Journal of Statistical Planning and Inference 143 (2013) 1887–1897

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Estimation of the population spectral distribution from a large dimensional sample covariance matrix

Weiming Li [a], Jiaqi Chen [b], Yingli Qin [c,1], Zhidong Bai [d,e,*,2], Jianfeng Yao [f]

[a] School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China
[c] Department of Statistics and Actuarial Science, University of Waterloo, Canada
[d] KLAS MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China
[e] Department of Statistics and Applied Probability, National University of Singapore, Singapore
[f] Department of Statistics and Actuarial Science, The University of Hong Kong, Hongkong, China

ARTICLE INFO

Article history:
Received 14 November 2012
Received in revised form
22 June 2013
Accepted 29 June 2013
Available online 18 July 2013

Keywords:
Empirical spectral distribution
High-dimensional data analysis
Marčenko–Pastur distribution
Large sample covariance matrices
Stieltjes transform

ABSTRACT

This paper introduces a new method to estimate the spectral distribution of a population covariance matrix from high-dimensional data. The method is founded on a meaningful generalization of the seminal Marčenko–Pastur equation, originally defined in the complex plane, to the real line. Beyond its easy implementation and the established asymptotic consistency, the new estimator outperforms two existing estimators from the literature in almost all the situations tested in a simulation experiment. An application to the analysis of the correlation matrix of S&P 500 daily stock returns is also given.

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 sizable to the sample size n, modern random matrix theory indicates that the sample covariance matrix

$$S_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^*$$

does not approach to Σ_p. For instance, in a simple case where $\Sigma_p = I_p$ (identity matrix), the eigenvalues of S_n will spread over an interval approximately equal to $(1 \mp \sqrt{p/n})^2$ around the unique population eigenvalue 1 of Σ_p (Bai and Yin, 1993; Marčenko and Pastur, 1967; Yin et al., 1988). Therefore, classical statistical procedures based on an approximation of Σ_p by S_n become inconsistent in such high dimensional data situations.

* Corresponding author. Tel.: +86 431 850 98161.
E-mail addresses: liwm601@gmail.com (W. Li), chenjq1016@gmail.com (J. Chen), yingli.qin@uwaterloo.ca (Y. Qin), baizd@nenu.edu.cn (Z. Bai), jeffyao@hku.hk (J. Yao).

[1] Yingli Qin is supported by the University of Waterloos Start-up Grant.
[2] Work of this author was partially supported by the NSF China Grant 11171057 and the Program for Changjiang Scholars and Innovative Research Team in University.

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.06.017

To be precise, let us recall that the *spectral distribution* (SD) G^A of an $m \times m$ Hermitian matrix (or real symmetric) A is the measure generated by its eigenvalues $\{\lambda_i^A\}$,

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

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

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

Following the random matrix theory, both sizes p and n grow to infinity. It is then natural to assume that H_p weakly converges to a limiting distribution H when $p \to \infty$. We refer this limiting SD H as the *population spectral distribution* (PSD) of the observation model.

The main observation is that under reasonable assumptions, when both dimensions p and n become large at a proportional rate say c, almost surely, the (random) SD G^{S_n} of the sample covariance matrix S_n will weakly converge to a deterministic distribution F, called *limiting spectral distribution* (LSD). Naturally this LSD F depends on the PSD H, but in general this relationship is complex and has no explicit form. The only exception is the case where all the population eigenvalues (σ_i) are unit, i.e. $\Sigma_p \equiv I_p$ $(H = \delta_1)$; the LSD F is then explicit known to be the Marčenko–Pastur (MP) distribution with an explicit density function. For a general PSD H, this relationship is expressed via an implicit equation, see Section 3, Eqs. (2.1) and (2.3).

An important question here is the recovering of the PSD H (or H_p) from the sample covariance matrix S_n. This question has a central importance in several popular statistical methodologies like Principal Component Analysis (Johnstone, 2001), Kalman filtering or Independent Component Analysis which all rely on efficient estimations of some population covariance matrices.

Recently, El Karoui (2008) has proposed a variational and nonparametric approach to this problem based on an appropriate distance function using the MP equation (2.1) below and a large dictionary made with base density functions and Dirac point masses. The proposed estimator is proved consistent in a nonparametric estimation sense assuming both the dictionary size and the number of observations n tend to infinity. However, no result on the convergence rate of the estimator, e.g. a central limit theorem, is given.

In another important work Rao et al. (2008), the authors propose to use a suitable set of empirical moments, say the first q moments: for $k=1,\ldots,q$, $\hat{\alpha}_k = p^{-1}\mathrm{tr}S_n^k = p^{-1}\sum_{l=1}^{p}\lambda_l^k$ where (λ_l) are the eigenvalues of S_n (assuming $p \le n$). Here a pure parametric approach is adopted and the PSD depends on a set of real parameters θ: $H = H(\theta)$. Therefore, when $n \to \infty$ and under appropriate normalization, the sample moments $(\hat{\alpha}_k)$ will have a Gaussian limiting distribution with asymptotic mean and variance $\{m_\theta, Q_\theta\}$ which are functions of the (unknown) parameters θ. In Rao et al. (2008), the authors propose an estimator $\hat{\theta}_R$ of the parameters by maximizing the asymptotic Gaussian likelihood of $\hat{\alpha} = (\hat{\alpha}_j)_{1 \le j \le q}$, with distribution $N_q(m_\theta, Q_\theta)$. Intensive simulations illustrate the consistency and the asymptotic normality of this estimator. However, their simulation experiments are limited to simplest situations and no theoretic result is provided concerning the consistency of the estimator. An important difficulty in this approach is that the functions m_θ and Q_θ have no explicit form.

In a recent work Bai et al. (2010), a modification of the procedure in Rao et al. (2008) is proposed to get a direct moments estimator based on the sample moments $(\hat{\alpha}_j)$. Compared to El Karoui (2008) and Rao et al. (2008), this moment estimator is simpler and much easier to implement. Moreover, the convergence rate of this estimator (asymptotic normality) is also established. A recent paper by the authors in Chen et al. (2011) has also analyzed the underlying order selection problem and proposed a solution based on the cross-validation principle.

However, despite all the above contributions, there is still a need for new methods of estimation. Actually, the general approach in El Karoui (2008) has several implementation issues that seem to be responsible for its relatively low performance as attested by the very simple nature of provided simulation results. This low efficiency is probably due to the use of a too general dictionary made with large number of discrete distributions and piece-wisely linear densities. Concerning the moment based methods in Rao et al. (2008) and Bai et al. (2010), we will see that their accuracies degrade drastically as the number of parameters to be estimated increases. Lastly, it is well known that the contour-integral based method in a related work Mestre (2008) is limited to a small class of discrete models where distinct population eigenvalues should generate non-overlapping clusters of sample eigenvalues.

The new approach developed in this paper can be viewed as a synthesis of the optimization approach in El Karoui (2008) and the parametric setup in Bai et al. (2010). On one hand, we adopt the optimization approach and will prove that it is in general preferable to the moment approaches. On the other hand, using a generic parametric approach for discrete PSDs as well as continuous PSDs, we are able to avoid the aforementioned implementation difficulties in El Karoui (2008). Another important contribution from the paper is that the optimization problem has been moved from the complex plane to the real line by considering a characteristic equation (MP equation) on the real line. The obtained optimization procedure is then much easier to implement.

The rest of the paper is organized as follows. In the next section, we provide an MP equation defined on the real line which will be the corner-stone of our estimation method. This method is developed in Section 3 and we prove its strong

consistency. Then, in Section 4, simulation experiments are carried out to compare the performance of three estimation methods under investigation. The last section collects proofs of main theorems.

2. Marčenko–Pastur equation on the real line

Throughout the paper, $A^{1/2}$ stands for any Hermitian square root of a non-negative definite Hermitian matrix A. Our model assumptions are as follows.

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}), $i,j\geq 1$ satisfying

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

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

Assumption (c). The SD H_p of Σ_p weakly converges to a probability distribution H as $p\to\infty$.

The Assumptions (a)–(c) are classical conditions for the celebrated MP theorem (Marčenko and Pastur, 1967; Silverstein, 1995; Bai and Silverstein, 2010). More precisely, under these assumptions, almost surely, as $n\to\infty$, the empirical SD $F_n:=G^{S_n}$ of S_n, weakly converges to a (nonrandom) generalized MP distribution F.

Unfortunately, except the simplest case where $H\equiv\delta_{\sigma^2}$, the LSD F has no explicit form and it is characterized as follows. Let $\underline{s}(z)$ denote the Stieltjes transform of $cF+(1-c)\delta_0$, which is a one-to-one map defined on the upper half complex plane $\mathbb{C}^+=\{z\in\mathbb{C}:\Im(z)>0\}$. This transform satisfies the following fundamental MP equation:

$$z=-\frac{1}{\underline{s}(z)}+c\int\frac{t}{1+t\underline{s}(z)}dH(t),\quad z\in\mathbb{C}^+. \tag{2.1}$$

The above MP equation excludes the real line from its domain of definition. As the first contribution of the paper, we fill this gap by an extension of the MP equation to the real line. The estimation method introduced in Section 3 will be entirely based on this extension.

The support of a distribution G is denoted by S_G and its complementary set by S_G^c. Since the ESD F_n is observed, we usually use $\underline{s}_n(z)$, the Stieltjes transform of $(p/n)F_n+(1-p/n)\delta_0$ to approximate $\underline{s}(z)$ defined in the MP equation on $\mathbb{C}^+$. Accordingly, for $u\in\mathbb{R}$, we define

$$\underline{s}_n(u)=-\frac{1-p/n}{u}+\frac{1}{n}\sum_{l=1}^{p}\frac{1}{\lambda_l-u}. \tag{2.2}$$

Denote by F^{c_n,H_p} the LSD associated with the MP equation (2.1) where c and H are replaced by $c_n=p/n$ and H_p, respectively, and let $U=\liminf_{n\to\infty}S^c_{F^{c_n,H_p}}\backslash\{0\}$. Then with probability one, for all large n, $\underline{s}_n(u)$ is well defined on the interior $\mathring{U}$ of U from the "no eigenvalues" conclusion in Bai and Silverstein (1998).

The set $\mathring{U}$ is in general a subset of $S_F^c\backslash\{0\}$, since there may be some "spiked" sample eigenvalues that have limits in S_F^c but not in $\mathring{U}$. For a discrete model, however, if $H_p=H$ is assumed for all large p then $\mathring{U}=S_F^c\backslash\{0\}$.

Theorem 2.1. *Assume that the Assumptions (a)–(c) hold. Then*

(1) *for any* $\mathring{U}$, $\underline{s}_n(u)$ *converges to* $\underline{s}(u)$, *almost surely*,
(2) *for any* $u\in S_F^c$, $\underline{s}=\underline{s}(u)$ *is a solution to equation*

$$u=-\frac{1}{\underline{s}}+c\int\frac{t}{1+t\underline{s}}dH(t), \tag{2.3}$$

(3) *the solution is also unique in the set* $B^+=\{\underline{s}\in\mathbb{R}\backslash\{0\}:du/d\underline{s}>0,(-\underline{s})^{-1}\in S_H^c\}$,
(4) *for any non-empty open interval* $(a,b)\subset B^+$, H *is uniquely determined by* $u(\underline{s})$, $\underline{s}\in(a,b)$.

The proof is given in the last section. Some remarks are in order:

1. Notice that since $(-\infty,0)\subset\mathring{U}\subset S_F^c$, there are infinitely many u-points such that $\underline{s}_n(u)$ almost surely converges to $\underline{s}(u)$.
2. The MP equation (2.3) can be inverted in the following sense: the knowledge of $u(\underline{s})$ on any interval in B^+ (see Fig. 1) uniquely determines the PSD H. The estimation method in Section 3 will be built on this property.

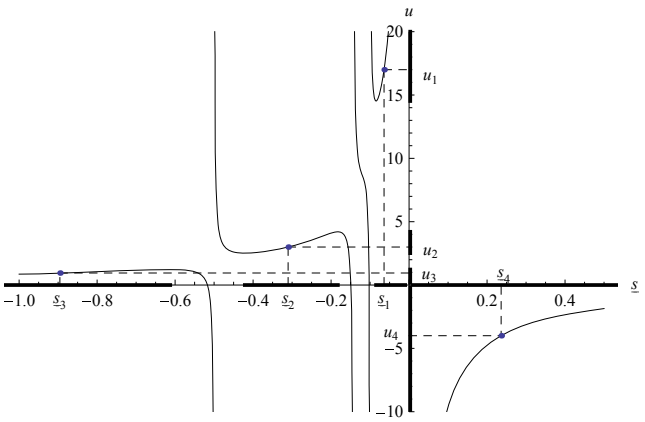

Fig. 1. The curve of $u=u(\underline{s})$ (solid thin), and the sets B^+ and S_F^c (solid thick) for $H=0.3\delta_2+0.4\delta_7+0.3\delta_{10}$ and $c=0.1$. $u_i=u(\underline{s}_i), \underline{s}_i\in B^+, i=1,2,3,4$.

3. Estimation

3.1. The method

We consider the estimation problem in a parametric setup. Suppose $H=H(\theta)$ is the limit of H_p with unknown parameter vector $\theta\in\Theta\subset\mathbb{R}^q$. The procedure of the estimation of H includes three steps:

S1. Choose a u-net $\{u_1,\ldots,u_m\}$ from $\mathring{\mathrm{U}}$, where u_j's are distinct and the size m is not less than q.
S2. For each u_j, calculate $\underline{s}_n(u_j)$ using (2.2) and plug the pair into the MP equation (2.3). Then, we obtain m approximate equations

$$u_j\simeq -\frac{1}{\underline{s}_n(u_j)}+\frac{p}{n}\int\frac{t\,dH(t,\theta)}{1+t\underline{s}_n(u_j)}:=\widehat{u}_j\left(\underline{s}_{nj},\theta\right)\quad (j=1,\ldots,m).$$

S3. Find the least squares solution of θ,

$$\widehat{\theta}_n=\arg\min_{\theta\in\Theta}\sum_{j=1}^{m}(u_j-\widehat{u}_j(\underline{s}_{nj},\theta))^2.$$

We name $\widehat{\theta}_n$ as the least squares estimate (LSE) of θ. Accordingly, $\widehat{H}=H(\widehat{\theta}_n)$ is called the LSE of H. Apparently, the estimates depend on the choice of the u-net, which is a central issue of this estimation procedure and will be discussed in Section 4.

A similar procedure of the estimation in the complex field was discussed in El Karoui (2008). In their first two steps, they chose a z-net (a list of complex numbers) from $\mathbb{C}^+$ and created a system of approximate equations. Then the PSD H is discretized as a weighted sum of a grid of pre-chosen mass points, and thus only the weight parameters are unknown and need to be estimated. This discretization tactfully converts the model estimation into a problem of convex optimization. However, the choice of the mass points and its accompanying calculations are added burdens especially when the range of the support of H is large.

In the parametric setup, the optimization procedure may encounter a tough issue as it is a problem of non-convex nonlinear global optimization. It can be verified that the optimization is hard to achieve for many cases when using a z-net from $\mathbb{C}^+$. This is one of our motivations to move the optimization from the complex field to the real field.

3.2. Consistency

We establish the strong consistency of our estimator in two models that are widely used in the literature. The estimates will be further studied in the simulation section.

The first model is made with discrete PSDs with finite support on $\mathbb{R}^+$, i.e.

$$H(\theta)=m_1\delta_{a_1}+\cdots+m_k\delta_{a_k},\quad \theta\in\Theta,$$

where $m_k=1-\sum_{i=1}^{k-1}m_i$, $\theta=(a_1,\ldots,a_k,m_1,\ldots,m_{k-1})$ are $(2k-1)$ unknown parameters and

$$\Theta=\left\{\theta\in\mathbb{R}^{2k-1}:m_i>0,\ \sum_{i=1}^{k}m_i=1;0<a_1<\cdots<a_k<+\infty\right\}.$$

Here, Eq. (2.3) can be simplified to

$$u = -\frac{1}{\underline{s}} + c\sum_{i=1}^{k}\frac{a_i m_i}{1+a_i\underline{s}}.$$

For the well-definition of the equation on Θ, we assume that the u-net satisfies

$$\inf_{\theta\in\Theta}\min_{i,j}|1+a_i\underline{s}(u_j)|\geq\delta, \tag{3.1}$$

where δ is some positive constant. It is clearly satisfied if all the u_j's are negative.

Theorem 3.1. *In addition to the Assumptions (a)–(c), suppose that the true value of the parameter θ_0 is an interior point of Θ and the condition (3.1) is fulfilled. Then, the LSE $\widehat{\theta}_n$ for the discrete model is strongly consistent, that is, almost surely, $\widehat{\theta}_n\to\theta_0$.*

Next we suppose that the PSD $H(\theta)$ has a probability density $h(t|\theta)$ with respect to Lebesgue measure. From Szegö (1959) (Chapters 2, 4), if $h(t|\theta)$ has finite moments of all order, it can be expanded in terms of Laguerre polynomials

$$h(t|\theta) = \sum_{j\geq 0} c_j\psi_j(t)e^{-t},$$

where

$$c_j = \int \psi_j(t)h(t|\theta)\,dt.$$

As discussed in Bai et al. (2010), we consider a family of $h(t|\theta)$ with finite expansion

$$h(t|\theta) = \sum_{j=0}^{q} c_j\psi_j(t)e^{-t} = \sum_{j=0}^{q}\alpha_j t^j e^{-t}, \quad t>0, \quad \theta\in\Theta,$$

where $\alpha_0 = 1-\alpha_1-\cdots-q!\alpha_q$, $\theta=(\alpha_1,\ldots,\alpha_q)$, and

$$\Theta = \{\theta\in\mathbb{R}^q : h(t|\theta)>0,\ t\in\mathbb{R}^+\}.$$

For this model, Eq. (2.3) becomes

$$u = -\frac{1}{\underline{s}} + c\sum_{j=0}^{q}\alpha_j\int\frac{t^{j+1}e^{-t}}{1+t\underline{s}}dt.$$

It is clear that the calculation of $\widehat{\theta}_n$ is here simple since the above equation is linear with respect to θ.

Theorem 3.2. *In addition to the Assumptions (a)–(c), suppose that the true value of the parameter θ_0 is an interior point of Θ. Then, the LSE $\widehat{\theta}_n$ for the continuous model is strongly consistent.*

4. Simulation experiments

In this section, simulations are carried out to compare our LSE with the approximate quasi-likelihood estimate in Rao et al. (2008) (referred as RMSE) and the moment estimate in Bai et al. (2010) (referred as BCY). We do not include the estimator of El Karoui (2008) in this study since this estimator is nonparametric using a suitable approximation dictionary while the LSE is based on a parametric form of unknown PSDs.

We study five different PSDs: three of them are discrete and two continuous. Samples are drawn from mean-zero real normal population with the dimensions $n=500$ and $p=100$, 500, 1000. Statistics are computed from 1000 independent replications.

To evaluate the quality of an estimate $\widehat{H}=H(\widehat{\theta})$, instead of looking at individual values $(\widehat{\theta}_i)$ of the parameters, we use a global distance, namely the Wasserstein distance $(\mathrm{WD}) = \int|Q_H(t)-Q_{\widehat{H}}(t)|dt$ where $Q_\mu(t)$ is the quantile function of distribution μ. The use of Wasserstein distance is motivated by the fact that it applies to both discrete and continuous distributions (unlike other common distance such as the Kullback–Leibler or L_2 distance).

For the LSE, we need to choose a u-net from $S^c_{F_n}\cap S^c_F\backslash\{0\}$. When H has finite support, the upper and lower bounds of $S_F\backslash\{0\}$ can be estimated respectively by $\lambda_{max}=\max\{\lambda_i\}$ and $\lambda_{min}=\min\{\lambda_i:\lambda_i>0\}$ where λ_i's are sample eigenvalues. As a consequence, we design a primary set:

$$\mathcal{U} = \begin{cases} (-10,0)\cup(0,0.5\lambda_{min})\cup(5\lambda_{max},10\lambda_{max}) & (\text{discrete model},\ p\neq n),\\ (-10,0)\cup(5\lambda_{max},10\lambda_{max}) & (\text{discrete model},\ p=n),\\ (-10,0) & (\text{continuous model}). \end{cases}$$

Next, we choose l equally spaced u-points from each individual interval of $\mathcal{U}$. We name this process as adaptive choice of u-net. Here we set $l=20$ for all cases considered in simulation, that is, for example we take $\{-10+10t/21, t=1,\ldots,20\}$ from the first interval. Hence, the sizes of the u-nets are respectively $m=3l=60$ for discrete models with $p\neq n$, $m=2l=40$ for those with $p=n$, and $m=l=20$ for continuous models.

*Case*1: $H=0.5\delta_1+0.5\delta_2$. This is a simple case as H has only two atoms with equal weights. Table 1 shows that all the three estimates are consistent, and their efficiencies are very close.

*Case*2: $H=0.3\delta_1+0.4\delta_3+0.3\delta_5$. In this case, we increase the order of H. Analogous statistics are summarized in Table 2. The results show that LSE clearly outperforms RMSE and BCY in the light of the Wasserstein distance. Particularly, RMSE and BCY have not converged yet with dimensions $n=500$ and $p=500$, 1000, while LSE only contains a small bias in such situations. This exhibits the robustness of our method with respect to the increase of the order.

*Case*3: $H=0.3\delta_1+0.4\delta_5+0.3\delta_{15}$. In this case, we increase the variance of H. Table 3 collects the simulation results. Compared with Table 2, RMSE and BCY deteriorate significantly while LSE remains stable. The average Wasserstein distances of LSE are (at least) a third less than those of RMSE and BCY for all p and n used. This demonstrates the robustness of our method with respect to the increase of the variance.

*Case*4: $h(t)=(\alpha_0+\alpha_1 t)e^{-t}, \alpha_1=1$. This is the simplest continuous model with only one parameter to be estimated. In this case, H is a gamma distribution with shape parameter 2 and scale parameter 1. Statistics in Table 4 show that all the three estimates have similar efficiency.

*Case*5: $h(t)=(\alpha_0+\alpha_1 t+\alpha_2 t^2+\alpha_3 t^3)e^{-t}, \alpha_1=\alpha_2=\alpha_3=1/9$. This model with three parameters becomes more difficult to estimate. From the results in Table 5, our LSE performs fairly well and again outperform these two moment based methods.

In summary, the LSE is superior to the RMSE and BCY estimators in all the tested situations. On the other hand, as expected, the performances of the RMSE and the BCY estimators are very close since they are all based on empirical moments (however, as explained in Bai et al. (2010), the BCY estimator is much easier to implement).

We do not compare the LSE estimator with that introduced in Mestre (2008) for the above cases, as the latter requires a separation assumption, which is really harsh. In fact, except the third case with dimension $p=100$, all others do not satisfy this separation condition. It is however known in the literature that Mestre's method is very efficient when this separation condition is satisfied.

Finally, we analyze the relationship between the size of a u-net and the efficiency of LSE. The average of Wasserstein distances of LSE with respect to different l values (the number of u-points picked from each individual interval) is plotted for Case 3 and Case 5, see Fig. 2. The results show that unless l is too small, the estimation efficiency remains remarkably stable with different values of l.

Another noteworthy phenomenon in Fig. 2 is that in Case 3, the Wasserstein distances increase with p as we have expected, while in Case 5 they actually decrease. This counter-intuitive phenomenon in Case 5 can be attributed mostly to

Table 1
Wasserstein distances of estimates for $H=0.5\delta_1+0.5\delta_2$.

Method	WD	$p/n=0.2$	$p/n=1$	$p/n=2$
LSE	Mean	0.0437	0.0601	0.0893
	S.D.	0.0573	0.0735	0.1077
RMSE	Mean	0.0491	0.0689	0.0859
	S.D.	0.0320	0.0482	0.0629
BCY	Mean	0.0500	0.0664	0.0871
	S.D.	0.0331	0.0466	0.0617

Table 2
Wasserstein distances of estimates for $H=0.3\delta_1+0.4\delta_3+0.3\delta_5$.

Method	WD	$p/n=0.2$	$p/n=1$	$p/n=2$
LSE	Mean	0.1589	0.3566	0.4645
	S.D.	0.1836	0.4044	0.5156
RMSE	Mean	0.2893	0.7494	0.8153
	S.D.	0.0966	0.2188	0.1080
BCY	Mean	0.2824	0.5840	0.7217
	S.D.	0.1769	0.2494	0.2156

Table 3
Wasserstein distances of estimates for $H=0.3\delta_1+0.4\delta_5+0.3\delta_{15}$.

Method	WD	$p/n=0.2$	$p/n=1$	$p/n=2$
LSE	Mean	0.1756	0.2524	0.5369
	S.D.	0.2105	0.3013	0.6282
RMSE	Mean	0.7090	1.4020	1.9160
	S.D.	0.0524	0.6501	0.2973
BCY	Mean	0.9926	1.5379	1.8562
	S.D.	0.5618	0.6875	0.7526

Table 4
Wasserstein distances of estimates for $h(t)=te^{-t}$.

Method	WD	$p/n=0.2$	$p/n=1$	$p/n=2$
LSE	Mean	0.0939	0.0441	0.0294
	S.D.	0.0704	0.0317	0.0229
RMSE	Mean	0.1126	0.0508	0.0346
	S.D.	0.0839	0.0393	0.0262
BCY	Mean	0.1168	0.0491	0.0348
	S.D.	0.0881	0.0361	0.0268

Table 5
Wasserstein distances of estimates for $h(t)=(t+t^2+t^3)e^{-t}/9$.

Method	WD	$p/n=0.2$	$p/n=1$	$p/n=2$
LSE	Mean	0.1895	0.0902	0.0740
	S.D.	0.1103	0.0526	0.0378
RMSE	Mean	0.3163	0.1515	0.1156
	S.D.	0.2062	0.0863	0.0670
BCY	Mean	0.3139	0.1554	0.1114
	S.D.	0.2007	0.0907	0.0624

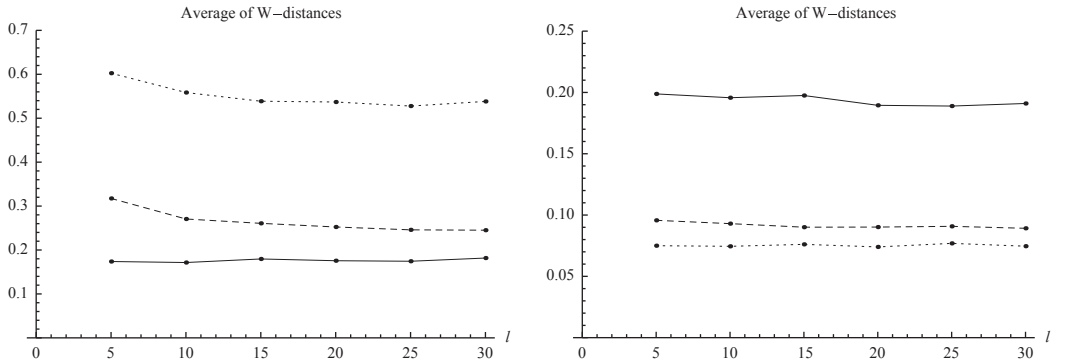

Fig. 2. The average of Wasserstein distances of LSE with respect to l ($l=5,10,\ldots,30$) for Case 3 (left) and Case 5 (right) with $p=100$, $n=500$ (solid lines), $p=100$, $n=500$ (dashed lines), and $p=1000$, $n=500$ (dotted lines).

the following facts. For discrete models, we could and do assume $H_p=H$ as H_p is really what we are interested in. For continuous models, however, this assumption is obviously invalid and H_p has to be deemed as a random distribution, i.e. an empirical distribution sampled from H. Hence, there are (random) model errors between the sampled H_p's and the theoretical model H. When the dimension p increases the model errors become small, which contributes to the decrease of the Wasserstein distances.

5. Application to S&P 500 daily stock data

In this section, we present a financial application of our estimation procedure in analysing an empirical correlation matrix of stock returns. We study a set of 488 U.S. stocks included in the S&P 500 index from September, 2007 to September 2011 (1001 trading days, 12 stocks have been removed because of missing values). Here, the data dimension is $p=488$ and the number of observations is $n=1000$.

Following Bouchaud and Potters (2011), we suppose that there is a PSD $H(\alpha)$ for the stock returns with an inverse cubic density $h(t|\alpha)$:

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

where $c=2(1-\alpha)^2$ and $a=2\alpha-1$. Notice that when $\alpha\to1^-$, the inverse cubic model tends to the MP case ($H=\delta_1$), so that this prior model is very flexible.

For the estimation procedure, we first remove the 6 largest sample eigenvalues which are deemed as spikes over the bulk of sample eigenvalues. As in Section 3, we use $l=20$ equally spaced u-points in $(-10,0)$. The LSE of α turns out to be $\widehat{\alpha}=0.4380$. The RMSE and BCY do not exist for this model for the reason that the moments of H do not depend on the unknown parameter.

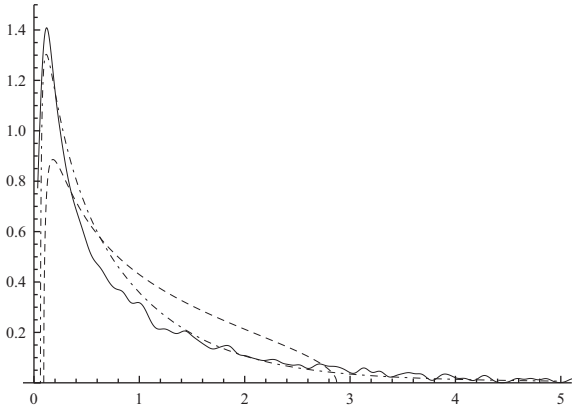

Fig. 3. The empirical density of the sample eigenvalues (plain black line), compared to the MP density (dashed line) and the limiting spectral density corresponding to the LSE estimate $h(t|0.4380)$ (dashed-dotted line).

Limiting spectral densities corresponding to the LSE $h(t|0.4380)$ and $H=\delta_1$ are shown in Fig. 3. We also plot the empirical spectral density of the correlation matrix, and the curve is smoothed by using a Gaussian kernel estimate with bandwidth $h=0.05$.

From Fig. 3, we could see that the MP density is far away from the empirical density curve. This confirms a widely believed fact that the correlation matrix may have more structure than just several spikes on top of the identity matrix. By contrast, the cubic model with $\alpha=0.4380$ yields a much more satisfying fit to the empirical density curve.

6. Proofs

We first recall useful results in three lemmas. The first one is provided in Silverstein (1995) and the two others in Silverstein and Choi (1995).

Lemma 6.1. *Assume that the Assumptions (a)–(c) hold. Then, almost surely, the empirical spectral distribution F_n converges in distribution, as $n\to\infty$, to a non-random probability measure F, whose Stieltjes transform $s=s(z)$ is a solution to the equation*

$$s=\int \frac{1}{t(1-c-czs)-z}\,dH(t).$$

The solution is also unique in the set $\{s\in\mathbb{C}: -(1-c)/z+cs\in\mathbb{C}^+\}$.

Lemma 6.2. *If $u\in S_F^c\backslash\{0\}$, then $\underline{s}=\underline{s}(u)$ satisfies*

$$(1)\underline{s}\in\mathbb{R}\backslash\{0\},\quad (2)(-\underline{s})^{-1}\in S_H^c,\quad (3)du/d\underline{s}>0.$$

Conversely, if $\underline{s}$ satisfies (1)–(3), then $u=u(\underline{s})\in S_F^c\backslash\{0\}$.

Lemma 6.3. *Set $B=\{\underline{s}\in\mathbb{R}\backslash\{0\}: (-\underline{s})^{-1}\in S_H^c\}$. Let $[\underline{s}_1,\underline{s}_2]$, $[\underline{s}_3,\underline{s}_4]$ be two disjoint intervals in B satisfying for all $\underline{s}\in[\underline{s}_1,\underline{s}_2]\cup[\underline{s}_3,\underline{s}_4]$, $du/d\underline{s}>0$. Then $[u_1,u_2],[u_3,u_4]$ are disjoint where $u_i=u(\underline{s}_i), i=1,2,3,4$.*

6.1. Proof of Theorem 2.1

The first conclusion follows from two convergence theorem. In fact, for any fixed $\mathring{u}$ there are $\varepsilon_0\in\mathbb{R}^+$ and $n_0\in\mathbb{Z}^+$ such that the neighborhood $U(u,\varepsilon_0)$ of u is a subset of $\cap_{n=n_0}^{\infty} S_{F_n}^c\backslash\{0\}$. This implies that for all $x\in S_{F_n}$ and $n>n_0$, we have $|1/(x-u)|<1/\varepsilon_0$. From this and Lebesgue's dominated convergence theorem, for any fixed $\mathring{u}$, almost surely,

$$\underline{s}_n(u)\to\underline{s}(u),$$

as $n\to\infty$ with $p/n\to c>0$. By Vitali's convergence theorem (Titchmarsh, 1939; Bai and Silverstein, 2004), we may conclude that $\underline{s}_n(u)$ converges almost surely for every $\mathring{u}$.

Next, we consider the second conclusion. For any fixed $u\in S_F^c$ and $\varepsilon>0$, let $z=u+\varepsilon i$, from Lemma 6.1, $\underline{s}(z)$ satisfies (2.1). On the other hand, according to Lemma 6.2, $(-\underline{s}(u))^{-1}\in S_H^c$ and thus $|t/(1+t\underline{s}(z))|$ is bounded on the set $\{(t,\varepsilon): S_H\times[0,1]\}$. Therefore, by Lebesgue's dominated convergence theorem and the continuity of the Stieltjes transform $s(z)$ for $z\in\mathbb{C}\backslash S_F$ (Silverstein and Choi, 1995), taking the limit as $\varepsilon\to0^+$ on both sides of (2.1) gets the conclusion.

Conclusion 3 follows from the results of Lemmas 6.2 and 6.3. In fact, let $u_{B^+}(\underline{s})$ be the restriction of $u(\underline{s})$ to B^+, then Lemma 6.2 shows that the range of $u_{B^+}(\underline{s})$ is $S_F^c\backslash\{0\}$. Lemma 6.3 indicates that $u_{B^+}(\underline{s})$ is also an injection. Therefore, $u_{B^+}(\underline{s})$ is a bijection from B^+ to $S_F^c\backslash\{0\}$.

As to the last conclusion, suppose $H_1(t)$ and $H_2(t)$ are two population spectral distribution functions satisfying, for all $\underline{s}\in(a,b)$,

$$\int \frac{t}{1+t\underline{s}}dH_1(t)=\int \frac{t}{1+t\underline{s}}dH_2(t). \tag{6.1}$$

We are going to show $H_1=H_2$ almost everywhere with respect to Lebesgue measure on $\mathbb{R}$.

For any $\underline{s}_0\in(a,b)$, $-1/\underline{s}_0$ is an interior point of S_H^c, then there is $\delta_0>0$ such that

$$U(-1/\underline{s}_0,\delta_0/|\underline{s}_0|)\subset S_H^c,$$

which implies $|1+t\underline{s}_0|>\delta_0$ for all $t\in S_H$. Choose $\varepsilon_0=\min\{|\underline{s}_0|\delta_0/(1+\delta_0), b-\underline{s}_0, \underline{s}_0-a\}$. Then, for any $\underline{s}\in U(\underline{s}_0,\varepsilon_0)$, $1+t\underline{s}$ has the same sign as $1+t\underline{s}_0$. Define

$$g(t,u)=\begin{cases}1 & (1+t\underline{s}_0>0,\ u>0),\\ -1 & (1+t\underline{s}_0<0,\ u<0),\\ 0 & (u(1+t\underline{s}_0)u\le 0).\end{cases}$$

We have then

$$\frac{t}{1+t\underline{s}}=\int_{-\infty}^{+\infty} g(t,u)e^{-(1/t+\underline{s})u}du,\ \underline{s}\in U(\underline{s}_0,\varepsilon_0).$$

Therefore, each side of (6.1) can be expressed as

$$\int \frac{t}{1+t\underline{s}}dH_i(t)=\int_{-\infty}^{+\infty}\int_0^{\infty} g(t,u)e^{-(1/t+\underline{s}_0)u}dH_i(t)e^{-(\underline{s}-\underline{s}_0)u}du. \tag{6.2}$$

It is clear that the left hand side of (6.2) is the Laplace transform of

$$\int_0^{\infty} g(t,u)e^{-(1/t+\underline{s}_0)u}dH_i(t).$$

By the uniqueness of Laplace transform, we have then

$$\int_0^{\infty} g(t,u)e^{-1/tu}dH_1(t)=\int_0^{\infty} g(t,u)e^{-1/tu}dH_2(t),$$

and thus $H_1=H_2$ almost everywhere.

6.2. Proof of Theorem 3.1

Define

$$\varphi(\theta)=\sum_{j=1}^{m}(u_j-u(s_j,\theta))^2,$$

where $s_j=\underline{s}(u_j)$ $(j=1,\dots,m)$. We first state and prove the following proposition.

Proposition 6.1. *If $u_1,\dots,u_m$ are distinct and $m\ge q=2k-1$, then $\varphi(\theta)=0$ for the discrete model has a unique solution θ_0 on Θ.*

Proof. Since $\underline{s}(u)$ is a bijective function from S_F^c to B^+ and $u_1,\dots,u_m$ are distinct, $s_1,\dots,s_m$ are also distinct.

Suppose there is a $\theta=(a_1,\dots,a_k,m_1,\dots,m_{k-1})$ such that $\varphi(\theta)=0$. Denote by $\theta_0=(a'_1,\dots,a'_k,m'_1,\dots,m'_{k-1})$ the true value of the parameter. We will show that $\theta=\theta_0$. Denote $b_i=1/a_i$ and $b'_i=1/a'_i$ $(i=1,\dots,k)$, we have then

$$\sum_{i=1}^{k}\frac{m_i}{s_j+b_i}=\sum_{i=1}^{k}\frac{m_i'}{s_j+b_i'}\quad (j=1,\dots,m). \tag{6.3}$$

Now look s_j as a parameter s and reduction to common factors leads to

$$(s+b_1')\cdots(s+b_k')\sum_{i=1}^{k}m_i\prod_{\ell\ne i}(s+b_\ell)=(s+b_1)\cdots(s+b_k)\sum_{i=1}^{k}m_i'\prod_{\ell\ne i}(s+b_\ell').$$

These are polynomials of degree $2k-1$; they coincide at $m\ge 2k-1$ different points $s=s_j$; they are then equal. Back to (6.3), we have now for all $s\ne -b_i,-b_i'$,

$$\sum_{i=1}^{k}\frac{m_i}{s+b_i}=\sum_{i=1}^{k}\frac{m_i'}{s+b_i'}.$$

Now each b_i should match one b_ℓ', because otherwise $b_i\ne b_\ell'$ for all ℓ and by letting $s\to -b_i$ we get a contradiction. So there is one b_ℓ' matches (then unique) for b_i. This proves also that $m_i=m_\ell'$. As the b_i are ordered, it is necessary that $b_\ell'=b_i'$ and hence also $m_i=m_i'$. □

Now let us begin the proof of Theorem 3.1. Recall that

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta}\sum_{j=1}^{m}\left(u_j+\frac{1}{\underline{s}_n(u_j)}-\frac{p}{n}\int\frac{tdH(t,\theta)}{1+t\underline{s}_n(u_j)}\right)^2 := \arg\min_{\theta\in\Theta}\varphi_n(\theta).$$

Under the assumption of the theorem, by the convergence of $\underline{s}_n(u_j)$ $(j=1,\ldots,m)$, $\varphi_n(\theta)$ is well defined on Θ for all large n. Moreover, for any fixed $\theta\in\Theta$, we have

$$\varphi_n(\theta)\to\varphi(\theta),$$

almost surely. Proposition 6.1 guarantees that $\theta=\theta_0$ is the unique solution to $\varphi(\theta)=0$ on Θ.

We claim that for almost all ω, there is a compact set $\overline{\Theta}=\overline{\Theta}(\omega)\subset\Theta$ which contains all $\widehat{\theta}_n(\omega)$ for large n. It is easy to see that for all large n, $\varphi_n(\theta)$ is uniformly bounded on $\overline{\Theta}$ and has continuous partial derivatives with respect to θ. By the Vitali's convergence theorem, we get

$$\sup_{\theta\in\overline{\Theta}}|\varphi_n(\theta)-\varphi(\theta)|\to 0. \tag{6.4}$$

For any $\varepsilon>0$, by the continuity of $\varphi(\theta)$, we have

$$\inf_{\substack{\|\theta-\theta_0\|>\varepsilon\\ \theta\in\overline{\Theta}}}\varphi(\theta)>\varphi(\theta_0)=0.$$

From this and (6.4), when n is large,

$$\inf_{\substack{\|\theta-\theta_0\|>\varepsilon\\ \theta\in\overline{\Theta}}}\varphi_n(\theta)>\varphi_n(\theta_0).$$

This proves that minimum point $\widehat{\theta}_n$ of $\varphi_n(\theta)$ for $\theta\in\overline{\Theta}$ must be in the ball $\{\|\theta-\theta_0\|\le\varepsilon\}$. Hence the convergence $\widehat{\theta}_n\to\theta_0$.

To complete the proof, it is sufficient to prove the claim, i.e. there is a compact set $\overline{\Theta}\subset\Theta$ such that for large n,

$$\inf_{\theta\in\overline{\Theta}^c}\varphi_n(\theta)>\varphi_n(\theta_0).$$

Suppose not. Then there exists a sequence $\{\theta_l, l=1,2,\ldots\}$ tending to the boundary $\partial\Theta$ of Θ such that $\lim_{l\to\infty}\varphi_n(\theta_l)\le\varphi_n(\theta_0)$. Under this situation, we only need to consider the following two cases.

The first is that $\{\theta_l\}$ has a convergent sub-sequence, i.e. $\theta_{l_k}\to\theta\in\partial\Theta$, as $k\to\infty$, then it follows that:

$$0\le\varphi(\theta)=\lim_{n\to\infty}\lim_{k\to\infty}\varphi_n(\theta_{l_k})\le\lim_{n\to\infty}\varphi_n(\theta_0)=\varphi(\theta_0)=0,$$

hence $\varphi(\theta)=0$. By a similar technique used in the proof of Proposition 6.1, we may get $\theta=\theta_0$, a contradiction.

The second is that $||\theta_l||=\left(\sum_{i=1}^{k}a_{il}^2+\sum_{i=1}^{k-1}m_{il}^2\right)^{1/2}\to\infty$. Then we immediately know there exists a_{il} such that $a_{il}\to\infty$, as $l\to\infty$. Without loss of generality, suppose that

$$\begin{cases} a_{il}\to\infty & (1\le i\le k_1),\\ \sum_{i=1}^{k_1}m_{il}\to m_0 & \\ a_{il}\to a_i<\infty & (k_1+1\le i\le k),\\ m_{il}\to m_i & (k_1+1\le i\le k-1). \end{cases}$$

We have then

$$0\le\lim_{n\to\infty}\lim_{l\to\infty}\varphi_n(\theta_l)\le\lim_{n\to\infty}\varphi_n(\theta_0)=\varphi(\theta_0)=0,$$

and thus

$$\lim_{n\to\infty}\lim_{l\to\infty}\varphi_n(\theta_l)=\sum_{j=1}^{m}\left(z_j-\frac{1-cm_0}{s_j}+c\sum_{i=k_1+1}^{k}\frac{a_im_i}{1+a_is_j}\right)^2=0.$$

If $m_0=0$ then the problem is similar to the first case. Assume $m_0\neq 0$. Denote $\theta_0=(a'_1,\ldots,a'_k,m'_1,\ldots,m'_{k-1})$, we have

$$\frac{m_0}{s_j}+\sum_{i=k_1+1}^{k}\frac{a_im_i}{1+a_is_j}=\sum_{i=1}^{k}\frac{a'_im'_i}{1+a'_is_j},$$

for $j=1,\ldots,m$. Now look s_j as a parameter s and multiplying common factors leads to

$$s\prod_{i_1=k_1+1}^{k}(1+a_{i_1}s)\prod_{i_2=1}^{k}(1+a'_{i_2}s)\left(\frac{m_0}{s}+\sum_{i=k_1+1}^{k}\frac{a_im_i}{1+a_is}\right)=s\prod_{i_1=k_1+1}^{k}(1+a_{i_1}s)\prod_{i_2=1}^{k}(1+a'_{i_2}s)\left(\sum_{i=1}^{k}\frac{a'_im'_i}{1+a'_is}\right).$$

These are polynomials of degree $2k-k_1\le 2k-1$; they coincide at $m\ge 2k-1$ different points $s=s_j$; they are then equal. Comparing their constant terms comes into conflict.

The proof is then complete.

6.3. Proof of Theorem 3.2

The proof of this theorem is similar to the proof of Theorem 3.1. We only present the following proposition.

Proposition 6.2. *If $u_1, \ldots, u_m$ are distinct and $m \geq q$, then $\varphi(\theta) = 0$ for the continuous model has a unique solution θ_0 on Θ.*

Proof. Suppose there is a $\theta = (\alpha_1, \ldots, \alpha_q)$ such that $\varphi(\theta) = 0$. Denote by $\theta_0 = (\alpha'_1, \ldots, \alpha'_k)$ the true value of the parameter. We will show that $\theta = \theta_0$.

Define $p(t, \beta) = \beta_0 + \beta_1 t + , \cdots, + \beta_q t^q$, where $\beta = (\beta_1, \ldots, \beta_q)$ and $\beta_0 = 1 - \sum_{j=1}^{q} j! \beta_j$. We have then

$$\int \frac{t}{1 + t s_j} p(t, \theta^*) e^{-t} dt = 0 \quad (j = 1, \ldots, m),$$

where $\theta^* = \theta - \theta_0$ and $s_j = \underline{s}(u_j)$.

Suppose $p(t, \theta^*) = 0$ has q_0 ($\leq q$) positive real roots $t_1 < \ldots < t_{q_0}$, and denote $t_0 = 0, t_{q_0+1} = +\infty$, then $p(t, \theta^*)$ maintains the sign in each interval (t_{i-1}, t_i) $(i = 1, \ldots, q_0 + 1)$. By mean value theorem, we have

$$0 = \int_0^{+\infty} \frac{t}{1 + t s_j} p(t, \theta^*) e^{-t} dt = \sum_{i=1}^{q_0+1} \frac{\xi_i}{1 + \xi_i s_j} \int_{t_{i-1}}^{t_i} p(t, \theta^*) e^{-t} dt \quad (j = 1, \ldots, m),$$

where $\xi_i \in (t_{i-1}, t_i)$ $(i = 1, \ldots, q_0 + 1)$.

Now look s_j as a parameter s and reduction to common factors leads to

$$0 = \sum_{i=1}^{q_0+1} \prod_{l \neq i} (1 + \xi_l s) \xi_i \int_{t_{i-1}}^{t_i} p(t, \theta^*) e^{-t} dt.$$

The left hand side is a polynomial of degree $q_0 - 1 \leq q - 1$ (the coefficient of $s^{q_0} = \prod_{j=1}^{q_0+1} \xi_j \int_0^{\infty} p(t, \theta^*) e^{-t} dt = 0$); the equation has $m \geq q$ different roots $s = s_j$; the polynomial is then zero. Let $s = -1/\xi_i$ $(i = 1, \ldots, q_0 + 1)$, we get

$$\int_{t_{i-1}}^{t_i} p(t, \theta^*) e^{-t} dt = 0 \quad (i = 1, \ldots, p_0 + 1),$$

which is followed by $p(t, \theta^*) = 0$, and thus $\theta^* = 0$. □

References

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.

Bai, Z.D., Silverstein, J.W., 1998. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. Annals of Probability 26, 316–345.

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

Bai, Z.D., Silverstein, J.W., 2010. Spectral Analysis of Large Dimensional Random Matrices, 2nd ed. Springer, New York.

Bai, Z.D., Yin, Y.Q., 1993. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. Annals of Probability 21, 1275–1294.

Bouchaud, J.P., Potters, M., 2011. Financial Applications of Random Matrix Theory: A Short Review. The Oxford Handbook of Random Matrix Theory. Oxford University Press, Oxford.

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.

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

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

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

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

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., Choi, S.I., 1995. Analysis of the limiting spectral distribution of large-dimensional random matrices. Journal of Multivariate Analysis 54, 295–309.

Szegö, G., 1959. Orthogonal Polynomials (revised ed.). American Mathematical Society, Providence.

Titchmarsh, E.C., 1939. The Theory of Functions, 2nd ed. Oxford University Press, London.

Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R., 1988. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. Probability Theory and Related Fields 78, 509–521.