\to\infty]{a.s.} 0. \tag{54}
$$

Both claims will be justified later. By assumption (13), we have, using Lemma 3.1, almost surely

$$
B_{i,p} \le pP\left(a_{np}^{-2}\sum_{t=1}^n \sum_j \tilde{c}_j^2 Z_{1,t-j}^2 > c\right) \xrightarrow[n\to\infty]{} c^{-\alpha/2}\left|\sum_j \tilde{c}_j^2\right|^{\alpha/2},
$$

and hence there exists a $C > 0$ such that $B_{i,p} \le C$ for all $i, p \in \mathbb{N}$ a.s. The elementary inequality $e^{\frac{-x}{1-x}} \le 1 - x \le e^{-x}$ $\forall x \in [0,1]$, equivalently $e^{\frac{-x^2}{1-x}} \le (1-x)e^x \le 1$ $\forall x \in [0,1]$, implies together with (54), for some $c_1 > 0$, that

$$
1 \ge \prod_{i=1}^p \left(1 - \frac{B_{i,p}}{p}\right) e^{\frac{B_{i,p}}{p}} \ge \prod_{i=1}^p e^{-\frac{B_{i,p}^2}{p^2 - pB_{i,p}}} \ge \prod_{i=1}^p e^{-\frac{B_{i,p}^2}{p^2 - pC}} \ge e^{\frac{-c_1}{p^2}\sum_{i=1}^p B_{i,p}^2} \xrightarrow[n\to\infty]{a.s.} 1.
$$

As a consequence we have that the product in (52) is asymptotically equivalent to

$$
\prod_{i=1}^p e^{-\frac{1}{p}B_{i,p}} = e^{-\frac{1}{p}\sum_{i=1}^p B_{i,p}} \xrightarrow[n\to\infty]{a.s.} e^{-B} = e^{-\int (1-e^{-f(x)})\nu(dx)},
$$

where the convergence follows from (53). This implies the almost sure convergence of the conditional Laplace functionals, therefore (51) holds conditionally on $(\theta_i)$. Using (13) one shows similarly as in the proof of Proposition 3.5, conditionally on $(\theta_i)$, that (51) implies (50). Taking the expectation yields that (50) also holds unconditionally.

*Proof of* (53) *and* (54). As a function in $x$, $pP(\sum_{t=1}^n Z_{1t}^2 > a_{np}^2 x)$ is decreasing and converges pointwise to the continuous function $x^{-\alpha/2}$ as $n \to \infty$. Therefore this convergence is uniform on compact intervals of the form $[x_0, \infty]$ with $x_0 > 0$. Now fix $x > 0$ and let $d_i = \sum_j c_j^2(\theta_i)$. Since $d_i \le d = \sum_j \tilde{c}_j^2 < \infty$ for all $i \in \mathbb{N}$, $\frac{x}{d_i} \ge \frac{x}{d} > 0$ is bounded from below, and thus

$$
\sup_{i\in\mathbb{N}} \left| pP\left(\sum_{t=1}^n Z_{1t}^2 > a_{np}^2 \frac{x}{d_i} \Big| d_i\right) - x^{-\alpha/2} d_i^{\alpha/2} \right| \xrightarrow[n\to\infty]{a.s.} 0. \tag{55}
$$

Since $(d_i)$ is an instantaneous function of the ergodic sequence $(\theta_i)$, it is also ergodic and thus

$$\frac{1}{p}\sum_{i=1}^{p} d_i^{\alpha/2} \xrightarrow[n\to\infty]{a.s.} E|d_1|^{\alpha/2}. \tag{56}$$

As a consequence of (55) and (56) we obtain

$$\left|\frac{1}{p}\sum_{i=1}^{p} pP\left(\sum_{t=1}^{n} Z_{1t}^2 > a_{np}^2 \frac{x}{d_i}\Big| d_i\right) - x^{-\alpha/2}E|d_1|^{\alpha/2}\right| \xrightarrow[n\to\infty]{a.s.} 0.$$

Then it is straightforward to show, as in the proof of Lemma 3.1, using (13), that

$$\frac{1}{p}\sum_{i=1}^{p} pP\left(\sum_{t=1}^{n}\sum_{j} c_j(\theta_i)Z_{1,t-j}^2 > a_{np}^2 x\Big|\theta_i\right) \xrightarrow[n\to\infty]{a.s.} x^{-\alpha/2}E|d_1|^{\alpha/2}.$$

The vague convergence of above sequence of measures implies $p^{-1}\sum_{i=1}^{p} B_{i,p} \to B$ almost surely. In exactly the same way one can show that $p^{-1}\sum_{i=1}^{p} B_{i,p}^2$ converges, thus $p^{-2}\sum_{i=1}^{p} B_{i,p}^2 \to 0$ a.s., which establishes (53) and (54) as claimed.

Let $2 \le \alpha < 4$. As before one can show, for any $m < \infty$, that

$$\frac{1}{p}\sum_{i=1}^{p} pP\left(\sum_{t=1}^{n}\sum_{j=-m}^{m} c_j(\theta_i)(Z_{i,t-j}^2 - \mu_Z) > a_{np}^2 x\Big|\theta_i\right) \xrightarrow[n\to\infty]{a.s.} x^{-\alpha/2}E|d_1^m|^{\alpha/2},$$

where $d_1^m = \sum_{j=-m}^{m} c_j^2(\theta_1)$. Hence, an adaptation of the proof of Lemma 3.2 yields, for the truncated process

$$X_{it,m} = \sum_{k=-m}^{m} c_k Z_{i,t-k}, \quad \mu_{X,m} = EX_{11,m}^2 = \sum_{j=-m}^{m} c_j^2 \mu_Z,$$

that, conditionally on the sequence $(\theta_i)$,

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\left(\sum_{t=1}^{n} X_{it,m}^2 - n\mu_{X,m}\right)} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha}\left(E\left|\sum_{j=-m}^{m} c_j^2(\theta_1)\right|^{\alpha/2}\right)^{2/\alpha}}. \tag{57}$$

It is only left to show that this result extends to the more general setting where $m = \infty$. By Proposition 3.5 it suffices to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\sum_{i=1}^{p} P\left(\left|\sum_{t=1}^{n}(X_{it}^2 - X_{it,m}^2 - (\mu_X - \mu_{X,m}))\right| > a_{np}^2\gamma\Big|(\theta_r)\right) = 0.$$

To proof this claim, follow the string of arguments of Proposition 3.5 and make use of the fact that

$$\left|\sum_{i=1}^{p} c_j(\theta_i)\right| \le p\tilde{c}_j.$$

□

*Proof of Theorem 3.* *Proof of (i).* If we condition on $(\theta_i)$, the proofs of Propositions 3.3 and 3.4 easily carry over to this more general setting when we make use of assumption (13). Taking the expectation then yields convergence in operator norm unconditionally. A combination of this together with Proposition 3.6 completes the proof.

*Proof of (ii).* Note that (56) is the only step in the proof of Proposition 3.6 where we use the ergodicity of the sequence $(\theta_i)$. But also if $(\theta_i)$ is just stationary, the ergodic theorem implies that the average in (56) converges to the random variable $Y = E\left(|d_1|^{\alpha/2}|\mathcal{G}\right)$, where $\mathcal{G}$ is the invariant $\sigma$-field generated by $(\theta_i)$. By construction, $Y$ depends on $\alpha$ and $c_j(\cdot)$, but it is independent of $(\Gamma_i)$, since $(\theta_i)$ is independent of $(Z_{it})$.

*Proof of (iii).* In this setting $(\theta_i)$ is a Markov chain which may not be stationary. But since we derive all results in the proof of Theorem 3 (i) conditionally on $(\theta_i)$ and then take the expectation, stationarity is in fact not needed. The theory on Markov chains, see [32], in particular their Theorem 17.1.7 for Markov chains on uncountable state spaces, shows that (56) holds if the expectation is taken with respect to the stationary distribution of the Markov chain. □

*Acknowledgements.* The authors are grateful to the editor and the anonymous referees whose comments improved the paper considerably.

R.D. was supported in part by NSF Grant DMS-1107031 and the Institute for Advanced Study at Technische Universität München. O.P thanks the Technische Universität München - Institute for Advanced Study, funded by the German Excellence Initiative, and the International Graduate School of Science and Engineering for financial support. R.S was partially supported by the Technische Universität München - Institute for Advanced Study, funded by the German Excellence Initiative

## References

[1] Anderson, G.W., Guionnet, A., Zeitouni, O.: An introduction to random matrices, *Cambridge Studies in Advanced Mathematics*, vol. 118. Cambridge University Press, Cambridge (2010)

[2] Anderson, T.W.: Asymptotic theory for principal component analysis. Ann. Math. Statist. **34**, 122–148 (1963)

[3] Auffinger, A., Ben Arous, G., Péché, S.: Poisson convergence for the largest eigenvalues of heavy tailed random matrices. Ann. Inst. Henri Poincaré Probab. Stat. **45**(3), 589–610 (2009). DOI 10.1214/08-AIHP188. URL http://dx.doi.org/10.1214/08-AIHP188

[4] Bai, Z., Jiang, D., Yao, J.F., Zheng, S.: Corrections to LRT on large-dimensional covariance matrix by RMT. Ann. Statist. **37**(6B), 3822–3840 (2009). DOI 10.1214/09-AOS694. URL http://dx.doi.org/10.1214/09-AOS694

[5] Bai, Z., Silverstein, J.W.: Spectral analysis of large dimensional random matrices, second edn. Springer Series in Statistics. Springer, New York (2010). DOI 10.1007/978-1-4419-0661-8. URL http://dx.doi.org/10.1007/978-1-4419-0661-8

[6] Bai, Z., Zhou, W.: Large sample covariance matrices without independence structures in columns. Statist. Sinica **18**(2), 425–442 (2008)

[7] Banna, M., Merlevède, F.: Limiting spectral distribution of large sample covariance matrices associated to a class of stationary processes (2013). URL http://hal.archives-ouvertes.fr/hal-00773140

[8] Belinschi, S., Dembo, A., Guionnet, A.: Spectral measure of heavy tailed band and covariance random matrices. Comm. Math. Phys. **289**(3), 1023–1055 (2009). DOI 10.1007/s00220-009-0822-4. URL http://dx.doi.org/10.1007/s00220-009-0822-4

[9] Ben Arous, G., Guionnet, A.: The spectrum of heavy tailed random matrices. Comm. Math. Phys. **278**(3), 715–751 (2008). DOI 10.1007/s00220-007-0389-x. URL http://dx.doi.org/10.1007/s00220-007-0389-x

[10] Bhatia, R.: Matrix analysis, *Graduate Texts in Mathematics*, vol. 169. Springer-Verlag, New York (1997)

[11] Billingsley, P.: Convergence of probability measures, second edn. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York (1999). DOI 10.1002/9780470316962. URL http://dx.doi.org/10.1002/9780470316962

[12] Biroli, G., Bouchaud, J.P., Potters, M.: On the top eigenvalue of heavy-tailed random matrices. Europhys. Lett. EPL **78**(1), Art. 10,001, 5 (2007). DOI 10.1209/0295-5075/78/10001. URL http://dx.doi.org/10.1209/0295-5075/78/10001

[13] Bloemendal, A., Virág, B.: Limits of spiked random matrices I. Probab. Theory Related Fields (2012). DOI 10.1007/s00440-012-0443-2. URL http://dx.doi.org/10.1007/s00440-012-0443-2. Online First

[14] Bordenave, C., Caputo, P., Chafaï, D.: Spectrum of large random reversible Markov chains: heavy-tailed weights on the complete graph. Ann. Probab. **39**(4), 1544–1590 (2011). DOI 10.1214/10-AOP587. URL http://dx.doi.org/10.1214/10-AOP587

[15] Bordenave, C., Caputo, P., Chafaï, D.: Spectrum of non-Hermitian heavy tailed random matrices. Comm. Math. Phys. **307**(2), 513–560 (2011). DOI 10.1007/s00220-011-1331-9. URL http://dx.doi.org/10.1007/s00220-011-1331-9

[16] Brockwell, P.J., Davis, R.A.: Time series: theory and methods, second edn. Springer Series in Statistics. Springer-Verlag, New York (1991). DOI 10.1007/978-1-4419-0320-4. URL http://dx.doi.org/10.1007/978-1-4419-0320-4

[17] Cizeau, P., Bouchaud, J.P.: Theory of Lévy matrices. Phys. Rev. E **50**, 1810–1822 (1994). DOI 10.1103/PhysRevE.50.1810. URL http://link.aps.org/doi/10.1103/PhysRevE.50.1810

[18] Davis, R., Hsing, T.: Point process and partial sum convergence for weakly dependent random variables with infinite variance. Ann. Probab. **23**(2), 879–917 (1995)

[19] Davis, R.A., Resnick, S.I.: Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. **13**(1), 179–195 (1985). URL http://links.jstor.org/sici?sici=0091-1798(198502)13:1<179:LTFMAO>2.0.CO;2-D&origin=MSN

[20] Dyson, F.J.: Statistical theory of the energy levels of complex systems. I. J. Mathematical Phys. **3**, 140–156 (1962)

[21] El Karoui, N.: Spectrum estimation for large dimensional covariance matrices using random matrix theory. Ann. Statist. **36**(6), 2757–2790 (2008). DOI 10.1214/07-AOS581. URL http://dx.doi.org/10.1214/07-AOS581

[22] Embrechts, P., Goldie, C.M.: On closure and factorization properties of subexponential and related distributions. J. Austral. Math. Soc. Ser. A **29**(2), 243–256 (1980)

[23] Geman, S.: A limit theorem for the norm of random matrices. Ann. Probab. **8**(2), 252–261 (1980). URL http://links.jstor.org/sici?sici=0091-1798(198004)8:2<252:ALTFTN>2.0.CO;2-4&origin=MSN

[24] Hachem, W., Loubaton, P., Najim, J.: The empirical eigenvalue distribution of a Gram matrix: from independence to stationarity. Markov Process. Related Fields **11**(4), 629–648 (2005)

[25] Heyde, C.C.: On large deviation probabilities in the case of attraction to a non-normal stable law. Sankhyā Ser. A **30**, 253–258 (1968)

[26] James, A.T.: Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist. **35**, 475–501 (1964)

[27] Johnstone, I.M.: On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist. **29**(2), 295–327 (2001). DOI 10.1214/aos/1009210544. URL http://dx.doi.org/10.1214/aos/1009210544

[28] Kallenberg, O.: Random measures, third edn. Akademie-Verlag, Berlin (1983)

[29] Longford, N.T.: Random coefficient models, *Oxford Statistical Science Series*, vol. 11. Oxford University Press, New York (1993)

[30] Marčenko, V.A., Pastur, L.A.: Distribution of eigenvalues in certain sets of random matrices. Mat. Sb. (N.S.) **72 (114)**, 507–536 (1967)

[31] Mehta, M.L.: Random matrices, *Pure and Applied Mathematics (Amsterdam)*, vol. 142, third edn. Elsevier/Academic Press, Amsterdam (2004)

[32] Meyn, S., Tweedie, R.L.: Markov chains and stochastic stability, second edn. Cambridge University Press, Cambridge (2009)

[33] Pan, G.: Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix. J. Multivariate Anal. **101**(6), 1330–1338 (2010). DOI 10.1016/j.jmva.2010.02.001. URL http://dx.doi.org/10.1016/j.jmva.2010.02.001

[34] Péché, S.: Universality results for the largest eigenvalues of some sample covariance matrix ensembles. Probab. Theory Related Fields **143**(3-4), 481–516 (2009). DOI 10.1007/s00440-007-0133-7. URL http://dx.doi.org/10.1007/s00440-007-0133-7

[35] Pfaffel, O., Schlemm, E.: Eigenvalue distribution of large sample covariance matrices of linear processes. Probab. Math. Statist. **31 (2)**, 313–329 (2011)

[36] Resnick, S.I.: Extreme values, regular variation and point processes. Springer Series in Operations Research and Financial Engineering. Springer, New York (2008). Reprint of the 1987 original

[37] Soshnikov, A.: A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. J. Statist. Phys. **108**(5-6), 1033–1056 (2002). DOI 10.1023/A:1019739414239. URL http://dx.doi.org/10.1023/A:1019739414239

[38] Soshnikov, A.: Poisson statistics for the largest eigenvalues in random matrix ensembles. In: Mathematical physics of quantum mechanics, *Lecture Notes in Phys.*, vol. 690, pp. 351–364. Springer, Berlin (2006). DOI 10.1007/3-540-34273-7_26. URL http://dx.doi.org/10.1007/3-540-34273-7_26

[39] Tao, T., Vu, V.: Random covariance matrices: Universality of local statistics of eigenvalues. Ann. Probab. **40**(3), 1285–1315 (2012)

[40] Tracy, C.A., Widom, H.: Level-spacing distributions and the Airy kernel. Comm. Math. Phys. **159**(1), 151–174 (1994). URL http://projecteuclid.org/getRecord?id=euclid.cmp/1104254495

[41] Wachter, K.W.: The strong limits of random matrix spectra for sample matrices of independent elements. Ann. Probability **6**(1), 1–18 (1978)

[42] Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. Ann. of Math. (2) **67**, 325–327 (1958)

[43] Yao, J.: A note on a Marčenko-Pastur type theorem for time series. Statist. Probab. Lett. **82**(1), 22–28 (2012). DOI 10.1016/j.spl.2011.08.011. URL http://dx.doi.org/10.1016/j.spl.2011.08.011