Ann Inst Stat Math (2015) 67:359–373
DOI 10.1007/s10463-014-0452-2

On generalized expectation-based estimation of a population spectral distribution from high-dimensional data

Weiming Li · Jianfeng Yao

Received: 24 January 2013 / Revised: 9 January 2014 / Published online: 12 April 2014

Abstract This paper discusses the problem of estimating the population spectral distribution from high-dimensional data. We present a general estimation procedure that covers situations where the moments of this distribution fail to identify the model parameters. The main idea is to use generalized functional expectations as a substitute for the moments. Beyond the consistency, we also prove a central limit theorem for the proposed estimator. Simulation experiments illustrate the implementation of the estimation procedure. An application to the analysis of the eigenvalues of the sample correlation matrix of S&P 500 daily stock returns is proposed.

Keywords Large sample covariance matrix · Eigenvalues distribution · Population spectral distribution · Empirical spectral distribution · Generalized expectation estimation

1 Introduction

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a sequence of i.i.d. zero-mean random vectors in $\mathbb{R}^p$ or $\mathbb{C}^p$, with a common population covariance matrix Σ_p. When the population size p is not negligible with respect to the sample size n, modern random matrix theory indicates that the sample covariance matrix $S_n = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^* / n$ does not approach Σ_p. Therefore,

J. Yao's research is partially supported by RGC grant HKU 705413P.

W. Li
School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: liwm601@gmail.com

J. Yao (✉)
Department of Statistics and Actuarial Science, The University of Hong Kong, Hongkong, China
e-mail: jeffyao@hku.hk

classical statistical procedures based on an approximation of Σ_p by S_n become inconsistent in such high-dimensional data situations.

The spectral distribution (SD) F^A of an $m \times m$ Hermitian (or real symmetric) matrix A is the measure generated by its eigenvalues $\{\lambda_i^A\}$,

$$F^A = \frac{1}{m} \sum_{i=1}^{m} \delta_{\lambda_i^A} ,$$

where δ_b denotes the Dirac point measure at b. Let $(\sigma_i)_{1 \leq i \leq p}$ be the p eigenvalues of the population covariance matrix Σ_p. We are particularly interested in the SD

$$H_p := F^{\Sigma_p} = \frac{1}{p} \sum_{i=1}^{p} \delta_{\sigma_i}.$$

This SD or its limit H (see below) is referred as the population spectral distribution (PSD) of the observation model.

The main observation is that for high-dimensional data, the observed SD $F_n := F^{S_n}$ of the sample covariance matrix is far from the PSD H_p. Indeed, under reasonable assumptions, when both dimensions p and n grow proportionally, almost surely, the empirical SD F_n weakly converges to a deterministic distribution F, called limiting spectral distribution (LSD), which in general has no explicit form but is expressed via an implicit equation (Marčenko and Pastur, 1967; Silverstein, 1995; Silverstein and Bai, 1995).

A natural question here is the recovery of the PSD H_p (or its limit H) from the sample covariance matrix S_n. This question has a central importance in such statistical methodologies as principal component analysis (Johnstone, 2001) and factor analysis that rely on an efficient estimation of some population covariance matrices.

Mestre (2008) introduces a method based on contour integration under an eigenvalue splitting condition. The estimation method has been employed in a so-called "information plus noise" model in Hachem et al. (2012). Most recently, Li and Yao (2013) has provided an extension of Mestre's method to situations where the eigenvalue splitting condition cannot be met. A consistent estimator of the PSD H is derived by solving a system of approximated moment equations. An interesting finding from this work is that when sample eigenvalues form a unique cluster, the generalized estimator is equivalent to a homogeneous estimator in Yao et al. (2012) and a full moment estimator in Bai et al. (2010). Some related references include El Karoui (2008), Rao et al. (2008), Chen et al. (2011), and Li et al. (2013).

However, except El Karoui (2008) and Li et al. (2013), all the cited estimation methods are based on the moments of the PSD H. It may happen and that has been a surprise, that these moments can not help to identify model parameters. Such an example is provided in Sect. 5 with the sample correlation matrix of stock returns, for which the underlying PSD H has a normalized unit mean and infinite variance whatever the values of the model parameter. Clearly, any estimation procedure based on the moments of H fails in such situations.

The main motivation of this work is to propose a new estimator to cover these intriguing situations. Inspired by the generalized method of moments, we consider empirical statistics linked to a class of general test functions. These test functions are usually smaller than the monomials x^j and thus expected to have a finite expectation with respect to the unknown PSD H. In the example of stock returns data, H has a infinite variance but test functions like $\sin(x)$ do have a finite integral with respect to H, which makes its estimation possible.

The rest of the paper is organized as follows. In the next section, we put forward a general estimator of a PSD H based on its functional expectations in a parametric setup. Asymptotic properties of the proposed estimator are discussed in Sect. 3, including consistency and asymptotic normality. In Sect. 4, a specific parametric model is investigated through simulation experiments. In the next section, our method is applied to analyze a correlation matrix of stock returns. Proofs of main theorems are collected in the last section.

2 Generalized expectation estimation

Let G be a measure on the real line, the support of G is denoted by S_G. The Stieltjes transform of G is

$$s_G(z) = \int \frac{1}{x-z} dG(x), \quad z \in \mathbb{C}^+,$$

which is a one-to-one map defined on the upper half complex plane $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im(z) > 0\}$. The transform can be trivially extended to $\mathbb{C} \backslash S_G$ by using the same functional form, which will be adopted throughout the paper.

Suppose that the underlying PSD H belongs to a parametric family:

$$\mathcal{H} = \{H(\theta) : \theta \in \Theta \subset \mathbb{R}^q\}.$$

Denote by c the limiting ratio of p/n, and F the LSD with respect to H and c. Let f be an analytic function on an open region containing the support S_F of F, and $H(f)$ be the expectation of f with respect to H, i.e.

$$H(f) = \int f(t) \mathrm{d}H(t).$$

We call this integral generalized expectation of the PSD H. It will be shown that $H(f)$ connects to F through the Stieltjes transform $\underline{s}(z)$ of $cF + (1-c)\delta_0$ by a contour integral:

$$H(f) = K(c, f) + \frac{1}{2\pi \mathrm{i} c} \oint_C z\underline{s}'(z) f(-1/\underline{s}(z)) \mathrm{d}z, \tag{1}$$

where $\underline{s}'(z)$ stands for the derivative of $\underline{s}(z)$, $K(c, f)$ is a constant related to c and f, and C is a positive oriented contour enclosing the support S_F (see Theorem 1). The analyticity assumption on f is necessary for this formula, since it is obtained by calculating the contour integral using the Cauchy integral theorem. When an empirical SD F_n is obtained, we may use the Stieltjes transform $\underline{s}_n(z)$ of $(p/n)F_n + (1-p/n)\delta_0$

and its derivative $\underline{s}_n'(z)$ to estimate $\underline{s}(z)$ and $\underline{s}'(z)$, respectively, in the formula (1), and then get an estimate

$$\widehat{H}(f) := K(p/n, f) + \frac{n}{p}\frac{1}{2\pi \mathrm{i}}\oint_C z\underline{s}_n'(z) f(-1/\underline{s}_n(z))\mathrm{d}z. \tag{2}$$

Now with the help of $H(f)$ and its estimate $\widehat{H}(f)$, we consider the estimation of the PSD H. Let $f_1, \ldots, f_q$ be analytic functions on an open region containing S_F, $\boldsymbol{\gamma} = (H(f_j))_{1\le j\le q}$ be a q dimensional vector of generalized expectations. In order to make θ identifiable from $\boldsymbol{\gamma}$, we assume that the vector function g from $\mathbb{R}^q$ to $\mathbb{R}^q$: $\theta \mapsto \boldsymbol{\gamma}$ is invertible in Θ. Under this assumption, the *generalized expectation estimator* (GEE) of θ is

$$\widehat{\theta}_n = g^{-1}(\widehat{\boldsymbol{\gamma}}_n),$$

where $\widehat{\boldsymbol{\gamma}}_n = (\widehat{H}(f_j))_{1\le j\le q}$ with the elements defined in (2).

There are several closely related estimators in the literature. In Mestre (2008), the author discussed a simple case where $f(z) = z$. In this case, $H(f)$ and its estimator become

$$H(f) = -\frac{1}{2\pi \mathrm{i}c}\oint_C z\underline{s}'(z)/\underline{s}(z)\mathrm{d}z, \quad \widehat{H}(f) = -\frac{n}{p}\frac{1}{2\pi \mathrm{i}}\oint_C z\underline{s}_n'(z)/\underline{s}_n(z)\mathrm{d}z,$$

respectively, where the second contour integral can be figured out by residue theorem. In the special case with monomials $f_j(z) = z^j$ $(j = 0, 1, \ldots q)$ and H is discrete with a finite support, the GEE has been discussed in Bai et al. (2010) and Yao et al. (2012), and has been extended by a localization method in Li and Yao (2013).

The generalization from monomials to general analytic functions proposed in this paper has important significance, in that it provides us a much wider class of statistics useful to the inference about H. A real data analysis presented in Sect. 5 is built on this generalization.

3 Asymptotic properties

In this section, we study the asymptotic properties of the expectations $\{\widehat{H}(f_j)\}$ and the GEE $\widehat{\theta}_n$. All these properties are based on the following assumptions.

Assumption (a). The sample and population sizes n, p both tend to infinity, and in such a way that $p/n \to c \in (0, \infty)$.

Assumption (b). There is a doubly infinite array of i.i.d. complex-valued random variables (w_{ij}), $i, j \ge 1$ satisfying

$$\mathbb{E}(w_{11}) = 0, \quad \mathbb{E}(|w_{11}|^2) = 1, \ \mathbb{E}(|w_{11}|^4) < \infty,$$

such that for each p, n, letting $W_n = (w_{ij})_{1\le i\le p, 1\le j\le n}$, the observation vectors can be represented as $\mathbf{x}_j = \Sigma_p^{1/2} w_{.j}$ where $w_{.j} = (w_{ij})_{1\le i\le p}$ denotes the j-th column of W_n.

Assumption (c). The PSD H_p of Σ_p weakly converges to a probability distribution H on $[0, \infty)$ as $n \to \infty$. Moreover, the sequence of spectral norms $(||\Sigma_p||)$ is bounded in p.

Assumptions (a)–(c) are classical conditions for the central limit theorem (CLT) of linear spectral statistics, see Bai and Silverstein (2004, 2010).

Theorem 1 *Under the assumptions* (a)–(c), *for each j $(1 \le j \le q)$,*

(i) *the generalized expectation $H(f_j)$ can be re-expressed as*

$$H(f_j) = K(c, f_j) + \frac{1}{2\pi \mathrm{i} c} \oint_C z\underline{s}'(z) f_j(-1/\underline{s}(z)) \mathrm{d}z,$$

where C is a positively oriented contour, taking values in $\mathbb{C}\backslash(S_F \cup \{0\})$ and enclosing the support S_F of F, and $K(c, f_j) = (1 - 1/c) f_j(0)$ if C enclosing 0, and zero otherwise;

(ii) *the empirical expectation $\widehat{H}(f_j)$ based on $\underline{s}_n(z)$ converges almost surely to $H(f_j)$.*

Theorem 2 *Under the assumptions* (a)–(c),

(i) *the random vector*

$$n\left(\widehat{H}(f_j) - H_p(f_j)\right)_{1 \le j \le q} \tag{3}$$

forms a tight sequence in n, where the centralization term $H_p(f_j)$ stands for the generalized expectation of H_p.

(ii) *If w_{11} and Σ_p are real and* $\mathrm{E}(w_{11}^4) = 3$, *then* (3) *converges weakly to a Gaussian distribution $N_q(\mu, \Phi)$, with mean vector*

$$\mu = \left(-\frac{1}{2\pi \mathrm{i}} \oint_C f_j(-1/\underline{s}(z)) \frac{\int t^2 \underline{s}'(z)^2 \mathrm{d}H(t)}{\underline{s}(z)(1 + \underline{s}(z))^3} \mathrm{d}z \right)_{1 \le j \le q}$$

and covariance matrix $\Phi = (\phi_{ij})_{q \times q}$ with

$$\phi_{ij} = \frac{-1}{4\pi^2 c^2} \oint_C \oint_{C'} f_i(-1/\underline{s}(z_1)) f_j(-1/\underline{s}(z_2)) k(z_1, z_2) \mathrm{d}z_1 \mathrm{d}z_2,$$

where $k(z_1, z_2) = 2\underline{s}'(z_1)\underline{s}'(z_2)/(\underline{s}(z_1) - \underline{s}(z_2))^2 - 2/(z_1 - z_2)^2$. The contours C and C' share the same properties and are assumed non-overlapping.

(iii) *If w_{11} is complex with* $\mathrm{E}(w_{11}^2) = 0$ *and* $\mathrm{E}(|w_{11}|^4) = 2$, *then* (ii) *also holds, except the mean vector is zero and the covariance matrix is $\Phi/2$.*

Proofs of Theorems 1 and 2 are presented in the last section.

Note that the centralization term in the above CLT conclusion is a quantity based on H_p, but not on its limit H, which is consistent with the main result in Bai and Silverstein (2004).

When applying the CLT, we may need to estimate the limiting mean vector (for real case) and the covariance matrix (for both real and complex cases). From the strong consistency of $\underline{s}_n(z)$, their natural estimators, denoted as $\widehat{\mu} = (\widehat{\mu}_j)_{1\le j\le q}$ and $\widehat{\Phi} = (\widehat{\phi}_{ij})_{q\times q}$, are respectively given by

$$\widehat{\mu}_j = -\frac{1}{2\pi \mathrm{i}}\oint_C f_j(-1/\underline{s}_n(z))\frac{\underline{s}_n'(z)^2\widehat{H}(f_0)}{\underline{s}_n(z)(1+\underline{s}_n(z))^3}\mathrm{d}z, \quad f_0(z) = z^2,$$
$$\widehat{\phi}_{ij} = \frac{-1}{4\pi^2c^2}\oint_C\oint_{C'} f_i(-1/\underline{s}_n(z_1))f_j(-1/\underline{s}_n(z_2))\widehat{k}(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2,$$

where $\widehat{k}(z_1,z_2) = 2\underline{s}_n'(z_1)\underline{s}_n'(z_2)/(\underline{s}_n(z_1)-\underline{s}_n(z_2))^2 - 2/(z_1-z_2)^2$. The consistency of the estimators follow immediately from the dominated convergence theorem.

Theorem 3 *In addition to the assumptions* (a)–(c)*, suppose that the true value of the parameter θ_0 is an inner point of Θ. Also, suppose that the function $g(\theta)$ is differentiable in a neighborhood of θ_0 and the Jacobian matrix $J(\theta) = \partial g/\partial\theta$ is invertible at θ_0. Then,*

(i) *the GEE $\widehat{\theta}_n$ is strongly consistent, i.e.*

$$\widehat{\theta}_n \to \theta_0, \quad a.s.,$$

(ii) *moreover, if the assumptions in* (ii) *or* (iii) *of Theorem* 2 *on w_{11} hold, then*

$$n(\widehat{\theta}_n - g^{-1}(\boldsymbol{\gamma}_p)) \xrightarrow{\mathcal{D}} N_q(J^{-1}(\theta_0)\mu(\theta_0), \Gamma(\theta_0)),$$

where $\boldsymbol{\gamma}_p = (H_p(f_j))_{1\le j\le q}$ and $\Gamma(\theta_0) = J^{-1}(\theta_0)\Phi(\theta_0)(J^{-1}(\theta_0))'$ with μ and Φ defined in Theorem 2*.*

This theorem follows from Theorems 1 and 2 by standard arguments: consistency of a method of moments type estimator for the first conclusion and application of the delta method for the second conclusion. Its proof is thus omitted.

4 A small Monte-Carlo study

In this section, we study a continuous PSD H which has a density of beta distribution, that is,

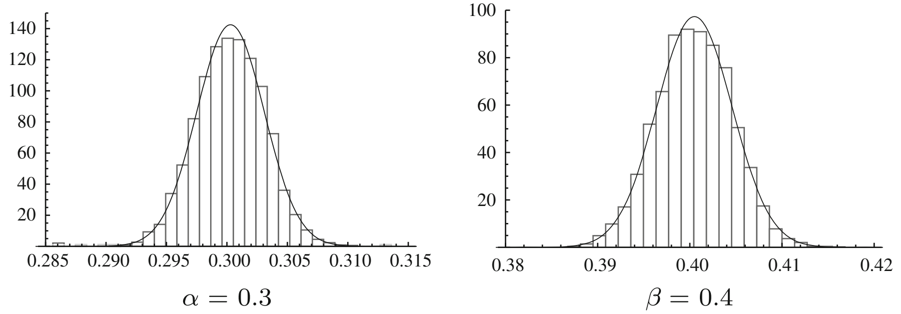
$$\begin{aligned} h(t|\theta) &= \frac{1}{B(\alpha,\beta)}t^{\alpha-1}(1-t)^{\beta-1}, \quad 0 < t < 1, \quad \theta\in\Theta, \\ &:= beta(t,\alpha,\beta), \end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function and $\Theta = \{(\alpha,\beta) : \alpha > 0, \ \beta > 0\}$.

By considering the relationship between the density and its shape parameters, we take $f_1(z) = z$ and $f_2(z) = z(1-z)$ for simplicity. Let $\boldsymbol{\gamma} = (H(f_1), H(f_2))$, we have

Table 1 Estimates for beta distribution with $\alpha = 0.3,\ \beta = 0.4$

Method	GEE		BCY		LSE	
Parameter	α	β	α	β	α	β
Mean	0.3003	0.4005	0.3002	0.4004	0.3002	0.4003
S.E.	0.0028	0.0041	0.0028	0.0041	0.0018	0.0028



$\alpha = 0.3$ $\beta = 0.4$

Fig. 1 The histogram estimates of α and β compared to the corresponding density curves of asymptotic normal distributions

$$g(\theta) = \left(\frac{\alpha}{\alpha + \beta}, \frac{\alpha\beta}{(\alpha + \beta)(1 + \alpha + \beta)} \right), \quad J(\theta) = \begin{pmatrix} \frac{\beta}{(\alpha+\beta)^2} & -\frac{\alpha}{(\alpha+\beta)^2} \\ \frac{\beta(\beta+\beta^2-\alpha^2)}{(\alpha+\beta)^2(1+\alpha+\beta)^2} & \frac{\alpha(\alpha+\alpha^2-\beta^2)}{(\alpha+\beta)^2(1+\alpha+\beta)^2} \end{pmatrix}.$$

It is easy to verify the invertibility of g and J, and hence the GEE is strongly consistent and asymptotically normal.

We numerically evaluate the performance of the GEE under this continuous model with $(\alpha, \beta) = (0.3, 0.4)$. Sample is drawn from standard complex normal distribution with the sample size $n = 1000$. The population eigenvalues are chosen as the $(p+1)$-quantiles of H with $p = 2000$. The independent replications are 5000.

For the purpose of comparison, we also calculate two other estimates, one is the moment estimator in Bai et al. (2010) (referred as BCY), and the other is the least-squares type estimator in Li et al. (2013) (referred as LSE). We also examine the CLT of the GEE by comparing the histograms of estimates with their theoretical asymptotic distributions. The results are collected in Table 1 and Fig. 1, respectively.

Results in Table 1 show that, with the chosen functions $f_1(z)$ and $f_2(z)$, the GEE and BCY are almost equivalent for the studied model, while the LSE is better from the view point of standard errors. All the three estimates are slightly, but systematically, biased as shown in the table, which is due to the fact $H_p \neq H$. The biases are inevitable for a finite p and will vanish when p approaches to infinity. Results in Figure 1 demonstrate that the histogram estimates match their asymptotic normal distributions very well.

5 Application to S&P 500 daily stocks data

We consider an empirical correlation matrix of daily returns from stocks listed in the Standard & Poor Index, and analyze the distribution of its eigenvalues. The time period is from September, 2007 to September 2011 covering 1001 trading days. As 12 stocks listed as by September 2011 do not have a complete history, they are removed from the analysis and in total 488 US stocks have been included. The total data matrix of the returns is then with data dimension $p = 488$ and sample size $n = 1000$. Next the 488×488 sample correlation matrix of these returns is computed and we obtain its 488 sample eigenvalues.

One may object that daily stock returns are commonly known to be uncorrelated but dependent in time, so that in a strict sense, our theoretical results where temporal independence has been assumed do not cover such situations. However, we will provide evidence below that the theory developed in this paper applies as well: indeed, the structure of the correlations between returns predicted by the theory matches very well the empirically observed one. Therefore, the present theory seems applicable to a wider class of high-dimensional data than the one assumed in the theoretical results (Theorems 1 and 3).

It is well known that for correlation matrices from stock returns or macro-economic time series, a few large eigenvalues detach from the bulk of the eigenvalues and they are termed as spikes, see Johnstone (2001). For the matrix under hands, we list below the eight largest eigenvalues and the eight smallest ones:

```
237.96, 17.763, 14.003, 8.7635, 5.2994, 4.8569, 4.3945, 3.5001,
   ....      ....        ....
0.0198, 0.0194, 0.0190, 0.0178, 0.0174, 0.0164, 0.0155, 0.0147.
```

Several spike eigenvalues are clearly presented here and the largest has a highly dominant value describing the general tendency over the time period of the US stock market (“market mode”). Moreover, analysis of these spike eigenvalues is definitely different from that of the bulk eigenvalues. Interested readers are referred to Johnstone (2001), Baik and Silverstein (2006), and Bai and Yao (2008) for theoretic backgrounds, and to Kritchman and Nadler (2008) and Passemier and Yao (2012) for recent advances on related inference theory.

We concentrate ourselves on the analysis of the bulk eigenvalues by removing the first 6 largest ones which are deemed as spike eigenvalues. The question we address here is: what is the structure of the eigenvalues at the population level that has led to these observed eigenvalues. To this end and following Bouchaud and Potters (2011) and Li et al. (2013), an inverse cubic density is assumed for the PSD H associated to the bulk eigenvalues, that is,

$$h(t|\alpha) = \frac{c}{(t-a)^3} I(t \geq \alpha), \quad 0 \leq \alpha < 1,$$

where $c = 2(1-\alpha)^2$ and $a = 2\alpha - 1$.

As already noticed, moment-based methods fail to estimate the parameter α in that the moments of $H(\alpha)$ can not identify the parameter: H has infinite variance and unit

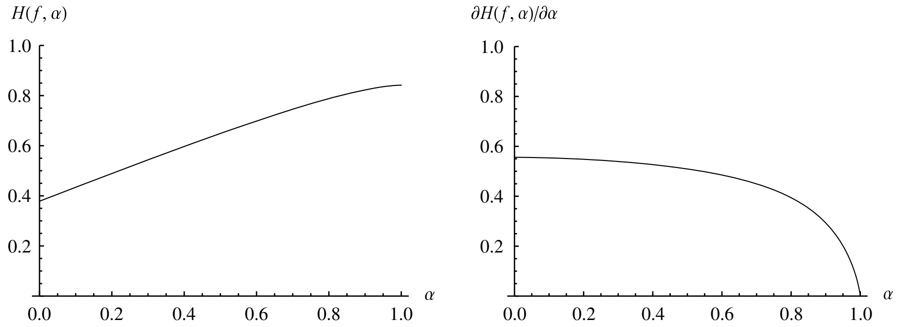


Fig. 2 Curves of $H(f,\alpha)$ (*left*) and $\partial H(f,\alpha)/\partial\alpha$ (*right*)

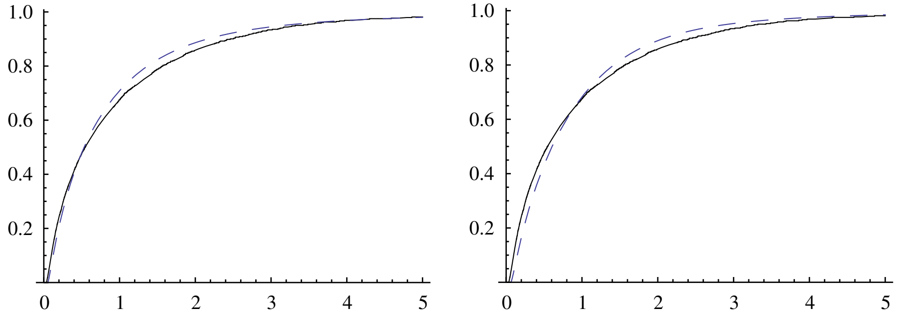


Fig. 3 The empirical distribution function of the sample eigenvalues (*plain black*) compared with the predicted LSD functions corresponding to $H(\widehat{\alpha})$ (*left*, *dashed blue*) and $H(\widehat{\alpha}')$ (*right*, *dashed blue*) (color figure online)

mean whatever the values of α. However, expectations of a suitably-chosen "test" function with respect to H can help to identify α.

Here, we provide an example with the test function $f(z) = \sin(z)$, that is, we consider the expectation

$$H(f,\alpha) = \int \sin(t)h(t|\alpha)\mathrm{d}t,$$

which exists and is increasing with respect to α (although $H(f,\alpha)$ has no analytic expression), see Fig. 2.

The estimate of the expectation is $\widehat{H}(f,\alpha) = 0.5546$ which indicates $\widehat{\alpha} = 0.3205$. Recall that the same data set has been analyzed in Li et al. (2013) and the LSE estimate is $\widehat{\alpha}' = 0.4380$. To assess these two estimates, we employ the Wasserstein distance $W = \int |Q_H(t) - Q_{\widehat{H}}(t)|\mathrm{d}t$ where $Q_\mu(t)$ is the quantile function of distribution μ. We calculate the distance between the ESD of the bulk eigenvalues and the predicted LSD derived from the estimate of $H(\alpha)$. It turns out that the distance is $d = 0.0824$ for $H(\widehat{\alpha})$ and is $d' = 0.1062$ for $H(\widehat{\alpha}')$. The ESD function and the predicted LSD functions are plotted in Fig. 3. The distances as well as the figure show that our method yields a better fit to the ESD.

Potential applications in the future of these findings can be done through an explicit factor modeling where factor scores and loadings, once estimated, will provide important information on the correlations, at the population level, between returns of the listed stocks.

6 Proofs

6.1 Lemmas

We present two lemmas where the conclusions will be used in the proof of our main theorem.

Lemma 1 *Under Assumptions* (a)–(c)*,*

(i) *the empirical spectral distribution F_n converges in distribution to a non-random distribution F, and the Stieltjes transform $\underline{s}(z)$ of $cF + (1-c)\delta_0$ satisfies the following equation:*

$$z = -\frac{1}{\underline{s}(z)} + c\int \frac{t}{1+t\underline{s}(z)}\mathrm{d}H(t)\ , \quad z \in \mathbb{C}^{+}, \tag{4}$$

(ii) *for any $z \in \mathbb{C}\backslash(S_F \cup \{0\})$ and sufficient large n, the Stieltjes transform $\underline{s}_n(z)$ of $(p/n)F_n + (1-p/n)\delta_0$ converges almost surely to $\underline{s}(z)$, which solves the equation* (4).

Proof The first conclusion is from Silverstein (1995) and the second is from Li et al. (2013). □

Let $\underline{s}_{c_n,H_p}(z)$ be the finite dimensional version of the Stieltjes transform $\underline{s}(z)$, which solves the following equation:

$$z = -\frac{1}{\underline{s}(z)} + c_n\int \frac{t}{1+t\underline{s}(z)}\mathrm{d}H_p(t), \quad z \in \mathbb{C}^{+},$$

with $c_n = p/n$. We are going to establish limiting results on

$$(Y_n(z), Z_n(z)) = n(\underline{s}_n(z) - \underline{s}_{c_n,H_p}(z), \underline{s}'_n(z) - \underline{s}'_{c_n,H_p}(z)),$$

when viewed as a random process defined on a contour $\mathcal{C}$ of the complex plane. The contour is described as follows. Let $v_0,\ x_l,\ x_r$ be any real numbers satisfying $v_0 > 0$, $x_l < \liminf_n \lambda_{\min}^{\Sigma_p} I_{(0,1)}(c)(1-\sqrt{c})^2$ and $x_l \neq 0$, and $x_r > \liminf_n \lambda_{\max}^{\Sigma_p}(1+\sqrt{c})^2$. Then

$$\mathcal{C} \equiv \{x_l + \mathrm{i}v : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + \mathrm{i}v : v \in [0, v_0]\},$$

where $\mathcal{C}_u = \{x + \mathrm{i}v_0 : x \in [x_l, x_r]\}$.

As $\underline{s}_n(z)$ and its derivative $\underline{s}'_n(z)$ may both converge to infinite when z is close to the real line. We introduce a truncated version $\{(\widehat{Y}_n(z), \widehat{Z}_n(z))\}$ of the original process following the same idea as in Bai and Silverstein (2004). More precisely, choose a sequence ε_n decreasing to zero satisfying for some $\alpha \in (0, 1)$

$$\varepsilon_n \geq n^{-\alpha}.$$

Let

$$\mathcal{C}_l = \begin{cases} \{x_l + \mathrm{i}v : v \in [n^{-1}\varepsilon_n, v_0]\}, & \text{if } x_l > 0; \\ \{x_l + \mathrm{i}v : v \in [0, v_0]\}, & \text{if } x_l < 0, \end{cases}$$

and

$$\mathcal{C}_r = \{x_r + \mathrm{i}v : v \in [n^{-1}\varepsilon_n, v_0]\}.$$

Write $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$, for $z = x + \mathrm{i}v$ we define

$$\widehat{Y}_n(z) = \begin{cases} Y_n(z), & \text{for } z \in \mathcal{C}_n; \\ Y_n(x + \mathrm{i}n^{-1}\varepsilon_n), & \text{for } z \in \mathcal{C} \backslash \mathcal{C}_n; \end{cases}$$

and

$$\widehat{Z}_n(z) = \begin{cases} Z_n(z), & \text{for } z \in \mathcal{C}_n; \\ Z_n(x + \mathrm{i}n^{-1}\varepsilon_n), & \text{for } z \in \mathcal{C} \backslash \mathcal{C}_n. \end{cases}$$

Obviously, $(Y_n(z), Z_n(z))$ agrees with $(\widehat{Y}_n(z), \widehat{Z}_n(z))$ on $\mathcal{C}_n$.

Lemma 2 *If the assumptions* (a)–(c) *hold, then:*

(i) *The process* $\{(\widehat{Y}_n(z), \widehat{Z}_n(z))\}$ *forms a tight sequence on* $\mathcal{C}$.
(ii) *If* w_{11} *and* Σ_p *are real and* $\mathrm{E}(w_{11}^4) = 3$, *then* $(\widehat{Y}_n(z), \widehat{Z}_n(z))$ *converges weakly to a Gaussian process* $(Y(z), Z(z))$, *with means*

$$\mathrm{E}Y(z) = \int \frac{ct^2 \underline{s}'(z)^2 \mathrm{d}H(t)}{\underline{s}(z)(1 + \underline{s}(z))^3}, \quad \mathrm{E}Z(z) = \frac{d}{\mathrm{d}z} \int \frac{ct^2 \underline{s}'(z)^2 \mathrm{d}H(t)}{\underline{s}(z)(1 + \underline{s}(z))^3}, \tag{5}$$

and covariance functions

$$\mathrm{Cov}(Y(z), Y(\tilde{z})) = \frac{2\underline{s}'(z)\underline{s}'(\tilde{z})}{(\underline{s}(z) - \underline{s}(\tilde{z}))^2} - \frac{2}{(z - \tilde{z})^2} := k(z, \tilde{z}), \tag{6}$$

$$\mathrm{Cov}(Y(z), Z(\tilde{z})) = \frac{\partial}{\partial \tilde{z}} k(z, \tilde{z}), \tag{7}$$

$$\mathrm{Cov}(Z(z), Z(\tilde{z})) = \frac{\partial^2}{\partial z \partial \tilde{z}} k(z, \tilde{z}), \tag{8}$$

where $\mathrm{Cov}(X, Y) \equiv \mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y)$.

(iii) *If* w_{11} *is complex with* $\mathrm{E}(w_{11}^2) = 0$ *and* $\mathrm{E}(|w_{11}|^4) = 2$*, then* (i) *also holds, except the means are zero and the covariance functions are* 1/2 *the function given in* (6)–(8).

Proof Firstly consider (i). According to Lemma 1.1 in Bai and Silverstein (2004), we know that the process $\{\widehat{Y}_n(z)\}$ forms a tight sequence on $\mathcal{C}$. Thus for any subsequence $\{\widehat{Y}_{n_k}(z)\}$, there exists a further sub-sequence $\{\widehat{Y}_{n_{k(j)}}(z)\}$ converging weakly to a limit, say $\widehat{Y}_{n_{k(0)}}(z)$, as $j \to \infty$. From the strong representation theorem (Skorohod, 1956; Dudley, 1985), there is a probability space on which we can define a sequence $\{\widetilde{Y}_{n_{k(j)}}(z)\}$ such that $\widetilde{Y}_{n_{k(j)}}(z)$ is identical in distribution to $\widehat{Y}_{n_{k(j)}}(z)$ for each $j = 0, 1, \ldots,$ and $\widetilde{Y}_{n_{k(j)}}(z)$ converges almost surely to $\widetilde{Y}_{n_{k(0)}}(z)$. Now using Vitali's convergence theorem (see Lemma 2.3 in Bai and Silverstein (2004)), we obtain that $\widetilde{Y}'_{n_{k(j)}}(z)$ converges almost surely to $\widetilde{Y}'_{n_{k(0)}}(z)$ for all $z \in \mathcal{C}$. Therefore, $(\widehat{Y}_{n_{k(j)}}(z), \widehat{Y}'_{n_{k(j)}}(z))$ converges weakly to $(\widehat{Y}_{n_{k(0)}}(z), \widehat{Y}'_{n_{k(0)}}(z))$, and thus $\{(\widehat{Y}_n(z), \widehat{Z}_n(z))\} = \{(\widehat{Y}_n(z), \widehat{Y}'_n(z))\}$ forms a tight sequence.

Considering (ii) and (iii), the convergence and the limiting covariance functions follow from the above arguments and Lemma 1.1 in Bai and Silverstein (2004), and thus we only need to calculate the mean function $\mathrm{E}Y(z)$ in (5).

In Bai and Silverstein (2004), it has been proved that

$$\mathrm{E}Y(z) = \frac{c\int \underline{s}^3(z)t^2(1+\underline{s}(z))^{-3}\mathrm{d}H(t)}{(1-c\int t^2\underline{s}^2(z)(1+t\underline{s}(z))^{-2}\mathrm{d}H(t))^2}. \tag{9}$$

On the other hand, taking the derivative of z on both sides of the equation (4), we have

$$\frac{\underline{s}^2(z)}{\underline{s}'(z)} = 1 - c\int \frac{t^2\underline{s}^2(z)}{(1+t\underline{s}(z))^2}\mathrm{d}H(t).$$

Substitute this to (9), we get

$$\mathrm{E}Y(z) = \int \frac{ct^2\underline{s}'(z)^2\mathrm{d}H(t)}{\underline{s}(z)(1+\underline{s}(z))^3}.$$

Then, the proof of the lemma is complete. □

Proof of Theorem 1 Let $u_n(z) = -1/\underline{s}_n(z)$, $u_{c_n,H_p}(z) = -1/\underline{s}_{c_n,H_p}(z)$, and $u(z) = -1/\underline{s}(z)$. Note that as $n \to \infty$, both $u_n(z)$ and $u_{c_n,H_p}(z)$ converge to $u(z)$ almost surely.

Write $D = \{u(z) : z \in C\}$. From the equation (4) and the fact that D encloses 0 if and only if C encloses 0, we have

$$\begin{aligned}\oint_D zf_j(u(z))\mathrm{d}\underline{s}(z) &= \oint_D \frac{f_j(u)}{u}\mathrm{d}u + c\int\oint_D \frac{tf_j(u)}{u(u-t)}\mathrm{d}u\mathrm{d}H(t)\\ &= (1-c)\oint_D \frac{f_j(u)}{u}\mathrm{d}u + c\int\oint_D \frac{f_j(u)}{u-t}\mathrm{d}u\mathrm{d}H(t)\\ &= -2\pi\mathrm{i}cK(c, f_j) + 2\pi\mathrm{i}cH(f_j),\end{aligned}$$

for $j = 1, \ldots, q$, where the last equation follows the residue theorem. Therefore, we get

$$\begin{aligned}H(f_j) &= K(c, f_j) + \frac{1}{2\pi\mathrm{i}c}\oint_D zf_j(u(z))\mathrm{d}\underline{s}(z)\\ &= K(c, f_j) + \frac{1}{2\pi\mathrm{i}c}\oint_C z\underline{s}'(z)f_j(-1/\underline{s}(z))\mathrm{d}z, \qquad (10)\end{aligned}$$

which is the first conclusion of the theorem.

The second conclusion follows from Lemma 1 and the dominated convergence theorem. □

Proof of Theorem 2 For simplicity, we prove the theorem by using the integral contour $C = \mathcal{C} \cup \overline{\mathcal{C}}$ where $\overline{\mathcal{C}} = \{x - \mathrm{i}v : x + \mathrm{i}v \in \mathcal{C}\}$. Then Lemma 2 holds on C.

From the formula (10),

$$H_p(f_j) = K(c_n, f_j) + \frac{1}{2\pi\mathrm{i}c_n}\oint_C z\underline{s}'_{c_n,H_p}(z)f_j(u_{c_n,H_p}(z))\mathrm{d}z,$$

for $j = 1, \ldots, q$. By the mean value theorem (Evard and Jafari, 1992),

$$\begin{aligned}&\widehat{H}(f_j) - H_p(f_j)\\ &\quad= \frac{1}{2\pi\mathrm{i}c_n}\oint_C z(\underline{s}'_n(z)f_j(u_n(z)) - \underline{s}'_{c_n,H_p}(z)f_j(u_{c_n,H_p}(z)))\mathrm{d}z\\ &\quad= \frac{1}{2\pi\mathrm{i}c_n}\oint_C z\Big((\underline{s}'_n(z) - \underline{s}'_{c_n,H_p}(z))f_j(u_n(z))\\ &\quad+ \underline{s}'_{c_n,H_p}(z)\Big\{\Re[f'_j(\xi_1(z))] + \Im[f'_j(\xi_2(z))]\Big\}(u_n(z) - u_{c_n,H_p}(z))\Big)\mathrm{d}z,\end{aligned}$$

where $\xi_1(z), \xi_2(z)$ are two points on the segment connecting $u_n(z)$ and $u_{c_n,H_p}(z)$.

Notice that, with probability one, for all n large,

$$\left|\oint_C Y_n(z) - \widehat{Y}_n(z)\mathrm{d}z\right| < K_1\varepsilon_n \quad \text{and} \quad \left|\oint_C Z_n(z) - \widehat{Z}_n(z)\mathrm{d}z\right| < K_2\varepsilon_n$$

for some constants K_1 and K_2, which both converge to zero as $n \to \infty$. In addition, for every $z \in \mathcal{C}$,

$$\begin{aligned}
&u_n(z),\ \xi_1(z),\ \xi_2(z),\ u_{c_n,H_p}(z) \to u(z),\\
&\underline{s}_n(z),\ \underline{s}_{c_n,H_p}(z) \to \underline{s}(z),\\
&\underline{s}'_{c_n,H_p}(z) \to \underline{s}'(z),
\end{aligned}$$

almost surely, as $n \to \infty$. From Lemma 2, we have

$$n(\widehat{H}(f_j) - H_p(f_j)) = \frac{1}{2\pi \mathrm{i} c} \oint_{C} z \left(\widehat{Z}_n(z) f_j(u(z)) + \widehat{Y}_n(z) u'(z) f_j'(u(z))\right) \mathrm{d}z + \delta_n,$$

with $\delta_n \to 0$ almost surely. Applying Lemma 2 again and the continuous mapping theorem, we get that the random vector

$$n\left(\widehat{H}(f_j) - H_p(f_j)\right)_{1 \le j \le q}$$

forms a tight sequence.

Moreover, under the assumptions in (ii) or (iii),

$$\begin{aligned}
n(\widehat{H}(f_j) - H_p(f_j)) \xrightarrow{\mathcal{D}} &\ \frac{1}{2\pi \mathrm{i} c} \oint_{C} z \left(Z(z) f_j(u(z)) + Y(z) u'(z) f_j'(u(z))\right) \mathrm{d}z\\
=&\ -\frac{1}{2\pi \mathrm{i} c} \oint_{C} Y(z) f_j(u(z)) \mathrm{d}z\\
:=&\ V_j, \quad j = 1, \ldots, q. \qquad (11)
\end{aligned}$$

where (11) is derived from the fact $Z(z) = Y'(z)$ and integration by part. It follows that

$$n\left(\widehat{H}(f_j) - H_p(f_j)\right)_{1 \le j \le q} \xrightarrow{\mathcal{D}} (V_j)_{1 \le j \le q} \qquad (12)$$

which is a Gaussian vector from the fact that Riemann sums corresponding to these integrals are multivariate Gaussian, and that weak limits of Gaussian vectors can only be Gaussian.

Applying the formulae (5), (6), and (11), for the real case, the limiting mean of (12) is given by

$$\mathrm{E}(V_j) = -\frac{1}{2\pi \mathrm{i}} \oint_{C} f_j(-1/\underline{s}(z)) \frac{\int t^2 \underline{s}'(z)^2 \mathrm{d}H(t)}{\underline{s}(z)(1 + \underline{s}(z))^3} \mathrm{d}z,$$

and the limiting covariance matrix is given by

$$\begin{aligned}
\mathrm{Cov}(V_i, V_j) &= -\frac{1}{4\pi^2 c^2} \oint_{C} \oint_{C'} f_i(u(z_1)) f_j(u(z_2)) \mathrm{Cov}(Y(z_1), Y(z_2)) \mathrm{d}z_1 \mathrm{d}z_2\\
&= -\frac{1}{4\pi^2 c^2} \oint_{C} \oint_{C'} f_i(u(z_1)) f_j(u(z_2)) k(z_1, z_2) \mathrm{d}z_1 \mathrm{d}z_2,
\end{aligned}$$

while for the complex case, $\mathrm{E}(V_j) = 0$ and $\mathrm{Cov}(V_i, V_j)$ is half of that in the real case. The proof of the theorem is complete. □

References

Bai, Z. D., Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probabilty*, *32*, 553–605.

Bai, Z. D., Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices* (2nd ed.). New York: Springer.

Bai, Z. D., Yao, J. F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré - Probabilité et Statistique*, *44*, 447–474.

Bai, Z. D., Chen, J. Q., Yao, J. F. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Australian & New Zealand Journal of Statistics*, *52*, 423–437.

Baik, J., Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, *97*, 1382–1408.

Bouchaud, J. P., Potters, M. (2011). Financial applications of random matrix theory: a short review. In The Oxford (Ed.), *Handbook of random matrix theory*. Oxford: Oxford University Press.

Chen, J. Q., Delyon, B., Yao, J. F. (2011). On a model selection problem from high-dimensional sample covariance matrices. *Journal of Multivariate Analysis*, *510*, 1388–1398.

Dudley, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. In A. Beck, R. Dudley, M. Hahn, J. Kuelbs, M. Marcus (Eds.), *Probability in Banach Spaces* (pp. 141–178). Berlin: Springer.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, *36*, 2757–2790.

Evard, J. C., Jafari, F. (1992). A complex rolle's theorem. *The American Mathematical Monthly*, *99*, 858–861.

Hachem, W., Loubaton, P., Mestre, X., Najim, J., Vallet, P. (2012). Large information plus noise random matrix models and consistent subspace estimation in large sensor networks. *Random Matrices: Theory and Applications*, *1*, 1150006.

Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, *29*, 295–327.

Kritchman, S., Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, *94*, 19–32.

Li, W. M., Yao, J. F. (2013). A local moment estimator of the spectrum of a large dimensional covariance matrix. *Statistica Sinica*. doi:10.5705/ss.2012.130.

Li, W. M., Chen, J. Q., Qin, Y. L., Yao, J. F., Bai, Z. D. (2013). Estimation of the population spectral distribution from a large dimensional sample covariance matrix. *Journal of Statistical Planning and Inference*, *143*(11), 1887–1897.

Marčenko, V. A., Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Matematicheskii Sbornik (New Series)*, *72*, 507–536.

Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transaction on Information Theory*, *54*, 5113–5129.

Passemier, D., Yao, J. F. (2012). On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices: Theory and Applications*, *1*, 1150002.

Rao, N. R., Mingo, J. A., Speicher, R., Edelman, A. (2008). Statistical eigen-inference from large wishart matrices. *Annals of Statistics*, *36*, 2850–2885.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *Journal of Multivariate Analysis*, *55*, 331–339.

Silverstein, J. W., Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *Journal of Multivariate Analysis*, *54*, 175–192.

Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Teoriya Veroyatnostei i ee Primenenia*, *1*, 289–319.

Yao, J. F., Kammoun, A., Najim, J. (2012). Eigenvalue estimation of parameterized covariance matrices of large dimensional data. *IEEE Transaction on Signal Processing*, *60*, 5893–5905.