Quantitative Finance, 2020
Vol. 20, No. 5, 769–781, https://doi.org/10.1080/14697688.2020.1711962

Random matrix models for datasets with fixed time horizons

G. L. ZITELLI *

Department of Mathematics, University of California, Irvine, CA 92617, USA

(*Received 15 January 2019; accepted 27 December 2019; published online 22 January 2020*)

This paper examines the use of random matrix theory as it has been applied to model large financial datasets, especially for the purpose of estimating the bias inherent in Mean-Variance portfolio allocation when a sample covariance matrix is substituted for the true underlying covariance. Such problems were observed and modeled in the seminal work of Laloux *et al.* [Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 1999, **83**, 1467] and rigorously proved by Bai *et al.* [Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Math. Finance*, 2009, **19**, 639–667] under minimal assumptions. If the returns on assets to be held in the portfolio are assumed independent and stationary, then these results are universal in that they do not depend on the precise distribution of returns. This universality has been somewhat misrepresented in the literature, however, as asymptotic results require that an arbitrarily long time horizon be available before such predictions necessarily become accurate. In order to reconcile these models with the highly non-Gaussian returns observed in real financial data, a new ensemble of random rectangular matrices is introduced, modeled on the observations of independent Lévy processes over a fixed time horizon.

Keywords: Random matrix theory; Sample covariance matrix; Lévy process; Marčenko–Pastur distribution; Empirical spectral distribution; Generalized gamma convolution

JEL Classification: C38, C60

1. Introduction

1.1. Motivation

The Marčenko–Pastur (M–P) law has been a key tool in the application of random matrix theory to wireless communications, finance, and statistics, appearing frequently as a first approximation when attempting to model the biases of statistics derived from large-dimensional datasets (see, among many others, Baik and Silverstein 2006, Bickel and Levina 2008, Bouchaud and Potters 2011, Couillet and Debbah 2011, Paul and Aue 2014). If N observations are employed to infer complex statistics about p features, the law succinctly addresses problems in the estimation of the underlying linear spectral statistics when p and N are comparably large such that $p/N = \lambda \sim O(1)$. Under appropriate normalization, the law states that the sample covariance matrix using N samples of a p-length isotropic random vector will exhibit a wide bulk of nonzero eigenvalues whose histogram asymptotically resembles the Marčenko–Pastur distribution m_λ with parameter $\lambda \in (0,\infty)$, whose continuous part is given by the density:

$$
\begin{aligned}
dm_\lambda(x) &= \frac{\sqrt{(\sigma_+ - x)(x-\sigma_-)}}{2\pi\lambda x}\,\mathrm{d}x, \\
x &\in [\sigma_-,\sigma_+], \quad \sigma_\pm = (1\pm\sqrt{\lambda})^2
\end{aligned}
\tag{1}
$$

Since the components of such a random vector are independent, the distribution m_λ can be thought of as the description of the bulk spectrum produced by purely noisy measurements. The subsequent theories of additive and multiplicative noise in the sample covariance of a system with a large and comparable number of observations and features has been quite successful, relying heavily on the shape of m_λ (see the survey of Capitaine and Donati-Martin 2016).

Applications to financial data have been investigated as early as Laloux *et al.* (1999), which directly compared the spectrum of a sample covariance matrix derived from the returns on S&P 500 assets over a 5-year period to the distribution m_λ for the appropriate shape $\lambda = p/N$. More recently, the multiplicative theory has been applied to the cleaning of sample covariance matrices for the purpose of making Mean-Variance portfolio theory more reliable (see Bai *et al.* 2009 and the recent survey of Bun *et al.* 2017). To provide a brief

*Email: gzitelli@uci.edu

account of the theory, suppose an individual wishes to invest in p risky assets whose returns over some fixed time increment Δt are well described by independent samples of a random $1 \times p$ vector $\mathbf{r}$ whose components have finite variance. The N observations of the returns can be used to construct an $N \times p$ (random) matrix of data $\mathbf{X}$ whose rows are independent and whose columns follow a fixed distribution. If $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix (unknown to the investor) associated to the random vector $\mathbf{r}$, then the change of basis $\mathbf{Y} = \mathbf{X}\boldsymbol{\Sigma}^{-1/2}$ creates a random matrix whose rows have been sampled from an isotropic random vector $\widetilde{\mathbf{r}} = \mathbf{r}\boldsymbol{\Sigma}^{-1/2}$. The M–P law (Marčenko and Pastur 1967) and subsequent multiplicative theory (Silverstein 1995, Bai and Zhou 2008) describe the asymptotic distribution of the eigenvalues of the sample covariance matrices derived from $\mathbf{Y}$ and $\mathbf{X}$, respectively. In particular, one might ask what can be said about the empirical spectral distribution (ESD, sometimes called the histogram measure) $\mu_{\mathbf{S}}$ of the sample covariance matrix $\mathbf{S}$, defined as

$$\mu_{\mathbf{S}} = \frac{1}{p} \sum_{\sigma \in \mathrm{Eig}[\mathbf{S}]} \delta_\sigma, \quad \mathbf{S} = \frac{1}{N}\mathbf{Y}^\dagger\mathbf{Y}$$

When the ratio $p/N \to \lambda \in (0, \infty)$ stabilizes and the isotropic random vectors $\widetilde{\mathbf{r}} = \widetilde{\mathbf{r}}_N$ have either fixed (for all N) i.i.d. components or have uniformly bounded fourth moments, the M–P law provides the following resolution to the problem: as $N \to \infty$, the histogram $\mu_{\mathbf{Y}}$ converges weakly to the Marčenko–Pastur distribution m_λ, which is the sum of a point mass $(1 - 1/\lambda)\delta_0$ (when $\lambda > 1$) and the continuous part given by (1). Properties of the corresponding ESD $\mu_{\mathbf{X}}$ can be deduced by observing that

$$\frac{1}{N}\mathbf{X}^\dagger\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{S}\boldsymbol{\Sigma}^{1/2} \sim \mathbf{S}\boldsymbol{\Sigma}$$

As a result, it is enough to study the eigenvalues of products of the form $\mathbf{S}\,\boldsymbol{\Sigma}$, where $\mathbf{S}$ is an M–P type random matrix.

The M–P law is attractive due to its universality and the speed at which the convergence $\mu_{\mathbf{S}} \to m_\lambda$ occurs. Universality is typically expressed in the standard presentation of the law, where the isotropic random vector $\widetilde{\mathbf{r}}$ has i.i.d. components (as is the case when $\mathbf{r}$ is multivariate Gaussian). Under these circumstances, the only assumption necessary for the law to hold is that the variance of these components be finite. At the same time, modestly sized matrices with N and p on the order of only 10^2 have been used as empirical evidence that $\mu_{\mathbf{S}} \to m_\lambda$ quite quickly. What is fascinating is that these two properties are not shown together, and ESD's constructed to demonstrate the effects of the law often use matrices with pseudo-random normal or Rademacher entries (see figures in Baik and Silverstein 2006, Burda *et al.* 2011, El Karoui 2013, Yao *et al.* 2015). Such distributions exhibit small kurtosis, in stark contrast to the highly leptokurtic distributions which typically appear when modeling asset returns.

Consider the application of M–P-inspired techniques in modeling a simple financial scenario. We begin with a large matrix $\mathbf{Y}$ of size 2000×500, representing the daily logarithmic returns on a collection of assets such as the S&P 500 over an 8 year window. Eigenvalue cleaning recipes (see the recent survey of Bun *et al.* 2017) and Mean-Variance portfolio bias estimation techniques such as those suggested by Bai *et al.* (2009) will then rely on the limiting measure for $\lambda = 1/4$. Indeed, the histogram of a 2000×500 matrix with i.i.d. standard normal entries will be strikingly similar to the limiting M–P distribution with $\lambda = 1/4$. However, universality of the law is not a statement about the fixed value $N = 2000$, and we can see in Figure 1 that a matrix composed of i.i.d. log-normal random variables, a distribution with finite moments of arbitrary order, exhibits eigenvalues outside of the predicted bulk. Such a matrix appears at some point in both models, where it is assumed that N is large enough so that the distribution of the entries is irrelevant.

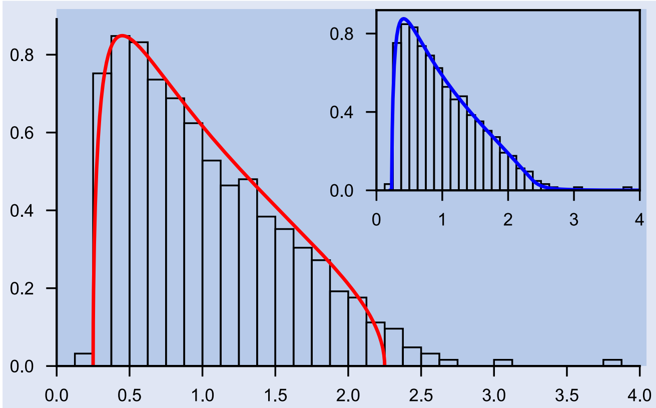

Figure 1. Histogram: 500 eigenvalues of the sample covariance matrix $\mathbf{S} = N^{-1}\mathbf{Y}^\dagger\mathbf{Y}$ of a random $N \times p = 2000 \times 500$ matrix $\mathbf{Y}$ with i.i.d. entries following the distribution of $Y = (Y' - \sqrt{e})/\sqrt{e^2 - e}$ where $Y' \sim \mathrm{LogNormal}(0, 1)$, having been normalized to have mean zero and unit variance. In red: The density of the M–P distribution m_λ for $\lambda = p/N = 1/4$. Inset: The histogram once again, compared to the limiting spectral distribution (in blue) for the log-normal sample Lévy covariance ensemble, as estimated in Section 4.2.

Methodological problems might also be considered when these models are applied to financial time series. If each row of the matrix ensemble is taken to be derived from daily returns on a collection of assets, then N denotes the number of days for which data has been collected. The asymptotic $N \to \infty$ therefore implies convergence over an arbitrarily long time horizon. In essence, although the individual entries of the matrix (representing daily returns) may be non-Gaussian, the theorem relies on the fact that N can be taken large enough so that the columns of $(1/\sqrt{N})\mathbf{Y}$ closely resemble the fluctuations of standard Brownian motion on the interval $[0, 1]$. The restriction of financial applications to fixed time horizons would suggest that this is unrealistic, as portfolio construction may well be employed over periods of time where the return process is still highly non-Gaussian.

An alternative approach for the purpose of modeling financial data is to design a random matrix ensemble whose entries describe the fluctuations of a collection of p stochastic processes over a fixed horizon $[0, T]$. Rather than the asymptotic $N \to \infty$ signifying additional observations beyond the horizon, we suppose that our observations are occurring at finer and finer discretizations of the interval $0 = t_0 < t_1 < \cdots < t_N = T$. If the rows of the matrix are taken to be i.i.d. (for each fixed N), then this corresponds to an equally spaced net $t_j = jT/N$, and the observations can be thought of as the fluctuations of p independent trajectories of a Lévy processes X_t. This coincides with a choice to model the log-price process of asset returns on a Lévy process, an idea which has attracted some interest (see selected chapters and discussions in Voit 2005, Jondeau *et al.* 2007, Jeanblanc *et al.* 2009, Pascucci 2011, Fischer 2014, Maejima 2015). We propose this as an alternative type of matrix ensemble, whose entries are i.i.d. but vary as $N \to \infty$ such that the sum of the entries in each column matches a fixed, infinitely divisible distribution.

Our departure from previous work in Random Matrix Theory can be understood along a few lines. The classical scenario investigated by Marčenko and Pastur (1967) involves a large random matrix $\mathbf{Y}$ of size $N \times p$, where N and p are both large but comparable, and considers the deterministic limiting histogram of its singular values as $N, p \to \infty$ with $p/N \to \lambda \in (0, 1)$. Initially, the assumptions on the entries of $\mathbf{Y}$ is that they are i.i.d. and follow some fixed, finite variance distribution for all N. This condition can be relaxed somewhat; up to rescaling (and observing instead the eigenvalues of $\mathbf{Y}^\dagger\mathbf{Y}$), we can consider an ensemble of matrices $\mathbf{Y}$ whose entries are i.i.d. following a changing distribution Y_N whose variance is $O(N^{-1})$, with the one additional assumption that the law of the entries satisfies Bai and Silverstein (2010, Theorem 3.10)

$$N \cdot \mathbb{E}\Big[\big|Y_N\big|^2 \mathbb{1}_{|Y_N|\geq\eta}\Big] \xrightarrow{N\to\infty} 0$$

for any $\eta > 0$. Any such ensemble falls into the Marčenko–Pastur basin of attraction. Therefore, in order to escape the M–P universe, we are interested in matrices whose entries are i.i.d. random variables drawn from changing distributions which become less Gaussian as N grows. We are strongly motivated by the conclusions in Carr *et al.* (2002) that the diffusion components of financial data are likely diversifiable, suggesting that noise and small idiosyncratic factors in the market may be more appropriately modeled by a pure point process. The SLCE does precisely this, taking the i.i.d. entries of our matrix to follow the distributions of $X_{T/N}$, where $T > 0$ is a fixed horizon parameter and X_t is a Lévy process. If the right tail of X_t is subexponential, then $\mathbb{P}[X_{T/N} > \eta] \sim (T/N)\Pi((\eta, \infty))$ as $\eta \to \infty$ (Sato 2013, Remark 25.14) for the Lévy measure Π (see (2)), and we have that

$$N \cdot \mathbb{E}\Big[\big|X_{T/N}\big|^2 \mathbb{1}_{|X_{T/N}|\geq\eta}\Big] \sim N \cdot \frac{T}{N}\int_\eta^\infty x^2 \, \mathrm{d}\Pi(x) \sim O(1)$$

This is assuming the variance of X_t even exists. This is significantly different from the Gaussian case, where if $\mathcal{L}(X_{T/N}) = N(0, T/N)$ is normally distributed then its tails are asymptotically

$$O(\sqrt{N}\,\mathrm{e}^{-\eta^2 N/2T}), \quad \eta \to \infty$$

Consequently, it is reasonable to expect that for a non-Gaussian Lévy process, the SLCE has the capacity to behave quite differently from the M–P law.

Similarly, one might compare our results to the theory of heavy-tailed random matrices, which was rigorously founded on the work of Ben Arous and Guionnet (2008) and Belinschi *et al.* (2009). This began with the study of large square matrices whose entries are heavy-tailed, lying in the domain of attraction of α-stable distributions. Some such banded matrices mimic the M–P case for a particular shape parameter $\lambda \in (0, 1)$, leading to a notion of heavy-tailed M–P type matrices. Connections with Free Probability were later made in Politi *et al.* (2010), where they considered "Free Wishart" matrices of the form

$$\left(\frac{1}{M}\sum_{j=1}^{M}\mathbf{U}_j\mathbf{L}_j\mathbf{U}_j^\dagger\right)\mathbf{P}_\lambda$$

Here the $\mathbf{L}_j$ are $N \times N$ matrices with i.i.d. heavy-tailed entries, while the $\mathbf{U}_j$ are independent Haar distributed unitary matrices, and $\mathbf{P}_\lambda$ is a projection onto a subspace of dimension λN. In both cases, the limiting eigenvalue distributions have unbounded right tails which decay like $1/|x|^{\alpha+1}$ as $x \to \infty$. Such a theory is particularly devastating for a model of asset prices, as it implies that noisy eigenvalues of large size may occur with a heavy-tailed frequency. Neither matrix model looks similar to the empirical bulk eigenvalues; a very different situation then that implied by Laloux *et al.* (1999).

In contrast to such previous models, the SLCE is parametrized by a Lévy process X_t, which can have a variety of tail behaviors, notably any feasible power-law tail decay via α-stable laws and Student's-t distributions (Sato 2013). This follows a general perspective on asset modeling established in Mantegna and Stanley (1994, 1995) and explored in texts like Voit (2005) and Jondeau *et al.* (2007), where classes of Lévy processes are used in place of Brownian motion. In this way, we are able to bridge the gap between the overly conservative M–P setting, which occurs when X_t is taken to be Brownian motion, and the wild heavy-tailed setting, which can be captured by taking X_t as the standard one-dimensional Lévy flight process.

This draws some comparison with the theory of Wishart–Student ensembles developed in Biroli *et al.* (2007b) and discussed further in Bouchaud and Potters (2011). Such results encompass random matrices of the form $(1/N)\mathbf{YTY}^\dagger$, where the $\mathbf{Y}$ are random matrices of M–P type and $\mathbf{T}$ is an $N \times N$ random diagonal matrix whose entries follow a Student's-t distribution. Such an ensemble was developed concurently during the investigation of matrices with i.i.d. power-law entries in Biroli *et al.* (2007a), however the explicit form of the product lends itself to techniques from Free Probability which make computation of the limiting distribution feasible. In fact, we take a similar approach in Section 4.1 when approximating the SLCE, although we consider instead transposed products of the form $(1/N)\mathbf{T}^{1/2}\mathbf{Y}^\dagger\mathbf{Y}\mathbf{T}^{1/2}$. Our departure from both perspectives lies in maintaining the i.i.d. property of the ensemble while using a changing distribution for the entries in order to prove a stable limiting spectrum without the use of Free Probability. Most importantly, however, the SLCE appears to match the scaling behavior observed in the S&P 500 and Nikkei 225 universes when passing from daily to intraday datasets, as discussed in Section 5.2, which has not been possible using any of the aforementioned matrix models.

The contribution of this paper is threefold. First, we outline a simple collection of random matrix ensembles based on this idea of observing independent Lévy processes over a discretization of a fixed interval $[0, T]$. Next, we prove the existence of a limiting spectral distribution in the case where the driving process X_t is essentially bounded, and demonstrate through Monte Carlo simulations that a similar limiting distribution emerges for modestly sized values of N (as in the M–P case) for more general processes as well. For the particular subclass of Extended Generalized Gamma Convolutions (EGGCs), there is a connection with M–P type products of the form **ST**, where **T** is now taken to be a random diagonal matrix with entries related to the quadratic variation of the Lévy process in question. This enables a fast algorithm for the approximation of the density of the limiting spectral distribution in the EGGC case, which is applicable even for processes that do not have a convenient analytic description. Finally, we compare the spectra of sample covariance matrices derived from daily and intraday returns on assets lying in the S&P 500 and Nikkei 225 indices, which show deviations from the M–P law under the latter timescale. Without sophisticated parameter fitting, we demonstrate an ad hoc specification of our ensemble that produces a spectrum in-silico quite similar to the Nikkei 225 constituents.

1.2. Preliminary notation

In what follows, we use $\mathcal{L}(\cdot)$ to denote the law of a random variable or probability measure on $\mathbb{R}$. An infinitely divisible (ID) random variable X is one which can, for any $n \geq 1$, be expressed as

$$X = X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}$$

where the collections $\{X_j^{(n)}\}_{j=1}^n$ are i.i.d. Many common distributions are ID (Sato 2013). The theory of ID random variables coincides with that of Lévy processes, defined as a càdlàg process X_t with stationary and independent increments. In particular, a random variable is ID if and only if it can be realized as the fluctuation of a Lévy process over a unit time interval, which is to say that $\mathcal{L}(X) = \mathcal{L}(X_1)$.

We let X_t denote a fixed Lévy process throughout, and write X for the corresponding infinitely divisible random variable such that $\mathcal{L}(X) = \mathcal{L}(X_1)$. The symbol $[X]_t$ is used for the quadratic variation process corresponding to X_t, which can be defined as the limit in probability of the discretized quadratic variation

$$\sum_{j=1}^{M} (X_{jt/M} - X_{(j-1)t/M})^2 \to [X]_t$$

where convergence occurs in probability as $M \to \infty$. As X_t is a semi-martingale, $[X]_t$ is guaranteed to exist for all $t \in (0, \infty)$. Properties of $[X]_t$ are discussed in Sections 2.1 and 2.3. If $[X]_t = \sigma^2 t$ for some constant $\sigma^2 > 0$, then it follows that X_t must be a Brownian motion process with variance σ^2.

The Stieltjes transform of a probability measure μ is denoted by

$$S_\mu(z) = \int_{\mathbb{R}} \frac{1}{x - z} \, \mathrm{d}\mu(x)$$

and defines an analytic function from the upper half plane $\mathbb{C}^+$ to its closure. If $\{\mu_n\}_{n \in \mathbb{N}}$ is a sequence of probability measures then the pointwise convergence in $\mathbb{C}^+$ of a sequence of Stieltjes transform $S_{\mu_n}(z) \to S_\mu(z)$ implies the vague (and therefore weak) convergence of probability measures $\mu_n \to \mu$.

The Kolmogorov–Smirnov statistic is defined for two distributions μ and ν in terms of their cumulative distribution functions (CDFs) F_μ and F_ν as

$$d_{\mathrm{K-S}}(\mu, \nu) = \sup_{x \in \mathbb{R}} |F_\mu(x) - F_\nu(x)|$$

When applied to datasets, the value $d_{\mathrm{K-S}}$ is reported using the empirical CDFs of each dataset.

2. Sample Lévy covariance ensembles

2.1. Classification of Lévy processes

By the Lévy–Khintchine formula (see Sato 2013, for references throughout this section), the characteristic function $\phi_{X_t}(\vartheta)$ of a Lévy process X_t can be expressed for all $t > 0$ as

$$\frac{1}{t} \log \phi_{X_t}(\vartheta) = \mathrm{i}\mu\vartheta - \frac{1}{2}\sigma^2\vartheta^2 + \int_{\mathbb{R}} \left[\mathrm{e}^{\mathrm{i}\vartheta x} - 1 - \frac{\mathrm{i}\vartheta x}{1 + x^2} \right] \mathrm{d}\Pi(x) \tag{2}$$

where Π is a Borel measure on $\mathbb{R}$ such that $\Pi(\{0\}) = 0$ and

$$\int_{\mathbb{R}} \frac{x^2}{1 + x^2} \, \mathrm{d}\Pi(x) < \infty$$

We say that X_t has a Lévy triplet (μ, σ, Π), and the measure Π is called the Lévy measure of X. The three terms in the expression correspond to the decomposition

$$X_t = \mu t + \sigma B_t + X'_t$$

where B_t is standard Brownian motion and X'_t is another Lévy process independent from B_t. If $\int_{\mathbb{R}} (|x|^3/(1 + x^2)) \, \mathrm{d}\Pi(x) < \infty$ then the mean of X'_t is finite and equal to $\int_{\mathbb{R}} (x^3/(1 + x^2)) \, \mathrm{d}\Pi(x)$.

The tail behavior of Π corresponds to the tail behavior of X'_t, which in turn determines the existence of moments of X. The higher order cumulants $\kappa_n[X]$ of X, if they exist, are in correspondence with the moments of Π such that

$$\begin{aligned} \kappa_2[X] &= \mathrm{var}[X] = \sigma^2 + \int_{\mathbb{R}} x^2 \, \mathrm{d}\Pi(x) \\ \kappa_n[X] &= \int_{\mathbb{R}} x^n \, d\Pi(x), \quad n \geq 3 \end{aligned} \tag{3}$$

From (3) we can see that the normalized excess kurtosis $\kappa_4[X_t]/(\kappa_2[X_t])^2$ of X_t, if it exists, decays like $O(t^{-1})$ as $t \to \infty$, reflecting the fact that X_t converges in shape to a normal distribution.

The behavior of Π in a neighborhood of zero determines the qualitative path properties of X'_t. When $\Pi(\mathbb{R}) = r < \infty$, the process X_t' can be identified with a compensated compound Poisson process

$$X_t' \sim \sum_{j=1}^{N_{r \cdot t}} Z_j - \mu' t \tag{4}$$

where N_t is a standard Poisson jump process and the Z_j are i.i.d. random variables with distributions given by the probability measure $r^{-1}\Pi$, with $\mu' = \int_{\mathbb{R}} (x/(1+x^2))\, \mathrm{d}\Pi(x)$. In fact, any Lévy process can be expressed as the limit of a sequence of compensated compound Poisson processes. By taking a sequence of independent Poisson random variables N_m, each with rate m, and a collection $\mathcal{L}(Z_{j,m}) = \mathcal{L}(X_{t/m})$ for all $m, j \in \mathbb{N}$ of independent random variables, it can be shown that

$$\mathcal{L}\left(\sum_{j=1}^{N_m} Z_{j,m}\right) \to \mathcal{L}(X_t) \tag{5}$$

where convergence is in distribution as $m \to \infty$.

When $\int_{(-1,1)} |x|\, \mathrm{d}\Pi(x) < \infty$, the Lévy process X_t' is almost surely a process of bounded variation over any finite interval. When this condition is met, (2) can be rewritten as

$$\frac{1}{t} \log \phi_{X_t}(\vartheta) = \mathrm{i}\tilde{\mu}\vartheta - \frac{1}{2}\sigma^2\vartheta^2 + \int_{\mathbb{R}} [\mathrm{e}^{\mathrm{i}\vartheta x} - 1]\, \mathrm{d}\Pi(x) \tag{6}$$

corresponding to a decomposition $X_t = \tilde{\mu}t + \sigma B_t + \tilde{X}_t$. In this case, the mean of $\tilde{X}_t$ (if it exists) will be equal to $\int_{\mathbb{R}} x\, \mathrm{d}\Pi(x)$. This form of (2) is useful for describing Lévy processes X_t whose distributions are nonnegative, called subordinators. Necessary and sufficient conditions for a process X_t to be a subordinator is that the process has finite variation, lacks a Brownian motion component ($\sigma = 0$), the Lévy measure Π is one-sided ($\Pi((-\infty, 0]) = 0$), and finally the drift $\tilde{\mu}$ is nonnegative. By the form of (6), it is clear that being a subordinator is time-invariant, so that X_t has a distribution supported on $[0, \infty)$ for some $t > 0$ if and only if it is supported there for all such t.

A Lévy process is called *essentially bounded* if the sequence $\sqrt[n]{|\kappa_n[X]|}$ for $n \in \mathbb{N}$ is bounded, or equivalently that the support of the measure Π is contained in some finite interval $[-B, B]$ with $B > 0$. By the definition, the property of being essentially bounded is time-invariant, so that X is essentially bounded if and only if X_t is essentially bounded for all $t > 0$. Essentially bounded processes have exponential moments of all orders, such that

$$\mathbb{E}[\mathrm{e}^{m|X|}] < \infty, \quad m > 0$$

This implies that, although essentially bounded Lévy processes do not have bounded tails, their long-term tail behavior is quite dampened.

Essentially bounded processes provide a modeling approach, in the vein of Mantegna and Stanley, to the problem of 'ultraslow' convergence of i.i.d. sums of random variables in the central limit theorem. In Mantegna and Stanley (1995), it was famously observed that scaling in the Standard and Poor's 500 index failed to exhibit heavy-tailed behavior for extreme outliers. As a example of a distribution with such properties, they defined their truncated Lévy flight in Mantegna and Stanley (1994) in terms of the an α-stable Lévy distribution, whose density is restricted to a large bounded set, although such a distribution fails to represent a Lévy process. The phrase 'truncated Lévy flight' was later adopted by Koponen (1995), whose distributions were used as the basis for the CGMY model of Carr *et al.* (2002). The class of essentially bounded processes follows a similar modification of Lévy flight, as the condition that Π has bounded support still produces a Lévy process that can have arbitrarily large kurtosis, while being more convenient for the purpose of density estimation.

In 1977, Thorin introduced a class of Lévy processes for the purpose of proving the infinite divisibility of many distributions (1977a, 1977b). Members of this class can be described in terms of their Lévy measures Π as follows. We say that an infinitely differentiable function $g : (0, \infty) \to \mathbb{R}$ is completely monotone (CM) if $(-1)^n g^{(n)}(x) \geq 0$ for all $x > 0$ and $n \geq 0$. Now suppose a Lévy process X_t has Lévy measure Π with density $d\Pi(x)$, given by

$$\mathrm{d}\Pi(x) = \begin{cases} x^{-1} g_+(x)\, \mathrm{d}x, & x > 0 \\ |x|^{-1} g_-(-x)\, \mathrm{d}x, & x < 0 \end{cases}$$

where the functions $g_\pm$ are CM. Then X_t is said to be in the Thorin class $T(\mathbb{R}^+)$, or a Generalized Gamma Convolution (GGC) if it is a subordinator, or equivalently that $g_- \equiv 0, \sigma = 0$, and $\mu - \int_0^\infty g_+(x)\, \mathrm{d}x \geq 0$. Otherwise, we say that X_t is in the extended Thorin class $T(\mathbb{R})$, or an Extended Generalized Gamma Convolution (EGGC).

Gaussian, α-stable, log-normal, Student's-t, Pareto, gamma, χ^2, and generalized inverse Gaussian can all be shown to be GGC or EGGC, thus making them ID. If B_t is a Brownian motion process without drift and G_t is a GGC subordinator then B_{G_t} is a symmetric EGGC, and all such symmetric EGGCs arise this way (Bondesson 1992). Many popular distributions suggested for the modeling of asset returns are EGGC, including the variance-gamma (VG) model of Madan and Seneta (1990), the normal-inverse Gaussian (NIG) model of Barndorff-Nielsen (1997), and tempered stable distributions and the CGMY model of Carr *et al.* (2002).

2.2. *Definition of the sample Lévy covariance ensemble*

Let X_t be a fixed Lévy process, and fix some time horizon $T \in (0, \infty)$ and shape parameter $\lambda \in (0, \infty)$. For each $N \geq 1$, let $p = p(N)$ denote a function $p : \mathbb{N} \to \mathbb{N}$ such that $p/N \to \lambda$ as $N \to \infty$. Now consider an $N \times p$ matrix $\mathbf{Y}_N$ with i.i.d. entries with law $\mathcal{L}(X_{T/N})$:

$$\mathbf{Y}_N = \underbrace{\left.\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{Np} \end{bmatrix}\right\}}_{p} N, \quad \mathcal{L}(y_{jk}) = \mathcal{L}(X_{T/N})$$

We will generally suppress N in the notation whenever possible. We then consider the sample covariance matrix

$$\mathbf{S} = \mathbf{Y}^{\dagger}\mathbf{Y}$$

The columns of $\mathbf{Y}$ can be thought of as sampled fluctuations of p i.i.d. Lévy processes with distributions X_t over the interval $[0, T]$, with discretization at N equally spaced points. The ensemble reduces to the M–P case when X is normally distributed. When X follows a Lévy α-stable distribution with $\alpha \in (0, 2)$, the scenario is encompassed by the theory of heavy-tailed random covariance matrices (see Ben Arous and Guionnet 2008, Belinschi *et al.* 2009). We say that the $p \times p$ matrices $\mathbf{S}$ follow a Sample Lévy Covariance Ensemble (SLCE) with parameters (X_t, T, λ).

As in the introduction, we define the ESD of $\mathbf{S}$ as

$$\mu_{\mathbf{S}} = \frac{1}{p} \sum_{\sigma \in \mathrm{Eig}[\mathbf{S}]} \delta_{\sigma}$$

The existence of a limiting distribution for $\mathbf{S}$ in the case when X_t is essentially bounded is guaranteed by the theory of localized vectors developed by Benaych-Georges and Cabanal-Duvillard (2012), and is presented in the following theorem.

THEOREM 2.1 *Let* $\mathbf{S}_N$ *denote a sequence of* $p \times p$ *SLCE matrices with parameters* (X, T, λ). *Suppose further that* X *is essentially bounded. Then there exists a distribution* $\mu_{(X,T,\lambda)}$, *parametrized by the triplet* (X, T, λ), *such that the weak convergence*

$$\mu_{\mathbf{S}_N} \to \mu_{(X,T,\lambda)}$$

occurs almost surely as $N \to \infty$. *The distribution* $\mu_{(X,T,\lambda)}$ *is weakly continuous in terms of its parameters* (X, T, λ), *where continuity in the first argument is taken with the topology of weak convergence of probability distributions. Furthermore,* $\mu_{(X,T,\lambda)}$ *depends only on the even cumulants of* X, *such that if* $X^{(1)}$ *and* $X^{(2)}$ *are two essentially bounded random variables with* $\kappa_{2n}[X^{(1)}] = \kappa_{2n}[X^{(2)}]$ *for all* $n \in \mathbb{N}$, *then* $\mu_{(X^{(1)},T,\lambda)} = \mu_{(X^{(2)},T,\lambda)}$. *Finally, if* X *is normally distributed, then* $\mu_{(X,T,\lambda)}$ *is some scaling of the M–P law. Otherwise, if* X *is not normally distributed, then* $\mu_{(X,T,\lambda)}$ *has unbounded support but with finite exponential moments of all orders.*

2.3. Convergence of sample variance to the quadratic variation process

We are interested in a random matrix ensemble $\mathbf{Y}$, where the columns of $\mathbf{Y}$ correspond to p independent experiments where X_t is sampled N times over the interval $[0, T]$ to produce observations

$$0 = X_0, X_{T/N}, X_{2\cdot T/N}, \ldots, X_{N\cdot T/N} = X_T$$

The entries $[\mathbf{Y}]_{jk} = y_{jk}$ are modeled on the fluctuations over the jth time interval, which is to say that

$$\mathcal{L}(y_{jk}) = \mathcal{L}(X_{j\cdot T/N} - X_{(j-1)\cdot T/N}) = \mathcal{L}(X_{T/N})$$

Suppose that we fix p and let $N \to \infty$, so that $\lambda \to 0^+$. In the usual setting, the M–P law converges weakly to the point mass at 1, reflecting the fact that the sample covariance of any two columns converges to 0, while the sample variance of each column converges to 1. For a non-Gaussian Lévy process, however, there are additional complications.

Let Z_j be i.i.d. random variables with $\mathcal{L}(Z_j) = \mathcal{L}(X_{T/N})$. When the time horizon $T > 0$ is fixed, the biased sample variance estimator on $Z_1, \ldots, Z_N$ is defined as

$$\widehat{\sigma}^2 = \frac{1}{T} \sum_{j=1}^{N} \left(Z_j - \frac{1}{N} \sum_{k=1}^{N} Z_k \right)^2 \tag{7}$$

Suppose X has finite fourth moment and variance equal to σ^2, then the expected value of $\widehat{\sigma}^2$ is $((N-1)/N)\sigma^2$ and its variance can be computed explicitly as

$$\mathrm{var}[\widehat{\sigma}^2] = \left(\frac{N-1}{N}\right)^2 \left(\frac{1}{T}\kappa_4[X] + \frac{2}{N-1}\sigma^4\right) \tag{8}$$

where $\kappa_4[X]$ is the fourth cumulant of X. In particular, as $N \to \infty$ the first term in (7) dominates and $\widehat{\sigma}^2$ converges in probability to the normalized quadratic variation process $(1/T)[X]_T$, whose variance (when it exists) is equal to $T^{-1}\kappa_4[X]$.

Since X_t is a semi-martingale, the quadratic variation process $[X]_t$ exists even when X_t lacks a well defined second moment. By (3), the only Lévy process with zero fourth cumulant is Brownian motion. It follows that in all other cases, the sample variance of each column of $\mathbf{X}$ will not converge almost surely to a non-random number, but rather converges in distribution to the normalized quadratic variation process $(1/T)[X]_T$. The quadratic variation can be computed explicitly for compensated compound Poisson processes, as in (4), as

$$\mathcal{L}([X']_t) = \mathcal{L}\left(\sum_{j=1}^{N_{r\cdot}} Z_j^2\right) \tag{9}$$

The quadratic variation of an arbitrary Lévy process can be described as the limiting distribution of a similar modification of (5). From this it is easy to see that all quadratic variation processes $[X]_t$ are subordinators. Furthermore, if X_t has finite moments of all orders then the cumulants of $[X]_t$ satisfy the relation

$$\kappa_n[[X]_t] = \kappa_{2n}[X_t], \quad n \geq 1$$

Of particular note is the fact that if Π is absolutely continuous with respect to Lebesgue measure so that $\mathrm{d}\Pi(x) = \rho(x)\,\mathrm{d}x$, then the Lévy measure $\widetilde{\Pi}$ of $[X]_t$ is also absolutely continuous and satisfies

$$\mathrm{d}\widetilde{\Pi}(x) = \frac{\rho(\sqrt{x}) + \rho(-\sqrt{x})}{2\sqrt{x}}\,\mathrm{d}x, \quad x > 0$$

This relationship can be used to explicitly describe $[X]_t$ in some cases where X_t is known. For instance, if X_t is an α-stable process for some $0 < \alpha < 2$ (of any skewness), then $[X]_t$ is a nonnegative $\alpha/2$-stable process.

We conclude this section by noting that if a limiting distribution for the ESD of our sample Lévy covariance matrix $\mathbf{S}$

exists for a particular choice of X, we expect it to converge to $[X]_T$ as $\lambda \to 0^+$. Such convergence is guaranteed rigorously in the case where X is essentially bounded by Theorem 2.1.

3. Simulations of limiting distributions with EGGC entries

We consider two examples of non-stable Lévy processes encountered in the financial modeling of asset returns. The first is the variance-gamma (VG) process X_t^{VG}, which can be realized as Brownian motion subordinated to a gamma process G_t^{gam}. The Lévy measure of the gamma process with unit mean and variance at time $t = 1$ can be expressed as

$$\mathrm{d}\Pi_{\mathrm{gam}}(x) = \frac{\mathrm{e}^{-x}}{x}\,\mathrm{d}x, \quad x > 0$$

The gamma process G_t^{gam} is such that $\mathcal{L}(G_1^{\mathrm{gam}}) = \mathrm{Exp}(1)$. If B_t is an independent standard Brownian motion process, the Lévy measure of the VG process $B_{G_t^{\mathrm{gam}}}$ can be found explicitly as

$$\mathrm{d}\Pi_{\mathrm{VG}}(x) = \frac{1}{|x|}\,\mathrm{e}^{-\sqrt{2}|x|}\,\mathrm{d}x, \quad x \neq 0$$

and it is well known that $\mathcal{L}(B_{G_1^{\mathrm{gam}}}) = \mathcal{L}(Z\sqrt{E})$ follows a Laplace distribution, where Z and E are independent with $\mathcal{L}(Z) = \mathcal{N}(0,1)$ and $\mathcal{L}(E) = \mathrm{Exp}(1)$.

The second process we consider is the normal-inverse Gaussian (NIG) process X_t^{NIG}, which can be realized as Brownian motion subordinated to an inverse Gaussian process G_t^{IG}. The Lévy measure of an inverse Gaussian process with unit mean and variance is given by

$$\mathrm{d}\Pi_{\mathrm{IG}}(x) = \frac{\mathrm{e}^{-x/2}}{2\pi x^{3/2}}\,\mathrm{d}x, \quad x > 0$$

The Lévy measure for the NIG process can be expressed in terms of modified Bessel functions (Pascucci 2011), but for the purpose of simulating random variables it is enough to know that

$$\mathcal{L}(X_t^{\mathrm{NIG}}) = \mathcal{L}(Z\sqrt{G_t^{\mathrm{IG}}})$$

where $\mathcal{L}(Z) = \mathcal{N}(0,1)$ is independent of G_t^{IG}.

Both processes are symmetric EGGCs, and so the entries of our matrices can be generated as the Hadamard product of a matrix with i.i.d. standard normal entries and one whose entries are taken to be the square root of i.i.d. random variables drawn from a gamma distribution or an inverse-Gaussian distribution, respectively. Data on the densities of the ESDs is collected by sampling matrices of various size

Table 1. Monte—Carlo simulations of SLCE spectra driven by variance-gamma (VG) and normal-inverse Gaussian (NIG) processes for various values of N and T. Each experiment consists aggregate eigenvalues from $10^6/p$ pseudo-random matrices. Processes have been normalized to have identical variance and kurtosis. Deviations from M—P law are reported.

Entries	N	T	$\#\{\sigma : \sigma < \sigma_{\min} = 0.25\}$	$\#\{\sigma : \sigma > \sigma_{\max} = 2.25\}$	$d_{\mathrm{K-S}}(\mu_N, m_\lambda)$	$d_{\mathrm{K-S}}(\mu_N, \mu_{2000})$
VG	20	1	0.418500×10^6	0.121512×10^6	0.432708	0.003008
	200	1	0.419560×10^6	0.121549×10^6	0.432980	0.000919
	2000	1	0.419553×10^6	0.122387×10^6	0.433081	NA
	20	10	0.098902×10^6	0.074011×10^6	0.129765	0.013368
	200	10	0.088608×10^6	0.071761×10^6	0.121758	0.001301
	2000	10	0.088087×10^6	0.071791×10^6	0.121542	NA
NIG	20	1	0.322939×10^6	0.104397×10^6	0.362288	0.008482
	200	1	0.322989×10^6	0.104298×10^6	0.362920	0.000984
	2000	1	0.322906×10^6	0.104413×10^6	0.363221	NA
	20	10	0.081187×10^6	0.070340×10^6	0.117020	0.018738
	200	10	0.067578×10^6	0.067129×10^6	0.109260	0.002326
	2000	10	0.065810×10^6	0.066821×10^6	0.107971	NA

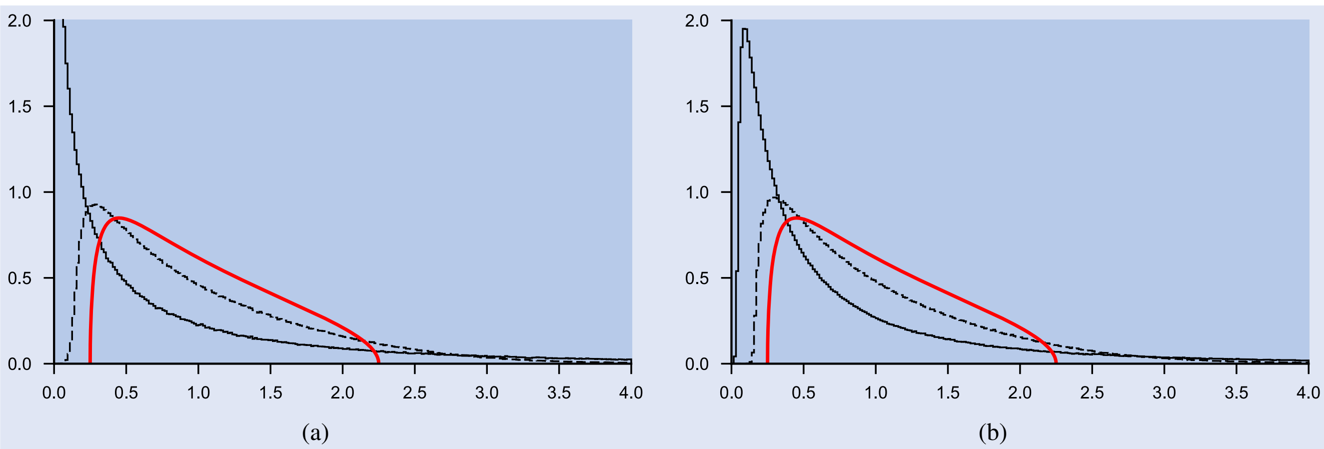

Figure 2. Histograms: One million eigenvalues aggregated from 2000 different 500×500 SLCE matrices $\mathbf{S} = \mathbf{Y}^\dagger \mathbf{Y}$, where $\mathbf{Y}$ is a pseudo-random $N \times p = 2000 \times 500$ matrix with (a) i.i.d. VG entries and (b) i.i.d. NIG entries, $T = 1$ (solid) and $T = 10$ (dashed). The excess kurtosis of individuals entries is $6 \times 10^3/T$. In red: The density of the M–P law for $\lambda = p/N = 1/4$.

(dependent on N), where $p = \lceil \lambda N \rceil$ for $\lambda = 1/4$. Experiments for different choices of λ show similar results; we display only $\lambda = 1/4$ for brevity and the convenience of the M–P distribution m_λ having support on the interval $[0.25, 2.25]$. For each choice of N, eigenvalues are aggregated from a total of $10^6/p$ Monte Carlo simulations in order to produce one million datapoints. Entries are normalized in order to be comparable to the M–P distribution m_λ.

Results are displayed in Table 1, which shows that the distribution of the eigenvalues deviates significantly from the M–P distribution m_λ. The VG process leads to sample covariance matrices which carry a huge portion of their spectrum to the left of the M–P bulk $[0.25, 2.25]$, as visualized in Figure 2(a). For $T = 1$, the density even appears to be unbounded. The normalization of both processes was chosen such that the excess kurtosis of individuals entries in the matrices can be computed as $3N/T$ for both ensembles, demonstrating that a limiting distribution is affected but not exclusively determined by the fourth cumulant of X_t. Shrinkage of the spectrum when comparing $T = 1$ and $T = 10$ is expected, as the entries necessarily become more Gaussian over longer horizons. For datasets of equal size 10^6, the threshold for rejecting the null hypothesis (that the two distributions are identical) with a confidence of 99% for the K–S test is that $d_{\text{K}-\text{S}}$ exceeds

$$\sqrt{\log\left(\frac{2}{1-0.99}\right)\frac{1}{10^6}} \approx 0.002302$$

Although this is under the assumption of independent samples, the repellent behavior of eigenvalues should, if anything, increase the accuracy of the test statistic.

4. Approximations of SLCE limiting distributions

4.1. Product model for limiting distributions

Motivated by Section 2.3, we consider the following candidate for the limiting distribution of the ESD $\mu_{\mathbf{S}}$ for the SLCE matrix $\mathbf{S} = \mathbf{Y}^\dagger \mathbf{Y}$ when X_t is a symmetric EGGC or, more generally, $\mathcal{L}([X]_t) \in T(\mathbb{R})$. Let $\widetilde{\mathbf{Y}}$ denote an $N \times p$ matrix with i.i.d. standard normal entries. We then consider products of the form

$$\frac{1}{N}\left(\widetilde{\mathbf{Y}}\mathbf{T}^{1/2}\right)^\dagger\left(\widetilde{\mathbf{Y}}\mathbf{T}^{1/2}\right) = \mathbf{T}^{1/2}\widetilde{\mathbf{S}}\mathbf{T}^{1/2} \sim \widetilde{\mathbf{S}}\mathbf{T}$$

where $\widetilde{\mathbf{S}} = N^{-1}\widetilde{\mathbf{Y}}^\dagger\widetilde{\mathbf{Y}}$ is a M–P type sample covariance matrix and $\mathbf{T}$ is a random $p \times p$ diagonal matrix whose entries are i.i.d. and drawn from the fixed distribution $[X]_T$.

As discussed in Section 2.1, when X_t is a symmetric EGGC then there exists some GGC subordinator G_t such that $\mathcal{L}(X_t) = \mathcal{L}(B_{G_t}) = \mathcal{L}(Z\sqrt{G_t})$, where B_t denotes a standard Brownian motion process without drift and $\mathcal{L}(Z) = \mathcal{N}(0,1)$, both independent from G_t. As such, the matrices $\mathbf{Y}$ can be written as the Hadamard product of a M–P type matrix like $\widetilde{Y}$ with an $N \times p$ matrix whose entries are i.i.d. and follow the distribution $\sqrt{G_{T/N}}$. As far as the author is aware, random matrix theory techniques have not seen much application to Hadamard products of matrices, apart from specialized results involving random sparse matrices. However, the product $\widetilde{\mathbf{S}}\mathbf{T}^{1/2}$, in which the columns of $\widetilde{\mathbf{Y}}$ are multiplied by the square root of the (random) variance of the process B_{G_t}, acts as a good approximation. Much more importantly, the form $\widetilde{\mathbf{S}}\mathbf{T}$ is amenable to approaches from random matrix theory (see Bai and Zhou 2008). Specifically, ESDs for random matrices of the form $\widetilde{\mathbf{S}}\mathbf{T}$ are known to have a limiting distribution with Stieltjes transform $S(z)$ given implicitly by

$$S(z) = \int_{\mathbb{R}} \frac{1}{w\big(1-\lambda-\lambda z S(z)\big) - z}\, \mathrm{d}H(w) \tag{10}$$

where H is the measure corresponding to the distribution of the diagonal entries of $\mathbf{T}$, which reduces to the M–P law when $H = \delta_1$.

4.2. Simulations when quadratic variation is unknown

Yao *et al.* (2015) propose a numerical scheme for approximating the density of $\mu_{\widetilde{\mathbf{S}}\mathbf{T}}$, which involves the computation of a modified form of (10) for the companion Stieltjes transform $\underline{S}(z) = -(1-\lambda)/z + \lambda S(z)$,

$$\underline{S}(z) = \frac{1}{-z + \lambda \int_{\mathbb{R}} \frac{w}{1+w\underline{S}(z)}\, \mathrm{d}H(w)} \tag{11}$$

Specifically, (11) has a unique fixed point for all $z \in \mathbb{C}^+$, and $(1/\lambda\pi)\,\mathrm{Im}(\underline{S}(x+\mathrm{i}\epsilon))$ converges to the continuous density of the limiting spectral distribution of $\mu_{\widetilde{\mathbf{S}}\mathbf{T}}$ as $\epsilon \to 0^+$. An approximation can be produced by fixing some small $\epsilon > 0$, and iterating the map

$$s \mapsto \frac{1}{-(x+\mathrm{i}\epsilon) + \lambda \int_{\mathbb{R}} \frac{w}{1+ws}\, \mathrm{d}H(w)} \tag{12}$$

Under such a scheme, the integral $\int_{\mathbb{R}}(w/(1+ws))\,\mathrm{d}H(w)$ can be evaluated numerically when an analytic description of $\mathcal{L}(\nu) = \mathcal{L}([X]_T)$ is known. Consider, however, the example presented in Figure 1, where the entries of a large 2000×500 matrix are i.i.d. lognormal random variables. As discussed in Section 2.1, the lognormal distribution is known to be GGC, so it can be realized as the distribution of a Lévy process at a fixed time. On the other hand, there is no convenient analytic expression for the distributions of the process for arbitrary $t > 0$, nor for the associated quadratic variation process. This is the case for many GGC and EGGC processes derived from generalized inverse Gaussian distributions (Bondesson 1992), such as the Student's-t distributions, generalized hyperbolic distributions, and skew generalized hyperbolic secant distributions (Fischer 2014), as membership in such classes is not time-invariant.

We propose the following approach to this problem. First, fix some desired p and N. Consider the SLCE with parameters $\lambda = p/N$, $T = 1$, and X_t chosen so that $\mathcal{L}(X_{1/N})$ matches the desired distribution of the entries, such as LogNormal$(0, 1)$. Then, although the quadratic variation $[X]_1$ may not have an analytic description, it will be closely approximately by the sample variance of N i.i.d. samples following the distribution $X_{1/N}$. We can now proceed by fixing some large $M \in \mathbb{N}$

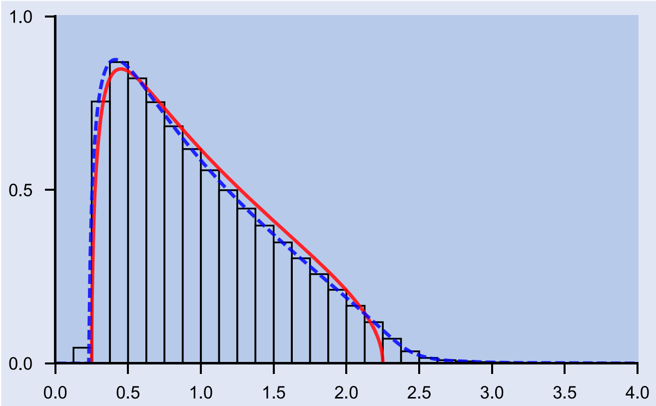

Figure 3. Histogram: One million eigenvalues aggregated from 2000 different 500×500 SLCE matrices $\mathbf{S} = \mathbf{Y}^{\dagger}\mathbf{Y}$, where $\mathbf{Y}$ is a pseudo-random $N \times p = 2000 \times 500$ matrix with i.i.d. LogNormal(0, 1) entries, normalized to have mean zero and unit variance. Excess kurtosis of the individual entries is approximately 1.109×10^2. In red: The density of the M–P distribution m_λ for $\lambda = p/N = 1/4$. In blue (dashed): The approximate density for the proposed limiting spectral distribution, calculated according to Algorithm 1 with $M = 10^6$, $\epsilon = 10^{-8}$, $\epsilon' = 10^{-6}$.

and considering a large array of M samples of the sample variance

$$\widehat{\sigma}^2 = [\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \ldots, \widehat{\sigma}_M^2]$$

defined as

$$\frac{N}{N-1}\widehat{\sigma}_j^2 = \frac{1}{N-1}\sum_{k=1}^{N}\left(y_{jk} - \frac{1}{N}\sum_{l=1}^{N} y_{jl}\right)^2$$

where the y_{jk} for $1 \leq j \leq M$ and $1 \leq k \leq N$ are i.i.d. samples of the chosen distribution. Each $\widehat{\sigma}_j^2$ is approximately distributed according to $[X]_1$. The integral in (12) can now be approximated using the discrete measure

$$\widehat{H} = \frac{1}{M}\sum_{j=1}^{M}\delta_{\widehat{\sigma}_j^2}$$

This method is summarized in Algorithm 1, and demonstrated in Figure 3. We see that those outliers in Figure 1 are not anomalous, and the eigenvalues of matrices of this type lie in a bulk which can be quite accurately predicted by this method. The bulk no longer exhibits a right edge, and a few rare eigenvalues as large as 12, 19, and 20 were observed in the random matrices generated. On the other hand, the smallest eigenvalues observed clustered around a left edge of about 0.2226, while the approximated values of $f(x)$ jump from 6.31×10^{-7} at $x = 0.232$ to 9.94×10^{-2} at $x = 0.234$.

5. SLCE for modeling intraday covariance noise

5.1. Empirical deviations from M–P law

We conclude with a few brief observations about the empirical structure of asset returns on daily and intraday timescales, and an example of the similar scaling of covariance noise which occurs under the SLCE model. We consider the universe of the S&P 500 (SPX) and Nikkei 225 (NKY) indices over two timeframes: an extended daily period of June 2013 through May 2017, and a shorter intraday minute-by-minute period from January through May of 2017. The daily timeframe provides 908 datapoints for the SPX versus 895 for the NKY, while the intraday timeframe exhibits approximately 40,000 min containing price-changing tick data on the SPX asset versus 31,000 on those in the NKY. Linear returns are computed from prices and then standardized by factoring out the cross sectional volatility, acting as a first approximation in order to isolate stationary behavior. Expected returns are not removed through subtracting the sample mean or by more sophisticated methods, as they are many orders of magnitude smaller than the volatility. If $\mathbf{r}_j$ for $j = 1, \ldots, N$ denotes the vector of returns over the jth period, then the standardized return vector $\underline{\mathbf{r}}_j$ is given by

$$\underline{\mathbf{r}}_j = \frac{\mathbf{r}_j}{\|\mathbf{r}_j\|}$$

An $N \times p$ matrix $\mathbf{X}$ is then composed of the N row vectors $\underline{\mathbf{r}}_j$, $j = 1, \ldots, N$. Figure 4 shows the histograms corresponding to the logarithms of the eigenvalues of the sample covariance matrix $N^{-1}\mathbf{X}^{\dagger}\mathbf{X}$.

Figure 4(a) and 4(c) are similar to those demonstrated in previous applications of random matrix theory to such datasets (for recent examples, see Livan *et al.* 2011, Singh and Xu 2016, Bun *et al.* 2017). Although the overlayed M–P distributions do not immediately coincide with the histograms, it is possible that a rescaling (represented by a horizontal shift of the solid red lines) might capture a decent portion of the bulk. On the other hand, a rescaling cannot widen or shrink

Algorithm 1 Approximate limiting spectral density

procedure ApproxDensity(x, p, N, M, ProbDist, ϵ, ϵ') ▷ Approximate the density $f(x)$
 $\lambda \leftarrow p/N$
 for $j = 1, \ldots, M$ **do** ▷ Sample the N-sample variance of ProbDist
  $\widehat{\sigma}^2[j] \leftarrow$ Var (Sample(ProbDist, size $= N$))
 $s \leftarrow i$
 $s_{\text{last}} \leftarrow i + i\epsilon'$
 while $|s - s_{\text{last}}| \geq \epsilon'$ **do** ▷ Stop when consecutive iterations are close
  $s_{\text{last}} \leftarrow s$
  $s \leftarrow \left(-x - i\epsilon + \lambda \times M^{-1} \times \sum_{j=1}^{M}\left[\widehat{\sigma}^2[j] \times \left(1 + \widehat{\sigma}^2[j] \times s_{\text{last}}\right)^{-1}\right]\right)^{-1}$
 return $\text{Im}(s)/\lambda\pi$ ▷ By Stieltjes inversion

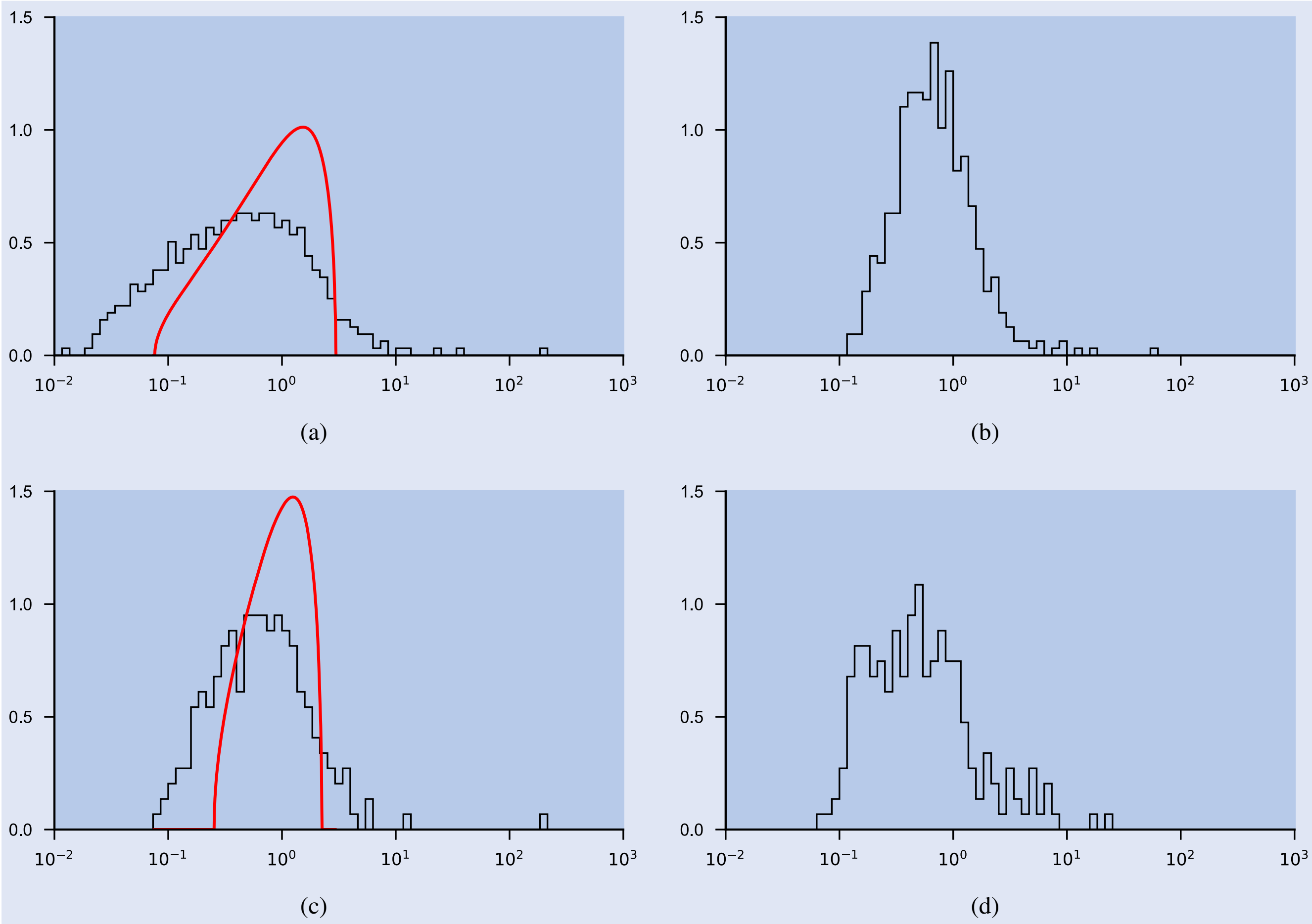

Figure 4. Log-plot of the empirical eigenvalues for sample covariance matrices of assets belonging to the S&P 500 (top row) and Nikkei 225 (bottom row) indices, for the periods June 2013–May 2017 (daily, left column) and January 2017–May 2017 (minute-by-minute, right column). Left column figures show the density of the logarithm of the M–P density (red) for appropriate values of λ. (a) 476 assets belonging to the S&P 500 Index with daily return data recorded from June 2013–May 2017. $N = 908$, $\lambda = 476/908 \approx 0.5242$. (b) Assets from (a) with intraday minute-by-minute data taken from January to May of 2017. $N = 40156$, $\lambda = 476/40156 \approx 0.0119$. (c) 221 assets belonging to the Nikkei 225 Index over the same period as (a), $N = 895$, $\lambda = 221/895 \approx 0.2469$. (d) Assets from (c) with intraday minute-by-minute data taken from January to May of 2017. $N = 30717$, $\lambda = 221/30717 \approx 0.0072$.

the densities on a logarithmic plot. The constant width of the logarithmic M–P distribution is equal to

$$w_\lambda = \log_{10}(1+\sqrt{\lambda})^2 - \log_{10}(1-\sqrt{\lambda})^2 = 2\log_{10}\frac{1+\sqrt{\lambda}}{1-\sqrt{\lambda}}$$

For Figure 4(a) and 4(c), the values w_λ are approximately 1.5914 and 0.9471, respectively. Compared to the lengths of the ticks in these figures, this is large enough to contain at least some of the bulk. For the right Figure 4(b) and 4(d), however, the values of w_λ are minuscule (0.1899 and 0.1477), and cannot account for any reasonable subset of the eigenvalues observed.

5.2. Covariance noise modeling for NKY using SLCE

Modeling using SLCE given a particular Lévy process can be done by selecting appropriate values of T, N, and p. In the daily dataset, the window is approximately 900 days long with a total of $N = 900$ datapoints, while for the second it is around 100 days long with a total of 31,000 (for NKY) datapoints. As a toy model, we consider the pure noise case where returns are stationary and independent, following identical NIG processes and their corresponding ensembles. Scaling is chosen by taking $T = 900\tau$ when modeling the first window and $T = 100\tau$ when modeling the second, where $\tau = 5 \times 10^{-3}$ is chosen so that the kurtosis of the entries is of the same order of magnitude as that observed in the NKY dataset. Figure 5(c) and 5(d) show the eigenvalues of a single sample from each model, along with the density estimated according to the techniques outlined in Section 4.2.

Unlike the M–P case, whose bulk nearly disappears in the intraday parameter range $\lambda \approx 0.0072$, the SLCE maintains a shape similar to NKY as scaling occurs. The approximated density for the ensemble on the minute-scale in Figure 5(d) is much closer to the shape of the bulk visible in the actual NKY eigenvalues in Figure 5(b). Under analysis motivated by the M–P distribution, one would necessarily conclude that nearly all eigenvalues in the minute-by-minute NKY data represent significant factors if all other assumptions on the returns held true, while for the SLCE it becomes unclear whether there are any at all.

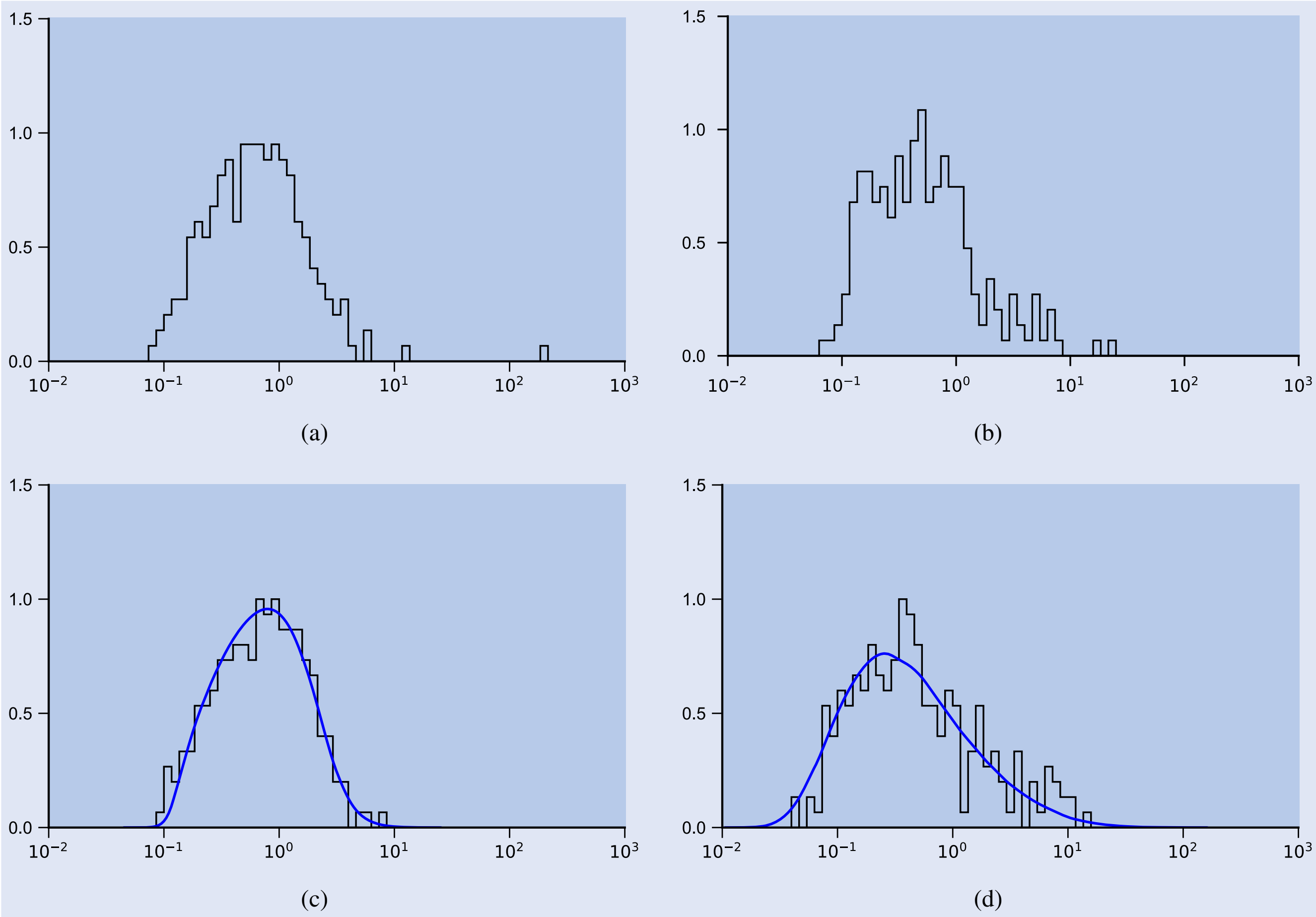

Figure 5. Log-plot of the empirical eigenvalues for sample covariance matrices of assets belonging to the Nikkei 225 (top row) indices, and randomly generated data (bottom row), for daily-scaled data (June 2013–May 2017, left column) and minute-scaled data (January 2017-May 2017, right column). (a) 221 assets belonging to the Nikkei 225 Index, $N = 895$, $\lambda = 221/895 \approx 0.2469$. (b) Assets from (a) with intraday minute-by-minute data taken from January to May of 2017. $N = 30717$, $\lambda = 221/30717 \approx 0.0072$. (c) Eigenvalues from a matrix drawn from the SLCE with i.i.d. NIG entries, $T = 0.005 \times 900$, where N and p match (a), along with the estimated density for the ensemble (solid blue line). (d) Eigenvalues from a matrix drawn from the SLCE with i.i.d. NIG entries, $T = 0.005 \times 100$, where N and p match (b), along with the estimated density for the ensemble (solid blue line).

6. Conclusion

This study has proposed a new type of random matrix model which is distinct from others in random matrix theory, including those which have seen applications in industry. In spite of this, it is our goal to convince the reader that the design of these ensembles is quite natural. One interpretation of (8) is that increasing the observations of a Brownian motion process will decrease the error of the sample variance, with no lower bound. This is true regardless of whether additional observations are made by extending the time horizon with discrete points in the future, or by refining the number of points sampled during the current horizon. Such a statement cannot be true for a non-Gaussian Lévy process. Despite the generous sufficient conditions for the M–P law, it must still be understood as a statement about matrices whose asymptotic behavior is Gaussian-like. Recent proofs and generalizations of the law provide support for this viewpoint (Yaskov 2016a, 2016b), where it is framed as a type of concentration phenomenon much like how large multivariate Gaussian random vectors cluster around the ellipse determined by their covariance matrices. The use of the M–P law to drive financial models may therefore be viewed as a first approximation with strong underlying normality assumptions.

Theorem 2.1 guarantees the existence of a limiting distribution in the case where the driving process X_t is essentially bounded. However, our experiments have been presented in the larger context of the EGGC class, as this appears to be the correct class for which the methods of Section 4.1 produce accurate results. This is particularly convenient because the GGC and EGGC classes have been well studied in the mathematical literature and contain those Lévy processes popular in the modeling of asset prices. More recently, Aoyama *et al.* (2008) introduced a class $M(\mathbb{R})$ of Lévy processes with the condition that the Lévy measure has a density of the form $\mathrm{d}\Pi(x) = |x|^{-1}g(x^2)\,\mathrm{d}x$ $(x \neq 0)$ for some CM function g. These correspond precisely to those symmetric Lévy processes X_t such that $[X]_t$ is GGC. As all symmetric EGGCs also lie in this Aoyama class $M(\mathbb{R})$, it encompasses many of the examples discussed in Section 2.1. This class is slightly larger than the class of symmetric EGGCs, containing members such as the process with Lévy measure

$$\mathrm{d}\Pi(x) = \frac{1}{|x|}\,\mathrm{e}^{-x^2}\,\mathrm{d}x, \quad x \neq 0$$

Unfortunately, examples produced in this gap $M(\mathbb{R})\backslash T(\mathbb{R})$ do not correspond to distributions with fast numerical sampling methods, and so it has not been feasible to include them in this discussion. It is the conjecture of the author that this wider class $M(\mathbb{R})$ is likely the appropriate setting for the numerical methods outlined in this paper.

Although not demonstrated here, additional simulations have provided evidence that Algorithm 1 is effective at approximating the limiting spectral distribution of the SLCE with Lévy process X_t taken as an α-stable process ($0 < \alpha < 2$), as these are EGGC. Because of the stability of such an X_t, this ensemble encompasses the case of large $N \times p$ matrices $\mathbf{Y}$ whose entries are i.i.d. (for all N) α-stable random variables, and describes the spectra of their heavy-tailed sample covariance matrices $N^{-2/\alpha}\mathbf{Y}^{\dagger}\mathbf{Y}$, sometimes called Wishart–Lévy matrices. Although analytic descriptions of these limiting distributions were established in Belinschi *et al.* (2009), they are only given implicitly except in a few specific cases. Numerical simulations have been conducted in the modified case of free Wishart–Lévy matrices (Politi *et al.* 2010), although it is known that the limiting spectral distributions of these ensembles deviate slightly (see, for example, Burda and Jurkiewicz 2011). As a result, Algorithm 1 appears to provide a convenient way to estimate these limiting distributions without the need to sample the spectra of thousands of pseudo-random matrices.

The similarities between Figures 5(a)–5(b) and 5(c)–5(d) are not the result of any complicated modeling or parameter fitting of the underlying asset behavior. The bottom figures are constructed under the (certainly false) hypothesis that fluctuations in the market are the result of complete noise, with no underlying covariance structure or factors. One interpretation of the remarkable similarities is that there are significantly fewer factors present in the market than were previously inferred by modeling noisy factors in principal component analysis on the M–P law. It would be interesting to see figures produced using well fitted Lévy processes based on higher frequency data, as techniques in this area have become quite advanced (see Feng and Lin 2013).

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

G. L. Zitelli http://orcid.org/0000-0002-6769-265X

References

Aoyama, T., Maejima, M. and Rosiński, J., A subclass of type G selfdecomposable distributions on $\mathbb{R}^d$. *J. Theoret. Probab.*, 2008, **21**, 14–34.
Bai, Z. and Silverstein, J.W., *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed., Springer Series in Statistics, 2010 (Springer: New York).
Bai, Z. and Zhou, W., Large sample covariance matrices without independence structures in columns. *Statist. Sinica*, 2008, **18**, 425–442.
Bai, Z., Liu, H. and Wong, W.K., Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Math. Finance*, 2009, **19**, 639–667.
Baik, J. and Silverstein, J.W., Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 2006, **97**, 1382–1408.
Barndorff-Nielsen, O.E., Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.*, 1997, **24**, 1–13.
Belinschi, S., Dembo, A. and Guionnet, A., Spectral measure of heavy tailed band and covariance random matrices. *Commun. Math. Phys.*, 2009, **289**, 1023–1055.
Ben Arous, G. and Guionnet, A., The spectrum of heavy tailed random matrices. *Commun. Math. Phys.*, 2008, **278**, 715–751.
Benaych-Georges, F. and Cabanal-Duvillard, T., Marčenko-Pastur theorem and Bercovici-Pata bijections for heavy-tailed or localized vectors. *ALEA Lat. Am. J. Probab. Math. Stat.*, 2012, **9**, 685–715.
Bickel, P.J. and Levina, E., Regularized estimation of large covariance matrices. *Ann. Statist.*, 2008, **36**, 199–227.
Biroli, G., Bouchaud, J.P. and Potters, M., On the top eigenvalue of heavy-tailed random matrices. *Europhys. Lett.*, 2007a, **78**, Art. 10001.
Biroli, G., Bouchaud, J.P. and Potters, M., The student ensemble of correlation matrices: eigenvalue spectrum and Kullback-Leibler entropy. *Acta Phys. Polon. B*, 2007b, **38**, 4009–4026.
Bondesson, L., *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*, Lecture Notes in Statistics Vol. 76, 1992 (Springer-Verlag: New York).
Bouchaud, J.P. and Potters, M., Financial applications of random matrix theory: a short review. In *The Oxford Handbook of Random Matrix Theory*, pp. 824–850, 2011 (Oxford University Press: Oxford).
Bun, J., Bouchaud, J.P. and Potters, M., Cleaning large correlation matrices: tools from random matrix theory. *Phys. Rep.*, 2017, **666**, 1–109.
Burda, Z. and Jurkiewicz, J., Heavy-tailed random matrices. In *The Oxford Handbook of Random Matrix Theory*, edited by G. Akemann, J. Baik, and P. Di Francesco, pp. 30–59, 2011 (Oxford University Press: Oxford).
Burda, Z.A., Jarosz, A., Nowak, M.A., Jurkiewicz, J., Papp, G. and Zahed, I., Applying free random variables to random matrix analysis of financial data. Part I: The Gaussian case. *Quant. Finance*, 2011, **11**, 1103–1124.
Capitaine, M. and Donati-Martin, C., Spectrum of deformed random matrices and free probability [online], 2016. Available online at: https://hal.archives-ouvertes.fr/hal-01346371 (accessed 21 December 2017).
Carr, P., Geman, H., Madan, D.B. and Yor, M., The fine structure of asset returns: An empirical investigation. *J. Bus.*, 2002, **75**, 305–332.
Couillet, R. and Debbah, M., *Random Matrix Methods for Wireless Communications*, 2011 (Cambridge University Press: Cambridge, MA).
El Karoui, N., On the realized risk of high-dimensional Markowitz portfolios. *SIAM J. Financial Math.*, 2013, **4**, 737–783.
Feng, L. and Lin, X., Inverting analytic characteristic functions and financial applications. *SIAM J. Financial Math.*, 2013, **4**, 372–398.
Fischer, M.J., *Generalized Hyperbolic Secant Distributions: With Applications to Finance*, Springer Briefs in Statistics, 2014 (Springer: Heidelberg).
Jeanblanc, M., Yor, M. and Chesney, M., *Mathematical Methods for Financial Markets*, Springer Finance, 2009 (Springer-Verlag London, Ltd.: London).
Jondeau, E., Poon, S.H. and Rockinger, M., *Financial Modeling Under Non-Gaussian Pistributions*, Springer Finance, 2007 (Springer-Verlag London, Ltd.: London).
Koponen, I., Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E*, 1995, **52**, 1197.

Laloux, L., Cizeau, P., Bouchaud, J.P. and Potters, M., Noise dressing of financial correlation matrices. *Phys. Rev. Lett.*, 1999, **83**, 1467–1470.
Livan, G., Alfarano, S. and Scalas, E., Fine structure of spectral properties for random correlation matrices: An application to financial markets. *Phys. Rev. E*, 2011, **84**, 016113.
Madan, D.B. and Seneta, E., The variance gamma (VG) model for share market returns. *J. Bus.*, 1990, **63**, 511–524.
Maejima, M., Classes of infinitely divisible distributions and examples. In *Lévy Matters V*, edited by L.N. Andersen, S. Asmussen, F. Aurzada, P.W. Glynn, M. Maejima, M. Pihlsgård, and T. Simon, Vol. 2149, pp. 1–65, 2015 (Springer: Cham).
Mantegna, R.N. and Stanley, H.E., Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.*, 1994, **73**, 2946–2949.
Mantegna, R.N. and Stanley, H.E., Scaling behaviour in the dynamics of an economic index. *Nature*, 1995, **376**, 46–49.
Marčenko, V.A. and Pastur, L.A., Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1967, **1**, 457.
Pascucci, A., *PDE and Martingale Methods in Option Pricing*, Bocconi & Springer Series Vol. 2, 2011 (Springer, Bocconi University Press: Milan).
Paul, D. and Aue, A., Random matrix theory in statistics: a review. *J. Statist. Plann. Inference*, 2014, **150**, 1–29.
Politi, M., Scalas, E., Fulger, D. and Germano, G., Spectral densities of Wishart-Lévy free stable random matrices. *Eur. Phys. J. B*, 2010, **73**, 13–22.
Sato, K.I., *Lévy Processes, Pnfinitely Divisible Distributions*, Cambridge Studies in Advanced Mathematics, Vol. 68, 2013 (Cambridge University Press: Cambridge), Translated from the 1990 Japanese original. Revised edition of the 1999 English translation.
Silverstein, J.W., Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 1995, **55**, 331–339.
Singh, A. and Xu, D., Random matrix application to correlations amongst the volatility of assets. *Quant. Finance*, 2016, **16**, 69–83.
Thorin, O., On the infinite divisibility of the lognormal distribution. *Scand. Actuar. J.*, 1977a, **1977**, 121–148.
Thorin, O., On the infinite divisibity of the Pareto distribution. *Scand. Actuar. J.*, 1977b, **1977**, 31–40.
Voit, J., *The Statistical Mechanics of Financial Markets*, 3rd ed., Texts and Monographs in Physics, 2005 (Springer-Verlag: Berlin).
Yao, J., Zheng, S. and Bai, Z., *Large Sample Covariance Matrices and High-dimensional Data Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 39, 2015 (Cambridge University Press: New York).
Yaskov, P., Necessary and sufficient conditions for the Marchenko-Pastur theorem. *Electron. Commun. Probab.*, 2016a, **21**, Paper No. 73, 1–8. doi:10.1214/16-ECP4748.
Yaskov, P., A short proof of the Marchenko-Pastur theorem. *C. R. Math. Acad. Sci. Paris*, 2016b, **354**, 319–322.

Appendix. Proof of Theorem 2.1

Proof Our goal is to show that the SLCE satisfies the conditions of Benaych-Georges and Cabanal-Duvillard (2012, Theorem 3.2) in some appropriate sense. We define

$$\mathbf{X}_N = \sqrt{N}\mathbf{Y}_N$$

where $\mathbf{Y}_N$ is the $N \times p$ matrix appearing in the SLCE. Then the distribution of the i.i.d. entries in $\mathbf{X}_N$ follow $\sqrt{N}X_{T/N}$. As the proof of Benaych-Georges and Cabanal-Duvillard (2012, Theorem 3.2) relies only on Benaych-Georges and Cabanal-Duvillard (2012, Theorem 2.6 and Proposition 2.7), which themselves only make assumptions about the even moments, we note that $\mu_{(\lambda,\mathbf{c})}$ only relies on λ and the sequence $\{c_{2n}\}_{n\in\mathbb{N}}$. We desire to show that

$$\frac{\mathbb{E}[|[\mathbf{X}]_{1,1}|^{2n}]}{N^{n-1}} = \frac{\mathbb{E}[(\sqrt{N}X_{T/N})^{2n}]}{N^{n-1}} = N \cdot \mathbb{E}[X_{T/N}^{2n}] = N \cdot m_{2n}[X_{T/N}]$$

converges, as $N \to \infty$, to some sequence c_{2n} for which $\sqrt[2n]{c_{2n}}$ is bounded.

By definition of the cumulants, we observe that the moments $m_{2n}[X_{T/N}]$ can be expressed as sums of products of the form

$$\prod_{j=1}^{2n} \kappa_j[X_{T/N}]^{k_j} = \prod_{j=1}^{2n} \left(\frac{T}{N}\right)^{k_j} \kappa_j[X_1]^{k_j}$$

with $k_j \in \{0, 1, 2, \ldots, 2n\}$ such that $\sum_{j=1}^{2n} j \cdot k_j = 2n$. This condition also implies that

$$1 \leq \sum_{j=1}^{2n} k_j \leq 2n$$

Consequently, the expression $N \cdot m_{2n}[X_{T/N}]$ can be written as the sum of terms of the form

$$\left(\frac{T}{N}\right)^{\sum_{j=1}^{2n} k_j - 1} \prod_{j=1}^{2n} \kappa_j[X_1]^{k_j}$$

The highest order term in N corresponds to the single choice $k_{2n} = 1$ and $k_j = 0$ for $j = 1, \ldots, 2n-1$, which occurs once in the expansion for $m_{2n}[X_T/N]$ as the term (with unit coefficient) $\kappa_{2n}[X_{T/N}]$. Therefore, we can write

$$N \cdot m_{2n}[X_{T/N}] = T\kappa_{2n}[X_1] + O(N^{-1}) = \kappa_{2n}[X_T] + O(N^{-1})$$

It follows that

$$\frac{\mathbb{E}[|[\mathbf{X}]_{1,1}|^{2n}]}{N^{n-1}} \xrightarrow{N\to\infty} \kappa_{2n}[X_T] = c_{2n}$$

Since X_t is essentially bounded, there exist constant $B > 0$ such that $\sqrt[2n]{\kappa_{2n}[X_T]} \leq B$, which shows that $\sqrt[2n]{c_{2n}}$ is bounded. Since

$$\mathbf{Y}_N^\dagger \mathbf{Y}_N = \frac{1}{N}\mathbf{X}_N^\dagger \mathbf{X}_N$$

it follows from Benaych-Georges and Cabanal-Duvillard (2012, Theorem 3.2) that the ESD for $\mathbf{S}_N$ has a limiting distribution which depends only on λ and the even moments $\kappa_{2n}[X_T]$. We write this distribution as $\mu_{(X,T,\lambda)}$, and note that it depends continuously on λ, and also continuously on T by virtue of the relationship between $\kappa_{2n}[X_T]$ and c_{2n}. Continuity and additional properties follow from the other results in Benaych-Georges and Cabanal-Duvillard (2012). ■