Statistica Sinica **32** (2022), 43-63
doi:https://doi.org/10.5705/ss.202019.0046

SPECTRAL DISTRIBUTION OF THE SAMPLE COVARIANCE OF HIGH-DIMENSIONAL TIME SERIES WITH UNIT ROOTS

Alexei Onatski and Chen Wang

University of Cambridge and University of Hong Kong

Abstract: We study the empirical spectral distributions of two sample-covariance-type matrices associated with high-dimensional time series with unit roots. The first matrix is $\boldsymbol{\mathcal{S}} = \boldsymbol{X}\boldsymbol{X}'/T$, where $\boldsymbol{X}$ is an $n \times T$ data set, with rows represented by n independent and identically distributed (i.i.d.) copies of T consecutive observations of a difference-stationary process. The second matrix is $\boldsymbol{\mathcal{W}} = n\int_0^1 \boldsymbol{W}_n(t)\,\boldsymbol{W}_n(t)'\,\mathrm{d}t$, where $\boldsymbol{W}_n(t)$ is an n-dimensional vector with i.i.d. Brownian motion components. We show that as n and T diverge to infinity proportionally, the two distributions weakly converge to non-random limits. The limit corresponding to $\boldsymbol{\mathcal{S}}$ has a density $\varphi(x)$ that decays as $x^{-3/2}$ when $x \to \infty$. The limit corresponding to $\boldsymbol{\mathcal{W}}$ is a Feller-Pareto distribution. An illustrative application is provided.

Key words and phrases: Empirical spectral distribution, Feller-Pareto distribution, non-stationary time series, sample covariance, Stieltjes transform.

1. Introduction

High-dimensional sample covariance matrices are fundamental to modern multivariate analysis. An important summary of such matrices is provided by the empirical distribution of their eigenvalues (ESD, or empirical spectral distribution). Graphically, one can associate the ESD with Cattell (1966) scree plot, an effective tool for detecting a low-dimensional structure in noisy data. When the sample size is much larger than the dimension of the data, the structure appears in the graph as a small number of eigenvalues that form a steep slope above the flat “scree region” resulting from the remaining “noise eigenvalues.”

In modern applications in which the data dimensionality is non-negligible relative to the sample size, the sample eigenvalues are much more spread out than their population counterparts, which obscures the slope-scree division (see Johnstone and Paul (2018) for a recent review of related results). In such situations, one needs a benchmark for the empirical distribution of the noise eigenvalues. A

Corresponding author: Alexei Onatski, Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, UK. E-mail: ao319@cam.ac.uk.

disagreement of the scree plot with the benchmark indicates the presence of the structure (see Luo et al. (2017); Melkiev, Guo and Wirjanto (2018) for recent applications).

A natural benchmark for the ESD of the noise is its limiting version (LSD, or limiting spectral distribution) when the dimensionality and the sample size diverge to infinity proportionally. For example, when the noise is independent and identically distributed (i.i.d.) and has a finite second moment, the corresponding LSD is the Marchenko-Pastur distribution (see Theorem 3.6 in Bai and Silverstein (2010)).

There are many extensions of this result to correlated noise. Typically, the LSD is described using its Stieltjes transform (see Section 2), which, in turn, is characterized as a solution to a functional equation. In time series settings, such extensions have been obtained by Jin et al. (2009); Burda et al. (2010); Pfaffel and Schlemm (2011); Yao (2012); Merlevède and Peligrad (2016), among others. However, none of these extensions can handle time series data with unit roots. The goal of this study is to fill in this gap.

Stochastically trending unit-root data are widespread in economics and finance. Here, principal components analyses in high-dimensional settings have attracted much recent attention from both applied and theoretical researchers (see, e.g., Bai and Ng (2004); Engel, Mark and West (2015); Banerjee, Marcellino and Masten (2017); Barigozzi, Lippi and Luciani (2018)). The decision on how many principal components to retain for an "adequate" data description is based on the eigenvalues of the sample covariance matrix or the scree plot. Zhang, Pan and Gao (2018) and Onatski and Wang (2021) analyzed a few of the largest eigenvalues of the sample covariance of nonstationary data. In this study, we derive the limit of the empirical distribution of all the eigenvalues, that is, the LSD.

We consider two different settings. In the first setting, the data are represented by an $n \times T$ matrix $\boldsymbol{X}$, the columns of which X_t satisfy the first difference equation $X_t - X_{t-1} = \varepsilon_t$, with the entries of the vector ε_t given by i.i.d. copies of a linear process. The corresponding sample covariance matrix is

$$\boldsymbol{\mathcal{S}} \equiv \frac{\boldsymbol{X}\boldsymbol{X}'}{T}.$$

In the second setting, the n-dimensional data $\boldsymbol{W}_n(t)$ are continuous over time $t \in [0, 1]$. The entries of vector $\boldsymbol{W}_n(t)$ are i.i.d. copies of the standard Brownian motion. The corresponding sample covariance is

$$\boldsymbol{\mathcal{W}} \equiv n \int_0^1 \boldsymbol{W}_n(t) \boldsymbol{W}_n(t)' \mathrm{d}t.$$

Here, the multiplication by n ensures the compatibility of the matrices $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{W}}$. Indeed, under suitable assumptions (e.g., Phillips and Solo (1992)) as $T \to \infty$, while n remains fixed, the probability limit of $(n/T)\,\boldsymbol{\mathcal{S}}$ is proportional to $\boldsymbol{\mathcal{W}}$. When n and T are growing proportionally, both $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{W}}$ have spectral norms of order $O\left(n^2\right)$. (Of course, the growth in T is relevant only to $\boldsymbol{\mathcal{S}}$.)

We establish the existence of the LSD for the matrix $\boldsymbol{\mathcal{S}}$, and show that its Stieltjes transform satisfies an equation similar to that obtained in the literature in the context of stationary data. A novel feature of this LSD is the unboundedness of its support from the right. In particular, we prove that the LSD has a density $\varphi(x)$ that behaves as $x^{-3/2}$ for $x \to \infty$. Because $x\varphi(x)$ is not integrable, we conclude that for any fixed $\delta > 0$, the portion of the data variation absorbed by the first δn principal components almost surely (a.s.) approaches 100% as n and T diverge to infinity proportionally. In other words, a number of principal components equal to any percentage (however small) of the data size eventually explain all the variation in the data.

For the matrix $\boldsymbol{\mathcal{W}}$, we obtain a much sharper and somewhat surprising result. Specifically, the density of its LSD is equal to

$$\psi(x) = \frac{1}{2\pi x^2}\sqrt{x - \frac{1}{16}}, \quad x > \frac{1}{16}. \tag{1.1}$$

If we denote the random variable with the above distribution as ξ, then $(16\xi)^{-1}$ follows a beta $B(1/2, 3/2)$ distribution. Therefore, ξ is a Feller-Pareto random variable (e.g., Arnold (2015)).

The rest of the paper is organized as follows. In Sections 2 and 3, we obtain the results for matrices $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{W}}$, respectively. Several Monte Carlo experiments are conducted in Section 4. Section 5 illustrates the theoretical results using financial data. Section 6 concludes the paper. All technical proofs are given in the Supplementary Material and the Appendix.

2. Matrix $\mathcal{S}$

Let

$$z_t = \sum_{k=0}^{\infty} \theta_k u_{t-k}, \tag{2.1}$$

where u_t, for $t \in \mathbb{Z}$, are i.i.d. Further, let ε_t, for $t \in \mathbb{Z}$, be an n-dimensional process that consists of n independent copies of the process z_t. Finally, let

$$X_t = X_0 + \sum_{j=1}^{t} \varepsilon_j, \tag{2.2}$$

with an arbitrary X_0.

Denote the $n \times T$ matrices with tth columns X_t and ε_t as $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$, respectively. Let $\boldsymbol{U}$ be a T-dimensional upper-triangular matrix of ones, and let l be a T-dimensional vector of ones. Then,

$$\boldsymbol{X} = X_0 l' + \boldsymbol{\varepsilon U}.$$

We are interested in the asymptotic behaviour of the ESD of the sample covariance matrix of $\boldsymbol{X}$, that is, of the matrix

$$\mathbf{S} \equiv \frac{\left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)'}{T} = \frac{\boldsymbol{\varepsilon U M U' \varepsilon'}}{T},$$

where $\boldsymbol{M}$ is the projector matrix on the space orthogonal to l. We consider the asymptotic regime where $n, T \to \infty$ such that $n/T \to c \in (0, \infty)$, which we abbreviate as $n, T \to_c \infty$.

Remark 1. For $\boldsymbol{\mathcal{S}} = \boldsymbol{X X'}/T$, we have $\operatorname{rank}(\mathbf{S} - \boldsymbol{\mathcal{S}}) \leq 1$. Therefore, by Theorem A.43 of Bai and Silverstein (2010), the LSDs of $\mathbf{S}$ and $\boldsymbol{\mathcal{S}}$ are equivalent (if they exist).

Denote the eigenvalues of $\mathbf{S}$ as $\lambda_1 \geq \cdots \geq \lambda_n$. Let $\mathbf{F}_{n,T}$ be the ESD of $\mathbf{S}$ with the cumulative distribution function (CDF)

$$\mathbf{F}_{n,T}(\lambda) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{\lambda_j \leq \lambda\},$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Below, we show that as $n, T \to_c \infty$, $\mathbf{F}_{n,T}$ a.s. weakly converges to a non-random distribution $\mathbf{F}$.

Recall that the Stieltjes transform of $\mathbf{F}$ at $z \in \mathbb{C}^+$ (the upper half of the complex plane) is defined as

$$\mathbf{m}(z) = \int \frac{1}{x - z} \mathrm{d}\mathbf{F}(x).$$

The distribution $\mathbf{F}$ can be uniquely recovered from its Stieltjes transform. Moreover, if $\lim_{z \to x} \Im \mathbf{m}(z) = \mathbf{m}_x$ exists, the distribution $\mathbf{F}$ has density at x equal to

$\mathbf{m}_x/\pi$ (e.g., Theorem B.8 of Bai and Silverstein (2010)).

Assumption 1. *The innovations* u_t *are i.i.d., with* $\mathbb{E}u_t = 0$, $\mathbb{E}u_t^2 = 1$, *and* $\mathbb{E}u_t^4 < \infty$.

Assumption 2. *The linear filter in* (2.1) *is not identically zero and is absolutely summable; that is,* $0 < \sum_{k=0}^{\infty} |\theta_k| < \infty$.

These assumptions yield, in particular, the continuity on $\omega \in [0, 2\pi]$ of the spectral density of z_t,

$$f(\omega) = \frac{1}{2\pi} \left| \sum_{k=0}^{\infty} \theta_k e^{\mathrm{i}k\omega} \right|^2 .$$

Because z_t satisfies (2.1), it is also true that it is a linearly regular process (i.e., it does not contain a perfectly linearly predictable component); hence, $f(\omega) > 0$ almost everywhere on $[0, 2\pi]$. There may, however, exist a subset of $[0, 2\pi]$ of zero Lebesgue measure, where $f(\omega) = 0$.

Let H be a Borel measure on $[0, \infty)$ defined as follows. For any Borel set $E \subseteq [0, \infty)$,

$$H(E) = \frac{1}{2\pi} \mu \left(\omega \in (0, 2\pi) : \frac{\pi f(\omega)}{1 - \cos \omega} \in E \right),$$

where $\mu(\cdot)$ is the Lebesgue measure.

In the Appendix, we obtain the following theorem on the limiting ESD of $\mathbf{S}$.

Theorem 1. *Suppose that assumptions* 1 *and* 2 *hold. Then, as* $n, T \to_c \infty$, *the ESD of* $\mathbf{S}$ *a.s. weakly converges to a distribution function* $\mathbf{F}(x)$, *such that, for any* $z \in \mathbb{C}^+$, *its Stieltjes transform* $\mathbf{m} \equiv \mathbf{m}(z)$ *is a unique in* $\mathbb{C}^+$ *solution to the equation*

$$z = -\frac{1}{\mathbf{m}} + \int_0^{\infty} \frac{x \mathrm{d}H(x)}{1 + cx\mathbf{m}},$$

which is equivalent to the equation

$$z = -\frac{1}{\mathbf{m}} + \frac{1}{2\pi} \int_0^{2\pi} \frac{\mathrm{d}\omega}{(1 - \cos \omega)(\pi f(\omega))^{-1} + c\mathbf{m}}. \tag{2.3}$$

Theorem 1 can be viewed as a generalization of Theorem 1 of Yao (2012) to difference-stationary processes. Our assumptions for the differenced process $\varepsilon_t = X_t - X_{t-1}$ match Yao's assumptions for stationary data. Yao's theorem is a special case of Theorem 1, where the process z_t has an MA unit root that "cancels" the unit root in the integrated version of z_t.

Recently, Merlevède and Peligrad (2016) extended Yao (2012) theorem in a different direction. They showed that the theorem holds for regular stationary stochastic processes with arbitrarily slowly decaying autocovariances. In particular, long-range dependent stationary processes that have a spectral density unbounded at zero are allowed. However, an essential requirement of Merlevède and Peligrad (2016) is the boundedness of the second moment of the data. This assumption is violated for time series with a unit root which we handle in Theorem 1. Whether and how our assumptions can be relaxed is an interesting question left for future research.

Theorem 1 is not explicit about the properties of the limiting spectral distribution $\mathbf{F}$. To analyze $\mathbf{F}$, we assume, following Yao (2012), that the spectral density $f(\omega)$ is bounded away from zero. In addition, we require that $f(\omega)$ has a continuous derivative in a neighborhood of $\omega = 0$.

Assumption 3. *$f(\omega) \neq 0$ for $\omega \in [0, 2\pi]$. In addition, $f(\omega)$ has a continuous derivative in a neighborhood of $\omega = 0$.*

Using the results of Silverstein and Choi (1995), it is fairly straightforward to show that, under Assumption 3, the support of $\mathbf{F}$ contains an interval that is unbounded from the right (x_0, ∞). The existence of the density $\varphi(x)$ on this interval follows from Theorem 1.1 of Silverstein and Choi (1995). That theorem, combined with our Theorem 1, implies that $\pi\varphi(x)$ is equal to the imaginary part of the unique m from the upper half of the complex plane satisfying

$$x = -\frac{1}{m} + \int \frac{x \mathrm{d}H(x)}{1 + cxm},$$

or, equivalently,

$$x = -\frac{1}{m} + \frac{1}{\pi} \int_0^{\pi} \frac{\mathrm{d}\omega}{(1 - \cos\omega)(\pi f(\omega))^{-1} + cm}. \tag{2.4}$$

In the Appendix, we study the solution of (2.4) in detail, and obtain the following theorem.

Theorem 2. *Under assumptions 1, 2, and 3, the limiting spectral distribution $\mathbf{F}$ has a density $\varphi(x)$ on (x_0, ∞) for some $x_0 > 0$. The density satisfies the following equation:*

$$\varphi(x) = \sqrt{\frac{f(0)}{2\pi c}} x^{-3/2} (1 + o(1)),$$

as $x \to \infty$.

3. Matrix $\mathcal{W}$

The goal of this section is to prove (1.1). The idea of the proof is to approximate $\boldsymbol{\mathcal{W}}$ using the sample covariance matrix of an n-dimensional random walk of large length T. Then, use Theorem 1 to characterize the LSD of the approximation as $n/T \to c > 0$. Finally, show that the LSD of $\boldsymbol{\mathcal{W}}$ can be obtained by sending c to zero.

Let η_t be an i.i.d. $N(0, I_n)$ vector, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_T]$, and $\boldsymbol{U}$ be a T-dimensional upper-triangular matrix of ones. Further, let $\boldsymbol{X} = \boldsymbol{\eta U}$ be an $n \times T$ matrix with rows that are independent Gaussian random walks. Let $\boldsymbol{S}_{n,T} = \boldsymbol{X X}'/T$ be the sample covariance matrix of the random walk. As $T \to \infty$ while n is held fixed, $\boldsymbol{X X}'/T^2$ converges in distribution to $\int_0^1 \boldsymbol{W}_n(t)\boldsymbol{W}_n(t)'\mathrm{d}t$; thus, it is natural to approximate $\boldsymbol{\mathcal{W}}$ using $(n/T)\,\boldsymbol{S}_{n,T}$.

Denote the empirical spectral distribution function of $(n/T)\boldsymbol{S}_{n,T}$ as $F^{(n/T)\boldsymbol{S}_{n,T}}(x)$. Then, Theorem 1 yields the following corollary.

Corollary 1. *As $n, T \to_c \infty$ with $c \in (0, \infty)$, $F^{(n/T)\boldsymbol{S}_{n,T}}(x)$ a.s. weakly converges to a distribution F_c, the Stieltjes transform of which, $m_c \equiv m_c(z)$, satisfies the following cubic equation:*

$$(zm_c + 1)^2 \left(c^2 m_c + 4\right) - m_c = 0. \tag{3.1}$$

Because the approximation of $\boldsymbol{\mathcal{W}}$ using $(n/T)\,\boldsymbol{S}_{n,T}$ becomes exact when $T \to \infty$, heuristically, we expect to obtain the Stieltjes transform m_0 of the limiting spectral distribution of $\boldsymbol{\mathcal{W}}$ as a solution to (3.1) with c set to zero. That is, we expect

$$m_0 = 4\,(zm_0 + 1)^2\,.$$

Employing the inversion formula for Stieltjes transforms (e.g., Theorem B.8 in Bai and Silverstein (2010)), we obtain the density stated in (1.1). To show how this heuristic argument can be formalized, we prove the following theorem in the Appendix.

Theorem 3. *As $n \to \infty$, the empirical distribution of the eigenvalues of $\boldsymbol{\mathcal{W}}$ weakly converges in probability to a distribution F_0 with density*

$$\psi(x) = \frac{1}{2\pi x^2}\sqrt{x - \frac{1}{16}}, \quad x > \frac{1}{16}.$$

4. Monte Carlo

In this section, we conduct Monte Carlo experiments on (1.1). Each experiment is based on 1,000 Monte Carlo replications. Three underlying distributions are considered for the simulated data: standard Gaussian, centered $\chi^2(1)/\sqrt{2}$, and uniform $(-\sqrt{3}, \sqrt{3})$. Here, the scalars $\sqrt{2}$ and $\sqrt{3}$ are introduced to standardize the variances of the data. Data are generated for $(n, T) = (20, 200)$ and $(n, T) = (20, 40)$. In each experiment, the eigenvalues of

$$\frac{n}{T}\boldsymbol{S}_{n,T} \equiv \frac{n\boldsymbol{\eta}\boldsymbol{U}\boldsymbol{U}'\boldsymbol{\eta}'}{T^2}$$

are simulated, with $\boldsymbol{U}$ being a T-dimensional upper-triangular matrix of ones and $\boldsymbol{\eta}$ an $n \times T$ matrix with i.i.d. entries taking one of the above three underlying distributions. For each experiment, the empirical CDF is plotted and then superimposed with the theoretical counterpart based on (1.1). The results are reported in Figure 1, from which we see that for various distributions and values of n and T, the empirical distribution matches the theoretical one well.

5. Illustration

In this section, we illustrate our theoretical results using financial data. We consider the logarithms of monthly stock prices for 33 North American companies for the period July 1963 to December 2018. We chose companies if their stocks were part of the S&P100 index at some point during this period. Furthermore, we excluded companies with missing data for some of the covered months.

To construct the log price time series, we use CRSP data on holding period returns without dividends (RETX variable). For the initial period, we set the stock prices equal to the PRC variable, which is the stock price for the last trading day of the month. Then, we construct the price series for the remaining time periods sequentially by multiplying the previous month's price by 1 + RETX. The prices obtained in this way are automatically adjusted for distribution events such as stock splits. Finally, we take the logarithm of the constructed series.

It is a standard assumption in finance that log stock prices follow random walks. Had they been independent, the ESD of the corresponding sample covariance matrix could have been well approximated by the distribution $\mathbf{F}(x)$ described in Theorem 1. Before computing the sample covariance, we need to standardize the random walks so that their innovations have unit standard deviations. Note that the cross-sectional dimension of our data, $n = 33$, is an order of magnitude smaller than their time dimension, $T = 666$. Hence, we expect that the ESD

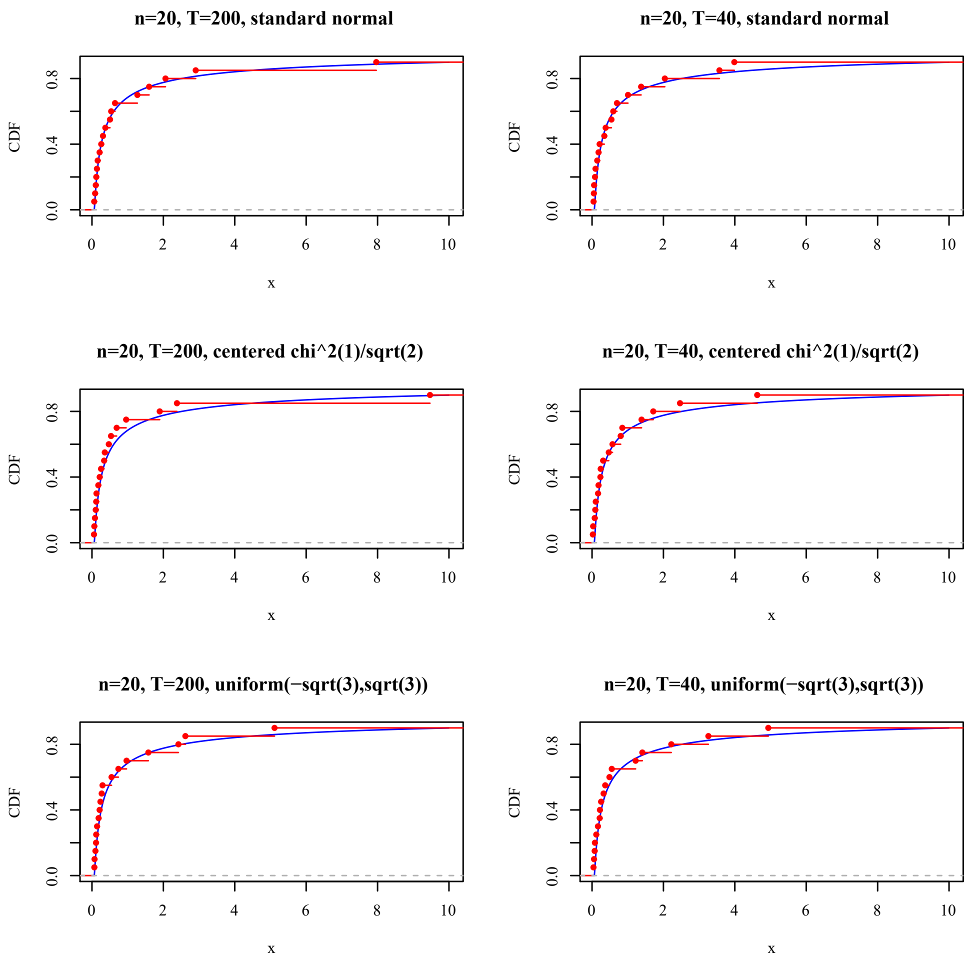

Figure 1. Comparison of the Monte Carlo empirical distribution with the Feller-Pareto distribution.

scaled by n/T is also well approximated by the Feller-Pareto distribution with density (1.1).

However, as is well known, financial log price series may contain common factors that correspond to sources of non-diversifiable risk. These common factors create cross-sectional dependence between log prices. This dependence leads to a gap between the ESD and the Feller-Pareto distribution. Such a gap can be easily depicted graphically as follows.

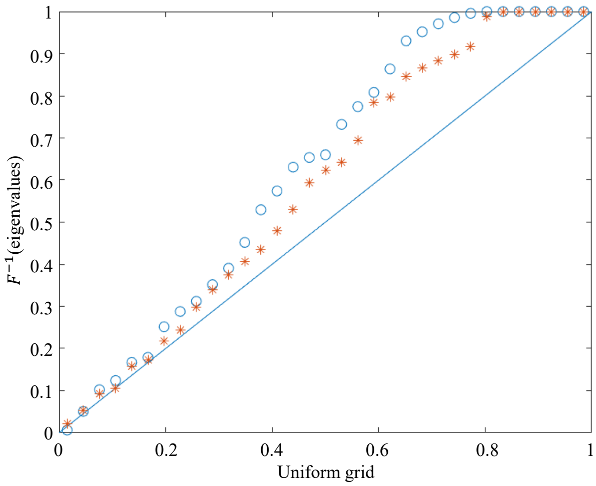

Figure 2. Comparison of ESD for financial log price data with the Feller-Pareto distribution F_B. Circles: raw data. Stars: residuals after extracting five Fama-French factors.

Let $\mathbf{X}$ be an $n\times T$ matrix of log prices, standardized so that changes in the log prices have unit standard deviation. Let $\boldsymbol{\lambda} \equiv (\lambda_1,\ldots,\lambda_n)'$ be an n-dimensional vector of the eigenvalues of $n\mathbf{X}\mathbf{X}'/T^2$, sorted in descending order of magnitude.

Recall that $1/(16\xi)$, where ξ is a Feller-Pareto random variable, is distributed as beta $B(1/2,3/2)$. Let $F_B(x)$ be the corresponding CDF. Had the rows of $\mathbf{X}$ been independent, we would have expected the entries of the vector $F_B^{-1}(\boldsymbol{\lambda}) \equiv (F_B^{-1}(\lambda_1),\ldots,F_B^{-1}(\lambda_n))'$ to be uniformly distributed on $[0,1]$. Then, a plot of these entries against a uniform grid of n points on $[0,1]$ would lie on the 45-degree line. A deviation would indicate dependence between the rows.

Figure 2 shows the plot (circle dots) for our 33 log price series. The circles lie above the 45-degree line, which means the quantiles of the ESD are larger than those of the Feller-Pareto distribution. As discussed above, this discrepancy is expected because different log prices are linked to the same sources of the non-diversifiable risk.

There is a large body of financial literature on modelling the sources of non-diversifiable risk. For example, we can do so using the five Fama-French factors in stock returns (see Fama and French (2015)). The values of the factors are available on Kenneth R. French's website. We can use the above methodology to visually assess the extent to which purging the data from these factors reduces the amount of cross-sectional dependence.

Specifically, we regress the first-differences of the log prices on the five factors, take the residuals, standardize them, and accumulate them over time. The data constructed in this way are free from the factors. The corresponding plot of the components of $F_B^{-1}(\boldsymbol{\lambda})$ against a uniform grid on $[0, 1]$ is shown in Figure 2 (star dots). As expected, the stars are closer to the 45-degree line than the circles. However, the discrepancy is not eliminated. This may be due to some remaining common industry factors. For example, five out of the 33 companies are pharmaceutical companies.

It may be interesting to further purge the data of commonalities until the entries of the vector $F_B^{-1}(\boldsymbol{\lambda})$ are reasonably uniform over $[0, 1]$. However, we feel that our goal of illustrating the above theoretical results has been achieved, and leave a more thorough investigation of financial data to future, more specialized research.

6. Conclusion

We have examined the limiting spectral distributions of sample covariance matrices of nonstationary time series data. We consider two situations. First, the data are given by n i.i.d. copies of a difference-stationary process. Second, the data are continuous over time, and are represented by n i.i.d. copies of a standard Brownian motion. In the first case, we show that the limiting spectral distribution has an unbounded support and a density $\varphi(x)$ decaying at infinity as $x^{-3/2}$. In the second case, we show that the limiting spectral distribution is Feller-Pareto. Our analysis of the first situation extends previous studies (e.g., Jin et al. (2009); Burda et al. (2010); Pfaffel and Schlemm (2011); Yao (2012); Merlevède and Peligrad (2016)) to nonstationary data with unit roots. Our theoretical results are illustrated using financial data in a quick check of cross-sectional dependence.

Supplementary Material

The online Supplementary Material (SM) provides proofs of Lemmas 1, 2, and 3 and the convergence of $\mathcal{L}(H_{\boldsymbol{\Gamma}}, H)$, which is required in the Appendix.

Acknowledgments

The authors would like to thank the associate editor and two referees for their helpful comments and suggestions. Chen Wang's research is supported by grant ECS 27308219 from the Research Grants Council of Hong Kong.

Appendix

A.1 Proof of Theorem 1

By the rank inequality (e.g. Theorem A.43 in Bai and Silverstein (2010)), it is sufficient to prove the a.s. weak convergence of the ESD of $\boldsymbol{\varepsilon A \varepsilon}'/T$, where $\boldsymbol{A} = \boldsymbol{MUMU'M}$. Let $\mathbf{\Gamma}$ be the $T \times T$ Toeplitz matrix with entries $\mathbf{\Gamma}_{ts} = \gamma_{t-s} \equiv Cov\left(z_s, z_t\right)$. We have

$$\frac{\boldsymbol{\varepsilon A \varepsilon}'}{T} = \frac{\boldsymbol{\eta}\mathbf{\Gamma}^{1/2}\boldsymbol{A}\mathbf{\Gamma}^{1/2}\boldsymbol{\eta}'}{T},$$

where $\boldsymbol{\eta} = \boldsymbol{\varepsilon}\mathbf{\Gamma}^{-1/2}$. Note that $\mathbf{\Gamma}$ must be invertible because z_t is assumed to have a representation (2.1), and thus, it is linearly regular. (Theorem 4.1 on p.569 of Doob (1953) implies that a linear process cannot be perfectly predicted from its past, which precludes singularity of $\mathbf{\Gamma}$. Had there been $a \in \mathbb{R}^T - \{0\}$ such that $a'\mathbf{\Gamma}a = 0$, $\text{var}(a'Z)$ would have been zero, and a perfect prediction would be possible, contradicting Theorem 4.1.)

Let $\mathcal{L}\left(\cdot, \cdot\right)$ be the Lévy distance between distribution functions. Let $F_{n,T}\left(\lambda\right)$ be the ESD function of $\boldsymbol{\varepsilon A \varepsilon}'/T$. Recall that Lévy distance metrizes the weak convergence. Therefore, it is sufficient to prove that $\mathcal{L}\left(F_{n,T}, \mathbf{F}\right) \to 0$ as $T \to \infty$. We will prove this by splitting $\mathcal{L}\left(F_{n,T}, \mathbf{F}\right)$ into several parts, using the triangle inequality (see (A.4) below).

In preparation for the proof, note that $\boldsymbol{A}$ is a circulant matrix (e.g. Onatski and Wang (2021)). In particular, it has the spectral decomposition

$$\boldsymbol{A} = \frac{\boldsymbol{\mathcal{F}}^{*}\text{diag}\left(0, \boldsymbol{D}\right)\boldsymbol{\mathcal{F}}}{T},$$

where $\boldsymbol{\mathcal{F}} = \left\{\exp\left(-\mathrm{i}\omega_{s-1}\left(t-1\right)\right)\right\}_{s,t=1}^{T}$ is the Fourier matrix of order T with $\omega_s = 2\pi s/T$, and

$$\boldsymbol{D} = \frac{1}{2}\text{diag}\left(\left(1-\cos\omega_1\right)^{-1}, \ldots, \left(1-\cos\omega_{T-1}\right)^{-1}\right).$$

Note that $\|\boldsymbol{A}\| = \left(1-\cos\omega_1\right)^{-1}/2$, which is unbounded asymptotically. To establish the convergence of the ESD of $\boldsymbol{\varepsilon A \varepsilon}'/T$, our strategy is to approximate $\boldsymbol{A}$ by a matrix $\bar{\boldsymbol{A}}$ with bounded norm such that Theorem 1.1 of Bai and Zhou (2008) can be applied to prove the convergence of the ESD of $\boldsymbol{\varepsilon}\bar{\boldsymbol{A}}\boldsymbol{\varepsilon}'/T$. To this end, for any $u > 0$, let

$$\cos_u \omega = \begin{cases} \cos\omega & \text{if } \frac{1}{2}(1-\cos\omega)^{-1} < u \\ 1-\frac{1}{2u} & \text{otherwise} \end{cases},$$

and let $\bar{\boldsymbol{D}}$ be obtained from $\boldsymbol{D}$ by replacing $\cos\omega_s$ with $\cos_u \omega_s$, $s = 1, \ldots, T-1$. Define

$$\bar{\boldsymbol{A}} = \frac{\boldsymbol{\mathcal{F}}^* \text{diag}\left(0, \bar{\boldsymbol{D}}\right) \boldsymbol{\mathcal{F}}}{T}.$$

It is straightforward to verify that $(1-\cos\omega)^{-1}/2 \leq \pi^2/\left(4\omega^2\right)$ for any $\omega \in [0, \pi]$. Therefore, we may have $\cos_u \omega_s \neq \cos\omega_s$ only for

$$s \leq \frac{T}{4\sqrt{u}} \quad \text{or} \quad s \geq T - \frac{T}{4\sqrt{u}}.$$

Hence,

$$\text{rank}\left(\boldsymbol{A} - \bar{\boldsymbol{A}}\right) \leq \frac{T}{2\sqrt{u}}. \tag{A.1}$$

Furthermore, since $\boldsymbol{\mathcal{F}}/\sqrt{T}$ is a unitary matrix,

$$\left\|\bar{\boldsymbol{A}}\right\| \leq u, \tag{A.2}$$

where $\|\cdot\|$ is the spectral norm.

Now let us consider $\boldsymbol{\Gamma}$. Let $\boldsymbol{C}$ be the so-called optimal or Cesàro circulant for matrix $\boldsymbol{\Gamma}$. That is, $\boldsymbol{C}$ is a T-dimensional circulant matrix with entries $\boldsymbol{C}_{st} = c_{s-t}$, where $c_0 = \gamma_0$ and

$$c_k = \frac{(T-k)\gamma_k + k\gamma_{k-T}}{T}, \quad k = 1, \ldots, T-1.$$

In Section S1 of SM, we prove the following lemma.

Lemma 1. *Assume that* $\{u_t\}$ *in* (2.1) *is a weakly stationary white noise sequence and that assumption* 2 *holds. Then,* $\|\boldsymbol{\Gamma} - \boldsymbol{C}\|_F^2 = o(T)$ *as* $T \to \infty$, *where* $\|\cdot\|_F$ *is the Frobenius norm.*

Let $H_{\boldsymbol{\Gamma}}$ be the ESD of $\boldsymbol{A}^{1/2}\boldsymbol{\Gamma}\boldsymbol{A}^{1/2}$, $\bar{H}_{\boldsymbol{\Gamma}}$ be the ESD of $\bar{\boldsymbol{A}}^{1/2}\boldsymbol{\Gamma}\bar{\boldsymbol{A}}^{1/2}$, $\bar{H}_{\boldsymbol{C}}$ be the ESD of $\bar{\boldsymbol{A}}^{1/2}\boldsymbol{C}\bar{\boldsymbol{A}}^{1/2}$, and H_u be the distribution such that, for any Borel set $E \subseteq [0, \infty)$,

$$H_u(E) = \frac{1}{2\pi}\mu\left(\omega \in (0, 2\pi) : \frac{\pi f(\omega)}{1-\cos_u \omega} \in E\right).$$

In the following, we show that $H_{\boldsymbol{\Gamma}}$ converges to H. Note that $H_{\boldsymbol{\Gamma}}$ is also the ESD of $\boldsymbol{A}^{1/2}\boldsymbol{\Gamma}\boldsymbol{A}^{1/2}$ because the eigenvalues of the latter matrix coincide with those of

$\mathbf{\Gamma}^{1/2}\boldsymbol{A}\mathbf{\Gamma}^{1/2}$. By the triangle inequality,

$$\mathcal{L}\left(H_{\mathbf{\Gamma}},H\right)\leq\mathcal{L}\left(H_{\mathbf{\Gamma}},\bar{H}_{\mathbf{\Gamma}}\right)+\mathcal{L}\left(\bar{H}_{\mathbf{\Gamma}},\bar{H}_{\boldsymbol{C}}\right)+\mathcal{L}\left(\bar{H}_{\boldsymbol{C}},H_{u}\right)+\mathcal{L}\left(H_{u},H\right).$$

In Section S2 of SM, we prove that each term on the right hand side converges to zero.

Let $\bar{F}_{n,T}(\lambda)$ be the ESD function of $\boldsymbol{\varepsilon}\bar{\boldsymbol{A}}\boldsymbol{\varepsilon}'/T\equiv\boldsymbol{\eta}\mathbf{\Gamma}^{1/2}\bar{\boldsymbol{A}}\mathbf{\Gamma}^{1/2}\boldsymbol{\eta}'/T$. In Section S3 of SM, we prove the following lemma.

Lemma 2. *For each $u>0$, $\bar{F}_{n,T}$ a.s. weakly converges to a distribution F_u with the Stieltjes transform $m_u\equiv m_u(z)$ equal to the unique solution in $\mathbb{C}^+$ of the equation*

$$z=-\frac{1}{m_u}+\int_0^\infty\frac{x\mathrm{d}H_u(x)}{1+cxm_u},$$

as $n,T\to_c\infty$.

Since, as follows from Lemma 2 and (S2.7) of SM, $\bar{F}_{n,T}$ a.s. weakly converge to $\mathbf{F}$ as $T,u\to\infty$, we must have

$$\mathcal{L}\left(F_u,\mathbf{F}\right)=o(1)\quad\text{as}\quad u\to\infty.\tag{A.3}$$

By the triangle inequality,

$$\mathcal{L}\left(F_{n,T},\mathbf{F}\right)\leq\mathcal{L}\left(F_{n,T},\bar{F}_{n,T}\right)+\mathcal{L}\left(\bar{F}_{n,T},F_u\right)+\mathcal{L}\left(F_u,\mathbf{F}\right).\tag{A.4}$$

By the rank inequality, $\mathcal{L}\left(F_{n,T},\bar{F}_{n,T}\right)\leq 1/\left(2\sqrt{u}\right)$. This fact, Lemma 2 and (A.3) imply Theorem 1.

A.2 Proof of Theorem 2

Consider a sequence x_j, $j=1,2,\ldots$ with $x_j\in(x_0,\infty)$ and $x_j\to\infty$ as $j\to\infty$. Let m_j be the corresponding sequence of the solutions from $\mathbb{C}^+$ of equation (2.4). Let

$$b=\max_{\omega\in[0,2\pi]}\left(1-\cos\omega\right)\left(\pi f\left(\omega\right)\right)^{-1}<\infty.$$

For any $\delta>0$ and all sufficiently large j, we must have $\Re\left(m_j\right)\in[-\delta-b,\delta]$ and $\Im\left(m_j\right)\in[0,\delta]$. Otherwise, there would exist a subsequence along which the right hand side of equation (2.4) remains bounded whereas the left hand side diverges to infinity. Since $[-\delta-b,\delta]\times[0,\delta]$ is compact, there must exist a convergent subsequence, which we will also denote as m_j, slightly abusing notation. Note that $\lim_{j\to\infty}\Im\left(m_j\right)=0$. Let us denote $\lim_{j\to\infty}\Re\left(m_j\right)$ as $\bar{m}$. We will now show that $\bar{m}=0$.

First, since δ is an arbitrary positive number, $\bar{m} \leq 0$. Suppose that $\bar{m} < 0$. The real and imaginary parts of equation (2.4) yield

$$x_j = \frac{1}{\pi} \int_0^\pi \frac{(1-\cos\omega)(\pi f(\omega))^{-1} \mathrm{d}\omega}{\left|(1-\cos\omega)(\pi f(\omega))^{-1} + cm_j\right|^2}, \tag{A.5}$$

$$0 = \frac{1}{|m_j|^2} - \frac{c}{\pi} \int_0^\pi \frac{\mathrm{d}\omega}{\left|(1-\cos\omega)(\pi f(\omega))^{-1} + cm_j\right|^2}. \tag{A.6}$$

Since $|(1-\cos\omega)(\pi f(\omega))^{-1}| \leq b$, we must have

$$x_j \leq \frac{b}{\pi} \int_0^\pi \frac{\mathrm{d}\omega}{\left|(1-\cos\omega)(\pi f(\omega))^{-1} + cm_j\right|^2} = \frac{b}{c|m_j|^2} \to \frac{b}{c\bar{m}^2} < \infty$$

which contradicts the fact that $x_j \to \infty$. Hence, $\lim_{j\to\infty} m_j = 0$.

Let $\bar{\omega} \in (0,\pi)$ be such that $f'(\omega)$ exists on $[0,\bar{\omega}]$, and

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{1-\cos\omega}{f(\omega)}\right) = \frac{\sin\omega\,(f(\omega) - \tan(\omega/2) f'(\omega))}{f^2(\omega)} \neq 0$$

for any $\omega \in (0,\bar{\omega}]$. The existence of such $\bar{\omega}$ follows from Assumption 3. Let us split the domain of the integration in the integral in (2.4) so that

$$\int_0^\pi \frac{\mathrm{d}\omega}{(1-\cos\omega)(\pi f(\omega))^{-1} + cm_j} = \int_0^{\bar{\omega}} \cdots + \int_{\bar{\omega}}^\pi \cdots \equiv I_1 + I_2.$$

Since $(1-\cos\omega)(\pi f(\omega))^{-1}$ is bounded away from zero on $[\bar{\omega},\pi]$, we have

$$|I_2| = O(1) \tag{A.7}$$

as $j \to \infty$.

Let $t = (1-\cos\omega)/(c\pi f(\omega))$. Then

$$I_1 = \frac{1}{c} \int_0^q \frac{\left(\phi(t)/\sqrt{t}\right) \mathrm{d}t}{t + m_j},$$

where $q = (1-\cos\bar{\omega})/(c\pi f(\bar{\omega}))$, $\phi(t)$ is continuously differentiable on $[0,q]$, and $\phi(0) = \sqrt{c\pi f(0)/2}$. Such integrals are well studied in mathematical literature and are called the integrals of Cauchy type. By Proposition II on p.75 of

Muskhelishvili (1968),

$$I_1 = \frac{\pi\phi(0)}{cm_j^{1/2}} + I_{10} = \frac{\pi\sqrt{c\pi f(0)/2}}{cm_j^{1/2}} + I_{10}, \tag{A.8}$$

where I_{10} is a holomorphic function of m_j in a neighborhood of 0, cut along the positive real semi-axis. Moreover,

$$|I_{10}| < \frac{C}{|m_j|^{\alpha}} \tag{A.9}$$

for some $\alpha < 1/2$ and $C > 0$.

Using (A.7), (A.8), and (A.9) in (2.4) yields

$$x_j = -\frac{1}{m_j} + \frac{\sqrt{c\pi f(0)/2}}{cm_j^{1/2}} + O(1) + O\left(|m_j|^{-\alpha}\right), \tag{A.10}$$

where $\alpha < 1/2$. Let $m_j = r_j \exp\{\mathrm{i}\varphi_j\}$ with $\varphi_j \in (0, \pi)$. The real and the imaginary parts of the equation (A.10) are

$$x_j = -r_j^{-1}\cos\varphi_j + \sqrt{\frac{\pi f(0)}{2c}} r_j^{-1/2}\cos\left(\frac{\varphi_j}{2}\right) + O(1) + O\left(r_j^{-\alpha}\right), \tag{A.11}$$

$$0 = r_j^{-1}\sin\varphi_j - \sqrt{\frac{\pi f(0)}{2c}} r_j^{-1/2}\sin\left(\frac{\varphi_j}{2}\right) + O(1) + O\left(r_j^{-\alpha}\right). \tag{A.12}$$

Since $r_j \to 0$ as $j \to \infty$, (A.12) implies that

$$\min\{\varphi_j, \pi - \varphi_j\} \to 0.$$

On the other hand, equation (A.11) shows that φ_j cannot be close to zero asymptotically, because then, x_j and $-r_j^{-1}\cos\varphi_j$ would be of different sign. Hence,

$$\lim_{j\to\infty}\varphi_j = \pi, \text{ and } r_j = \frac{1}{x_j} + o\left(\frac{1}{x_j}\right).$$

Using this in (A.12), we obtain

$$2x_j\cos\left(\frac{\varphi_j}{2}\right) - \sqrt{\frac{\pi f(0)}{2c}} x_j^{1/2} = o\left(x_j\cos\left(\frac{\varphi_j}{2}\right)\right) + o\left(x_j^{1/2}\right),$$

or, equivalently,

$$\cos\left(\frac{\varphi_j}{2}\right) = \frac{1}{4}\sqrt{\frac{2\pi f(0)}{c}} x_j^{-1/2} + o\left(\cos\left(\frac{\varphi_j}{2}\right)\right) + o\left(x_j^{-1/2}\right),$$

so that

$$\cos\left(\frac{\varphi_j}{2}\right) = \frac{1}{4}\sqrt{\frac{2\pi f(0)}{c}} x_j^{-1/2}\left(1+o(1)\right)$$

and

$$\Im\left(m_j\right) \equiv r_j \sin\varphi_j = \frac{1}{2}\sqrt{\frac{2\pi f(0)}{c}} x_j^{-3/2}\left(1+o(1)\right).$$

This yields Theorem 2.

A.3 Proof of Corollary 1

The spectral density of η_t equals $1/\left(2\pi\right)$. Therefore, by Theorem 1, $F^{\boldsymbol{S}_{n,T}}$ a.s. weakly converges to a distribution function $\mathbf{F}(x)$ whose Stieltjes transform $\mathbf{m} \equiv \mathbf{m}(z)$ satisfies

$$\begin{aligned} z &= -\frac{1}{\mathbf{m}} + \frac{1}{2\pi}\int_0^{2\pi} \frac{\mathrm{d}\omega}{2\left(1-\cos\omega\right) + c\mathbf{m}} \\ &= -\frac{1}{\mathbf{m}} + \frac{1}{2\pi i}\oint_{|s|=1} \frac{\mathrm{d}s}{\left[2\left(1-\left(s+s^{-1}\right)/2\right) + c\mathbf{m}\right] s} \\ &= -\frac{1}{\mathbf{m}} - \frac{1}{2\pi i}\oint_{|s|=1} \frac{\mathrm{d}s}{s^2 - \left(c\mathbf{m}+2\right)s + 1}. \end{aligned}$$

The integrand has two poles at $s_{1,2} = (c\mathbf{m}+2 \pm \sqrt{\left(c\mathbf{m}+2\right)^2 - 4})/2$. As $s_1 s_2 = 1$, we must have one of them inside the contour and the other outside. Therefore, we have

$$z = -\frac{1}{\mathbf{m}} \pm \frac{1}{s_1 - s_2} = -\frac{1}{\mathbf{m}} \pm \frac{1}{\sqrt{\left(c\mathbf{m}+2\right)^2 - 4}},$$

with the choice of $+$ or $-$ determined by which of $s_{1,2}$ is inside the contour. Rearranging, we obtain

$$c\left(z + \frac{1}{\mathbf{m}}\right)^2 \left(c\mathbf{m}+4\right)\mathbf{m} - 1 = 0.$$

On the other hand, we have

$$m_c(z) = \int \frac{1}{x-z}\mathrm{d}F_c\left(x\right) = \int \frac{1}{cx - z}\mathrm{d}\mathbf{F}\left(x\right) = \frac{1}{c}\mathbf{m}\left(\frac{z}{c}\right).$$

Substituting this into previous equation completes the proof.

A.4 Proof of Theorem 3

Let $F_{n,\infty}$ be the ESD of $\boldsymbol{\mathcal{W}}$. It suffices to show that for any $\delta > 0$, there exists some n_0 such that for any $n > n_0$,

$$\mathbb{P}\left(\mathcal{L}\left(F_0, F_{n,\infty}\right) < \delta\right) > 1 - \delta, \tag{A.13}$$

where $\mathcal{L}\left(\cdot,\cdot\right)$ is the Lévy distance. The idea is to introduce "intermediate" distributions F_γ, F_{n,T_γ}, and F_{n,T_∞} such that the Lévy distances $\mathcal{L}\left(F_0, F_\gamma\right)$, $\mathcal{L}(F_\gamma, F_{n,T_\gamma})$, $\mathcal{L}\left(F_{n,T_\gamma}, F_{n,T_\infty}\right)$, and $\mathcal{L}\left(F_{n,T_\infty}, F_{n,\infty}\right)$ are small, and then use the triangle inequality to establish (A.13).

Let F_γ be the limiting spectral distribution of $(n/T)\,\boldsymbol{S}_{n,T}$ as $p, T \to_\gamma \infty$. The Stieltjes transforms of F_γ and of F_0 satisfy the cubic equation (3.1) with c set to γ and to 0, respectively. Note that F_c continuously varies with c. Therefore, we can choose $\gamma \in (0,1)$ so small such that

$$\mathcal{L}\left(F_0, F_\gamma\right) < \frac{\delta}{4}. \tag{A.14}$$

Next, let T_γ be the smallest integer satisfying $n/T_\gamma \leq \gamma$, let $\boldsymbol{\eta}_\gamma$ be an $n \times T_\gamma$ matrix with i.i.d. standard normal entries, and let $\boldsymbol{U}_\gamma$ be the T_γ-dimensional upper triangular matrix of ones. Note that

$$\boldsymbol{S}_{n,T_\gamma} = \frac{\boldsymbol{\eta}_\gamma \boldsymbol{U}_\gamma \boldsymbol{U}_\gamma' \boldsymbol{\eta}_\gamma'}{T_\gamma},$$

and that

$$\left(\boldsymbol{U}_\gamma \boldsymbol{U}_\gamma'\right)^{-1} = \begin{pmatrix} 1 & -1 & & \cdots & & \\ -1 & 2 & -1 & \cdots & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix}.$$

This is a rank one perturbation of

$$\boldsymbol{A}_\gamma^{-1} = \begin{pmatrix} 2 & -1 & & \cdots & & \\ -1 & 2 & -1 & \cdots & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix},$$

which has eigenvalues

$$\lambda_k = 2 - 2\cos\frac{k\pi}{T_\gamma + 1}, \quad k = 1, \ldots, T_\gamma.$$

Let $\boldsymbol{A}_\gamma = \boldsymbol{Q}\boldsymbol{\Delta}_\gamma\boldsymbol{Q}'$ be a spectral decomposition of $\boldsymbol{A}_\gamma$. We have

$$\boldsymbol{\Delta}_\gamma = \mathrm{diag}\left(\left(2 - 2\cos\frac{k\pi}{T_\gamma + 1}\right)^{-1}\right)_{k=1,\ldots,T_\gamma}.$$

By the invariance of the Gaussianity under orthogonal transformations, matrix $\boldsymbol{\xi}_\gamma \equiv \boldsymbol{\eta}_\gamma\boldsymbol{Q}$ has i.i.d. $N(0,1)$ entries. Therefore, up to a low-rank perturbation (which can be ignored for the purpose of the analysis of the LSD), $\boldsymbol{S}_{n,T_\gamma}$ is distributed as $\boldsymbol{\xi}_\gamma\boldsymbol{\Delta}_\gamma\boldsymbol{\xi}_\gamma'/T_\gamma$.

Define $\boldsymbol{M}_{n,T_\gamma} = n\boldsymbol{\xi}_\gamma\boldsymbol{\Delta}_\gamma\boldsymbol{\xi}_\gamma'/(T_\gamma+1)^2$ and denote the empirical spectral distribution of $\boldsymbol{M}_{n,T_\gamma}$ as F_{n,T_γ}. Then, by Corollary 1, we have F_{n,T_γ} a.s. weakly converges to F_γ when $n, T \to_\gamma \infty$. Hence, almost surely, $\mathcal{L}\left(F_\gamma, F_{n,T_\gamma}\right) \to 0$ as $n \to \infty$ while $\gamma \in (0,1)$ is kept fixed. Therefore, there exists n_γ such that for any $n > n_\gamma$,

$$\mathbb{P}\left(\mathcal{L}\left(F_\gamma, F_{n,T_\gamma}\right) < \frac{\delta}{4}\right) > 1 - \frac{\delta}{4}. \tag{A.15}$$

Next, let $T_\infty > T_\gamma$. For any fixed n, we have $\mathcal{L}\left(F_{n,T_\infty}, F_{n,\infty}\right) \to 0$ in probability as $T_\infty \to \infty$. It is because $\boldsymbol{M}_{n,T_\infty}$ converges in distribution to $\boldsymbol{\mathcal{W}}$ as $T_\infty \to \infty$. Combining this with (A.14) and (A.15) yields: for any $\delta > 0$, there exists a $\gamma_\delta \in (0,1)$ s.t. for any positive $\gamma < \gamma_\delta$ there is an n_γ s.t. for any $n > n_\gamma$ there is a T_p s.t. for any $T_\infty > T_p$, one has

$$\mathbb{P}\left(\mathcal{L}\left(F_0, F_\gamma\right) + \mathcal{L}\left(F_\gamma, F_{n,T_\gamma}\right) + \mathcal{L}\left(F_{n,T_\infty}, F_{n,\infty}\right) < \frac{3\delta}{4}\right) > 1 - \frac{\delta}{2}.$$

It only remains to show that, for any $\delta > 0$, there exists a $\tilde{\gamma}_\delta \in (0,1)$ s.t. for any positive $\gamma < \tilde{\gamma}_\delta$ there is a $\tilde{n}_\gamma$ s.t. for any $n > \tilde{n}_\gamma$ and any $\tilde{T}_n$ there exists $T_\infty > \tilde{T}_n$ s.t.

$$\mathbb{P}\left(\mathcal{L}\left(F_{n,T_\gamma}, F_{n,T_\infty}\right) < \frac{\delta}{4}\right) > 1 - \frac{\delta}{2}.$$

The following lemma is proven in Section S4 of SM.

Lemma 3. *For any $\tau > 0$ there exists $\gamma_\tau \in (0,1)$ s.t. for any positive $\gamma < \gamma_\tau$, there is a $\tilde{n}_\gamma$ s.t. for any $n > \tilde{n}_\gamma$ and any $\tilde{T}_n$, there exists $T_\infty > \tilde{T}_n$ s.t. with* *probability larger than* $1-\tau$,

$$\|\boldsymbol{M}_{n,T_\gamma} - \boldsymbol{M}_{n,T_\infty}\| \le K\gamma,$$

where K *is an absolute constant that does not depend on* γ.

By Theorem A.45 of Bai and Silverstein (2010),

$$\mathcal{L}\left(F_{n,T_\gamma}, F_{n,T_\infty}\right) \le \|\boldsymbol{M}_{n,T_\gamma} - \boldsymbol{M}_{n,T_\infty}\|.$$

Therefore, given Lemma 3, for any $\tau > 0$ there exists $\gamma_\tau > 0$ s.t. for any positive $\gamma < \gamma_\tau$, there is a $\tilde{n}_\gamma$ s.t. for any $n > \tilde{n}_\gamma$ and any $\tilde{T}_n$, there exists $T_\infty > \tilde{T}_n$ s.t.

$$\mathbb{P}\left(\mathcal{L}\left(F_{n,T_\gamma}, F_{n,T_\infty}\right) < K\gamma\right) > 1-\tau.$$

Finally, the proof of Theorem 3 is completed with $\tau = \delta/2$ and $\tilde{\gamma_\delta} = \min\{\gamma_\tau, \delta/(4K)\}$.

References

Arnold, B. C. (2015). *Pareto Distributions.* 2nd Edition. Chapman & Hall/CRC, Boca Raton.

Bai, J. and Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica* **72**, 1127–1177.

Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Matrices.* 2nd Edition. Springer.

Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica* **18**, 425–442.

Banerjee, A., Marcellino, M. and Masten, I. (2017). Structural FECM: Cointegration in large-scale structural FAVAR models. *Journal of Applied Econometrics* **32**, 1069–1086.

Barigozzi, M., Lippi, M. and Luciani, M. (2018). Non-stationary dynamic factor models for large datasets. *arXiv 1602.02398v3.*

Burda, Z., Jarosz, A., Nowak, M. A. and Snarska, M. (2010). A random matrix approach to VARMA processes. *New Journal of Physics* **12**, 075036.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research* **1**, 245–276.

Doob, J. L. (1953). *Stochastic Processes.* John Wiley & Sons, Inc.

Engel, C., Mark, N. C. and West, K. D. (2015). Factor model forecasts of exchange rates. *Econometric Reviews* **34**, 32–55.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116**, 1–22.

Jin, B., Wang, C., Miao, B. and Lo Huang, M. N. (2009). Limiting spectral distribution of large-dimensional sample cavaraince matrices generated by VARMA. *Journal of Multivariate Analysis* **100**, 2112–2125.

Johnstone, I. and Paul, D. (2018). PCA in high dimensions: An orientation. In *Proceedings of the IEEE* **106**, 1277–1292.

Luo, L., Han, B., Chen, J., Sheng, G. and Jiang, X. (2017). Partial discharge detection and recognition in random matrix theory paradigm. *IEEE Acess* **5**, 8205–8213.

Melkiev, D., Guo, D. and Wirjanto, T. S. (2018). Applications of random-matrix theory and nonparametric change-point analysis to three notable systemic crises. *Quantitative Finance and Economics* **2**, 413–467.

Merlevède, F. and Peligrad, M. (2016). On the empirical spectral distribution for matrices with long memory and independent rows. *Stochastic Processes and their Applications* **126**, 2734–2760.

Muskhelishvili, N. I. (1968). *Singular Integral Equations, Boundary-Value Problems in the Theory of Functions, and Some of Their Applications to Mathematical Physics*. 3rd Edition. Nauka.

Onatski, A. and Wang, C. (2021). Spurious Factor Analysis. *Econometrica* **89**, 591–614.

Pfaffel, O. and Schlemm, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes. *Probability and Mathematical Statistics* **31**, 313–329.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics* **20**, 971–1001.

Silverstein, J. W. and Choi, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis* **54**, 295–309.

Yao, J. (2012). A note on a Marchenko-Pastur type theorem for time series. *Statistics and Probability Letters* **82**, 22–28.

Zhang, B., Pan, G. and Gao, J. (2018). CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *The Annals of Statistics* **46**, 2186–2215.

Alexei Onatski

Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, UK.

E-mail: ao319@cam.ac.uk

Chen Wang

Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam, Hong Kong.

E-mail: stacw@hku.hk

(Received February 2019; accepted May 2020)