Working Paper No. 407

Eigenvectors of some large sample covariance matrices ensembles

Olivier Ledoit and Sandrine Péché

March 2009

Eigenvectors of some large sample covariance matrices ensembles

Olivier Ledoit[*]

Institite for Empirical Research in Economics

University of Zurich

CH-8006 Zurich, Switzerland

oledoit@iew.uzh.ch

Sandrine Péché[†]

Institut Fourier

Université Grenoble 1

100 rue des Maths, BP 74

38402 Saint-Martin-d'Hères, France

sandrine.peche@ujf-grenoble.fr

March 24, 2009

Abstract

We consider sample covariance matrices $S_N = \frac{1}{p}\Sigma_N^{1/2} X_N X_N^* \Sigma_N^{1/2}$ where X_N is a $N \times p$ real or complex matrix with i.i.d. entries with finite 12th moment and Σ_N is a $N \times N$ positive definite matrix. In addition as $N \to \infty$ the spectral measure of Σ_N almost surely converges to some limiting probability distribution and $p/N \to \gamma > 0$. We quantify the relationship between sample and population eigen*vectors*, by studying the asymptotics of functionals of the type $\frac{1}{N}\text{Tr}\left(g(\Sigma_N)(S_N - zI)^{-1})\right)$, where I is the identity matrix, g is a bounded function and z is a complex number. This is then used to compute the asymptotically optimal bias correction for sample eigenvalues, paving the way for a new generation of improved estimators of the covariance matrix and its inverse.

1 Introduction and Overview of the Main Results

1.1 Model and results

Consider p independent samples $C_1, \ldots, C_p$, all of which are $N \times 1$ real or complex vectors. In this paper, we are interested in the large-N-limiting spectral properties of the sample covariance matrix

$$S_N = \frac{1}{p} C C^*, \quad C = [C_1, C_2, \ldots, C_p],$$

when we assume that the sample size $p = p(N)$ satisfies $p/N \to \gamma$ as $N \to \infty$ for some $\gamma > 0$. This framework is known as large-dimensional asymptotics. Throughout the paper, $\mathbf{1}$ denotes the indicator function of a set, and we make the following assumptions: $C = \Sigma_N^{1/2} X_N$ where

- (H_1) X_N is a $N \times p$ matrix of real or complex iid random variables with zero mean, unit variance, and 12th absolute central moment bounded by a constant B independent of N and p;

- (H_2) the population covariance matrix Σ_N is a N-dimensional random Hermitian positive definite matrix independent of X_N;

- (H_3) $p/N \to \gamma > 0$ as $N \to \infty$;

- (H_4) $(\tau_1, \ldots, \tau_N)$ is a system of eigenvalues of Σ_N, and the empirical distribution function (e.d.f.) of population eigenvalues given by $H_N(\tau) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{[\tau_j, +\infty)}(\tau)$ converges a.s. to a nonrandom limit $H(\tau)$ at every point of continuity of H. H defines a probability distribution function, whose support $\mathsf{Supp}(H)$ is included in the compact interval $[h_1, h_2]$ with $0 < h_1 \leq h_2 < \infty$.

The aim of this paper is to investigate asymptotic properties of the eigenvectors of such sample covariance matrices. Before exposing our results, we briefly review some known results about spectral properties of sample covariance matrices under large-dimensional asymptotics.

In the whole paper we denote by $((\lambda_1^N, \ldots, \lambda_N^N); (u_1^N, \ldots, u_N^N))$ a system of eigenvalues and orthonormal eigenvectors of the sample covariance matrix $S_N = \frac{1}{p} \Sigma_N^{\frac{1}{2}} X_N X_N^* \Sigma_N^{\frac{1}{2}}$. Without loss of generality, we assume that the eigenvalues are sorted in decreasing order: $\lambda_1^N \geq \lambda_2^N \geq \cdots \geq \lambda_N^N$. Superscripts will be omitted when no confusion is possible. We also denote by $(v_1, \ldots, v_N)$ a system of orthonormal eigenvectors of Σ_N.

First the asymptotic behavior of the eigenvalues is now quite well understood. The "global behavior" of the spectrum of S_N for instance is characterized through the e.d.f., defined as: $F_N(\lambda) = N^{-1} \sum_{i=1}^{N} \mathbf{1}_{[\lambda_i, +\infty)}(\lambda)$, $\forall \lambda \in \mathbb{R}$. The e.d.f. is usually described through its Stieltjes transform. We recall that the Stieltjes transform of a nondecreasing function G is defined by $m_G(z) = \int_{-\infty}^{+\infty} (\lambda - z)^{-1} dG(\lambda)$ for all z in $\mathbb{C}^+$, where $\mathbb{C}^+ = \{z \in \mathbb{C}, \operatorname{Im}(z) > 0\}$. The use of the Stieltjes transform is motivated by the following inversion formula: given any nondecreasing function G, one has that $G(b) - G(a) = \lim_{\eta \to 0^+} \pi^{-1} \int_a^b \operatorname{Im}[m_G(\xi + i\eta)] d\xi$, which holds if G is continuous at a and b.

The first fundamental result concerning the asymptotic global behavior of the spectrum has been obtained by Marčenko and Pastur in [21]. Their result has been later precised e.g. in [4, 14, 16, 28, 29]. In the next Theorem, we recall their result (which was actually proved in a more general setting than that exposed here) and quote the most recent version as given in [27].

Let $m_{F_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} = \frac{1}{N} \operatorname{Tr}\left[(S_N - zI)^{-1}\right]$, where I denotes the $N \times N$ identity matrix.

Theorem 1.1 ([21]). *Under Assumptions* (H_1) *to* (H_4)*, one has that for all* $z \in \mathbb{C}^+$*,* $\lim_{N \to \infty} m_{F_N}(z) = m_F(z)$ *a.s. where*

$$\forall z \in \mathbb{C}^+, \qquad m_F(z) = \int_{-\infty}^{+\infty} \left\{\tau \left[1 - \gamma^{-1} - \gamma^{-1} z \, m_F(z)\right] - z\right\}^{-1} dH(\tau). \qquad (1.1)$$

Furthermore, the e.d.f. of the eigenvalues of the sample covariance matrix given by $F_N(\lambda) = N^{-1} \sum_{i=1}^{N} \mathbf{1}_{[\lambda_i, +\infty)}(\lambda)$ *converges a.s. to the nonrandom limit* $F(\lambda)$ *at all points of continuity of* F*.*

In addition, [11] show that the following limit exists :

$$\forall \lambda \in \mathbb{R} - \{0\}, \lim_{z \in \mathbb{C}^+ \to \lambda} m_F(z) \equiv \check{m}_F(\lambda). \qquad (1.2)$$

They also prove that F has a continuous derivative which is given by $F' = \pi^{-1} \operatorname{Im}[\check{m}_F]$ on $(0, +\infty)$. More precisely, when $\gamma > 1$, $\lim_{z \in \mathbb{C}^+ \to \lambda} m_F(z) \equiv \check{m}_F(\lambda)$ exists for *all* $\lambda \in \mathbb{R}$, F has a continuous derivative F' on *all* of $\mathbb{R}$, and $F(\lambda)$ is identically equal to zero in a neighborhood of $\lambda = 0$. When $\gamma < 1$, the proportion of sample eigenvalues equal to zero is asymptotically $1 - \gamma$. In this case, it is convenient to introduce the e.d.f. $\underline{F} = \left(1 - \gamma^{-1}\right) \mathbf{1}_{[0, +\infty)} + \gamma^{-1} F$,

4

which is the limit of e.d.f. of the eigenvalues of the p-dimensional matrix $p^{-1}X_N^*\Sigma_N X_N$. Then $\lim_{z \in \mathbb{C}^+ \to \lambda} m_{\underline{F}}(z) \equiv \breve{m}_{\underline{F}}(\lambda)$ exists for *all* $\lambda \in \mathbb{R}$, $\underline{F}$ has a continuous derivative $\underline{F}'$ on *all* of $\mathbb{R}$, and $\underline{F}(\lambda)$ is identically equal to zero in a neighborhood of $\lambda = 0$. When γ is exactly equal to one, further complications arise because the density of sample eigenvalues can be unbounded in a neighborhood of zero; for this reason we sometimes have to rule out the possibility that $\gamma = 1$.

Further studies have complemented the a.s. convergence established by the Marčenko-Pastur theorem (see e.g. [1, 5, 6, 7, 9, 15, 22] and [2] for more references). The Marčenko-Pastur equation has also generated a considerable amount of interest in statistics [13, 19], finance [18, 17], signal processing [12], and other disciplines. We refer the interested reader to the recent book by Bai and Silverstein [8] for a throrough survey of this fast-growing field of research.

As we can gather from this brief review of the literature, the Marčenko-Pastur equation reveals much of the behavior of the *eigenvalues* of sample covariance matrices under large-dimensional asymptotics. It is also of utmost interest to describe the asymptotic behavior of the *eigenvectors*. Such an issue is fundamental to statistics (for instance both eigenvalues and eigenvectors are of interest in Principal Components Analysis), communication theory, wireless communication, finance. The reader is referred to [3], Section 1 for more detail and to [10] for a statistical approach to the problem and a detailed exposition of statistical applications.

Actually much less is known about eigenvectors of sample covariance matrices. In the special case where $\Sigma = I$ and the X_{ij} are i.i.d. standard (real or complex) Gaussian random variables, it is well-known that the matrix of sample eigenvectors is Haar distributed (on the orthogonal or unitary group). To our knowledge, these are the only ensembles for which the distribution of the eigenvectors is explicitly known. It has been conjectured that for a wide class of non Gaussian ensembles, the matrix of sample eigenvectors should be "asymptotically Haar distributed", provided $\Sigma = I$. Note that the notion "asymptotically Haar distributed" needs to be defined. This question has been investigated by [24], [25], [26] followed by [3]. Therein a random matrix U is said to be asymptotically Haar distributed if Ux is asymptotically uniformly distributed on the unit sphere for any non random unit vector x. [26] and [3] are then able to prove the conjecture under various sets of assumptions on the X_{ij}'s.

In the case where $\Sigma \neq I$, much less is known (see [3]). One expects that the distribution of

5

the eigenvectors is far from being rotation-invariant. This is precisely the aspect with which this paper is concerned.

In this paper, we present another approach to study eigenvectors of sample covariance matrices. Roughly speaking, we study "functionals" of the type

$$\forall z \in \mathbb{C}^+, \qquad \Theta_N^g(z) \;=\; \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} \sum_{j=1}^N |u_i^* v_j|^2 \times g(\tau_j) \tag{1.3}$$

$$=\; \frac{1}{N} \mathsf{Tr} \left[(S_N - zI)^{-1} g(\Sigma_N) \right],$$

where g is any real-valued univariate function satisfying suitable regularity conditions. By convention, $g(\Sigma_N)$ is the matrix with the same eigenvectors as Σ_N and with eigenvalues $g(\tau_1), \ldots, g(\tau_N)$. These functionals are generalizations of the Stieltjes transform used in the Marčenko-Pastur equation. Indeed, one can rewrite the Stieltjes transform of the e.d.f. of sample eigenvalues as:

$$\forall z \in \mathbb{C}^+, \qquad m_{F_N}(z) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} \sum_{j=1}^N |u_i^* v_j|^2 \times 1. \tag{1.4}$$

The constant 1 that appears at the end of Equation (1.4) can be interpreted as a weighting scheme placed on the population eigenvectors: specifically, it represents a flat weighting scheme. The generalization we here introduce puts the spotlight on how the sample covariance matrix relates to the population covariance matrix, or even *any function of* the population covariance matrix.

Our main result is given in the following Theorem.

Theorem 1.2. *Assume that conditions* $(H_1) - (H_4)$ *are satisfied. Let g be a (real-valued) bounded function defined on $[h_1, h_2]$ with finitely many points of discontinuity. Then there exists a nonrandom function Θ^g defined over $\mathbb{C}^+$ such that $\Theta_N^g(z) = N^{-1} \mathsf{Tr} \left[(S_N - zI)^{-1} g(\Sigma_N) \right]$ converges a.s. to $\Theta^g(z)$ for all $z \in \mathbb{C}^+$. Furthermore, Θ^g is given by:*

$$\forall z \in \mathbb{C}^+, \; \Theta^g(z) = \int_{-\infty}^{+\infty} \left\{ \tau \left[1 - \gamma^{-1} - \gamma^{-1} z m_F(z) \right] - z \right\}^{-1} g(\tau) dH(\tau). \tag{1.5}$$

One can first observe that as we move from a flat weighting scheme of $g \equiv 1$ to any arbitrary weighting scheme $g(\tau_j)$, the integration kernel $\left\{ \tau \left[1 - \frac{1}{\gamma} - \frac{z \, m_F(z)}{\gamma} \right] - z \right\}^{-1}$ remains

unchanged. Therefore, our Equation (1.5) generalizes Marčenko and Pastur's foundational result. Actually the proof of Theorem 1.2 follows from some of the arguments used in [27] to derive the Marchenko-Pastur equation. This proof is postponed until Section 2.

The generalization of the Marčenko-Pastur equation we propose allows to consider a few unsolved problems regarding the overall relationship between sample and population covariance matrices. Let us consider two of these problems, which are investigated in more detail in the two next subsections.

The first of these questions is: how do the eigenvectors of the sample covariance matrix deviate from those of the population covariance matrix? By injecting functions g of the form $\mathbf{1}_{(-\infty,\tau)}$ into Equation (1.5), we quantify the asymptotic relationship between sample and population eigenvectors. This is developed in more detail in Section 1.2.

Another question is: how does the sample covariance matrix deviate from the population covariance matrix as a whole, and how can we modify it to bring it closer to the population covariance matrix? This is an important question in Statistics, where a covariance matrix estimator that improves upon the sample covariance matrix is sought. By injecting the function $g(\tau) = \tau$ into Equation (1.5), we find the optimal asymptotic bias correction for the eigenvalues of the sample covariance matrix in Section 1.3. We also perform the same calculation for the *inverse* covariance matrix (an object of great interest in Econometrics and Finance), this time by taking $g(\tau) = 1/\tau$.

This list is not intended to be exhaustive. Other applications may hopefully be extracted from our generalized Marčenko-Pastur equation.

1.2 Sample vs. Population Eigenvectors

As will be made more apparent in Equation (1.8) below, it is possible to quantify the asymptotic behavior of sample eigenvectors in the general case by selecting a function g of the form $\mathbf{1}_{(-\infty,\tau)}$ in Equation (1.5). Let us briefly explain why.

First of all, note that each sample eigenvector u_i lies in a space whose dimension is growing towards infinity. Therefore, the only way to know "where" it lies is to project it onto a known orthonormal basis that will serve as a reference grid. Given the nature of the problem, the most meaningful choice for this reference grid is the orthonormal basis formed by the population eigenvectors $(v_1, \ldots, v_N)$. Thus we are faced with the task of characterizing the asymptotic

7

behavior of $u_i^* v_j$ for all $i, j = 1, \ldots, N$, i.e. the projection of the sample eigenvectors onto the population eigenvectors. Yet as every eigenvector is identified up to multiplication by a scalar of modulus one, the argument (angle) of $u_i^* v_j$ is devoid of mathematical relevance. Therefore, we can focus instead on its square modulus $|u_i^* v_j|^2$ without loss of information.

Another issue that arises is that of scaling. Indeed as

$$\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |u_i^* v_j|^2 = \frac{1}{N^2} \sum_{i=1}^{N} u_i^* \left(\sum_{j=1}^{N} v_j v_j^* \right) u_i = \frac{1}{N^2} \sum_{i=1}^{N} u_i^* u_i = \frac{1}{N},$$

we study $N |u_i^* v_j|^2$ instead, so that its limit does not vanish under large-N asymptotics. The indexing of the eigenvectors also demands special attention as the dimension goes to infinity. We choose to use an indexation system where "eigenvalues serve as labels for eigenvectors", that is u_i is the eigenvector associated to the ith largest eigenvalue λ_i.

All these considerations lead us to introduce the following key object:

$$\forall \lambda, \tau \in \mathbb{R}, \qquad \Phi_N(\lambda, \tau) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |u_i^* v_j|^2 \, \mathbf{1}_{[\lambda_i, +\infty)}(\lambda) \times \mathbf{1}_{[\tau_j, +\infty)}(\tau). \qquad (1.6)$$

This bivariate function is right continuous with left-hand limits and nondecreasing in each of its arguments. It also verifies $\lim_{\substack{\lambda \to -\infty \\ \tau \to -\infty}} \Phi_N(\lambda, \tau) = 0$ and $\lim_{\substack{\lambda \to +\infty \\ \tau \to +\infty}} \Phi_N(\lambda, \tau) = 1$. Therefore, it satisfies the properties of a bivariate cumulative distribution function.

Remark 1. *Our function Φ_N can be compared with the object introduced in [3]: $\forall \lambda \in \mathbb{R}$, $F_1^{S_N}(\lambda) = \sum_{i=1}^{N} |u_i^* x_N|^2 \, \mathbf{1}_{[\lambda_i, +\infty)}(\lambda)$, where $(x_N)_{N=1,2,\ldots}$ is a sequence of nonrandom unit vectors satisfying the non-trivial condition $x_N^* (\Sigma_N - zI)^{-1} x_N \to m_H(z)$. This condition is specified so that projecting the sample eigenvectors onto x_N effectively wipes out any signature of non-rotation-invariant behavior. The main difference is that Φ_N projects the sample eigenvectors onto the population eigenvectors instead.*

From Φ_N we can extract precise information about the sample eigenvectors. The average of the quantities of interest $N |u_i^* v_j|^2$ over the sample (resp. population) eigenvectors associated with the sample (resp. population) eigenvalues lying in the interval $[\underline{\lambda}, \overline{\lambda}]$ (resp. $[\underline{\tau}, \overline{\tau}]$) is equal

to:

$$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} N \, |u_i^* v_j|^2 \, \mathbf{1}_{[\underline{\lambda},\overline{\lambda}]}(\lambda_i) \times \mathbf{1}_{[\underline{\tau},\overline{\tau}]}(\tau_j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{1}_{[\underline{\lambda},\overline{\lambda}]}(\lambda_i) \times \mathbf{1}_{[\underline{\tau},\overline{\tau}]}(\tau_j)}$$

$$= \frac{\Phi_N(\overline{\lambda},\overline{\tau}) - \Phi_N(\overline{\lambda},\underline{\tau}) - \Phi_N(\underline{\lambda},\overline{\tau}) + \Phi_N(\underline{\lambda},\underline{\tau})}{[F_N(\overline{\lambda}) - F_N(\underline{\lambda})] \times [H_N(\overline{\tau}) - H_N(\underline{\tau})]}, \tag{1.7}$$

whenever the denominator is strictly positive. Since $\underline{\lambda}$ and $\overline{\lambda}$ (resp. $\underline{\tau}$ and $\overline{\tau}$) can be chosen arbitrarily close to each other (as long as the average in Equation (1.7) exists), our goal of characterizing the behavior of sample eigenvectors would be achieved in principle by determining the asymptotic behavior of Φ_N. This can be deduced from Theorem 1.2 thanks to the inversion formula for the Stieltjes transform: for all $(\lambda, \tau) \in \mathbb{R}^2$ such that Φ_N is continuous at (λ, τ)

$$\Phi_N(\lambda, \tau) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{\lambda} \mathsf{Im}\left[\Theta_N^g(\xi + i\eta)\right] d\xi, \tag{1.8}$$

which holds in the special case where $g = \mathbf{1}_{(-\infty, \tau)}$. We are now ready to state our second main result.

Theorem 1.3. *Assume that conditions* $(H_1) - (H_4)$ *hold true and let* $\Phi_N(\lambda, \tau)$ *be defined by (1.6). Then there exists a nonrandom bivariate function* Φ *such that* $\Phi_N(\lambda, \tau) \xrightarrow{a.s.} \Phi(\lambda, \tau)$ *at all points of continuity of* Φ. *Furthermore, when* $\gamma \neq 1$, *the function* Φ *can be expressed as:* $\forall (\lambda, \tau) \in \mathbb{R}^2, \quad \Phi(\lambda, \tau) = \int_{-\infty}^{\lambda} \int_{-\infty}^{\tau} \varphi(l, t) \, dH(t) \, dF(l)$, *where*

$$\forall (l, t) \in \mathbb{R}^2 \quad \varphi(l, t) = \begin{cases} \dfrac{\gamma^{-1} l t}{(at - l)^2 + b^2 t^2} & \text{if } l > 0 \\ \dfrac{1}{(1 - \gamma)[1 + \check{m}_{\underline{F}}(0)\, t]} & \text{if } l = 0 \text{ and } \gamma < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{1.9}$$

and a *(resp.* b*) is the real (resp. imaginary) part of* $1 - \gamma^{-1} - \gamma^{-1} l \, \check{m}_F(l)$.

Equation (1.9) quantifies how the eigenvectors of the sample covariance matrix deviate from those of the population covariance matrix under large-dimensional asymptotics. The result is explicit as a function of m_F.

To illustrate Theorem 1.3, we can pick any eigenvector of our choosing, for example the one that corresponds to the first (i.e. largest) eigenvalue, and plot how it projects onto the population eigenvectors (indexed by their corresponding eigenvalues). The resulting graph is
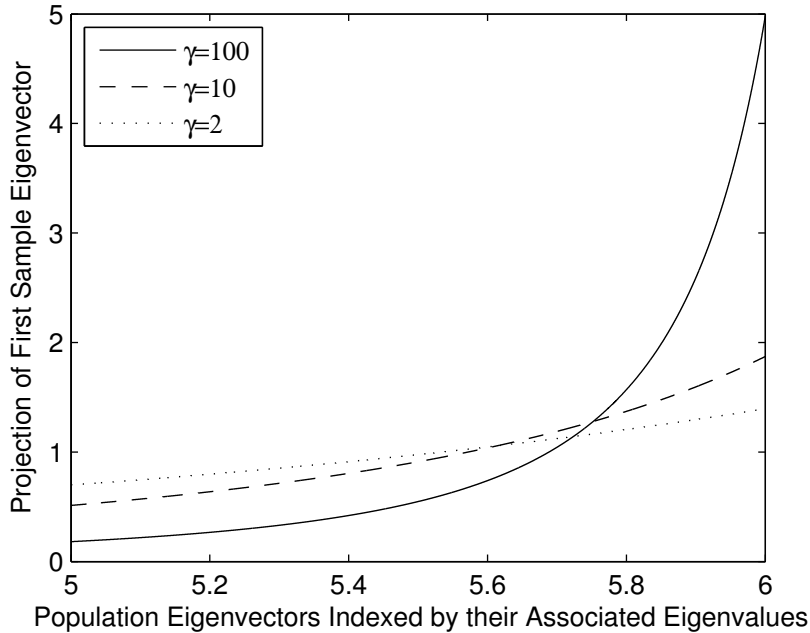
Figure 1: Projection of first sample eigenvector onto population eigenvectors (indexed by their associated eigenvalues). We have taken $H' = \mathbf{1}_{[5,6]}$.

shown in Figure 1. This is a plot of $\varphi(l, t)$ as a function of t, for fixed l equal to the supremum of $\mathsf{Supp}(F)$. It is the asymptotic equivalent to plotting $N|u_1^* v_j|^2$ as a function of τ_j. It looks like a density because, by construction, it must integrate to one. As soon as the sample size is of the order of 10 times the number of variables, we can see that the first sample eigenvector starts deviating quite strongly from the first population eigenvectors. This should have precautionary implications for Principal Component Analysis (PCA), where the number of variables is often so large that it is difficult to make the sample size more than ten times bigger.

Obviously, Equation (1.9) would enable us to draw a similar graph for any sample eigenvector (not just the first one), and for any γ and H verifying the assumptions of Theorem 1.3. Preliminary investigations reveal some unexpected patterns. For example: one might have thought that the sample eigenvector associated with the median sample eigenvalue would be closest to the population eigenvector associated with the median population eigenvalue; but in general this is not true.

10

1.3 Asymptotically Optimal Bias Correction for the Sample Eigenvalues

We now bring the two preceding results together to quantify the relationship between the sample covariance matrix and the population covariance matrix *as a whole.* As will be made clear in Equation (1.12) below, this is achieved by selecting the function $g(\tau) = \tau$ in Equation (1.5). The objective is to see how the sample covariance matrix deviates from the population covariance matrix, and how we can modify it to bring it closer to the population covariance matrix. The main problem with the sample covariance matrix is that its eigenvalues are too dispersed: the smallest ones are biased downwards, and the largest ones upwards. This is most easily visualized when the population covariance matrix is the identity, in which case the limiting spectral e.d.f. of sample eigenvalues F is known in closed form (see Figure 2). We
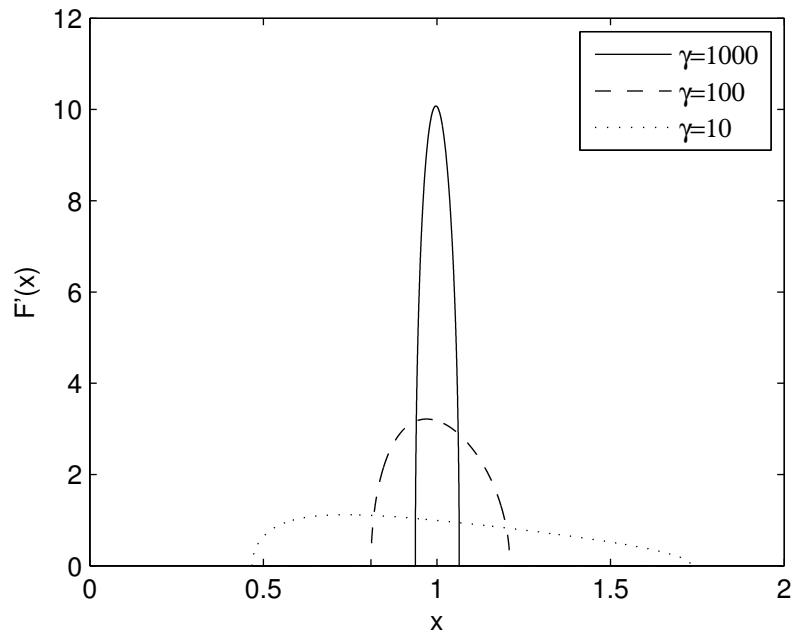

Figure 2: Limiting density of sample eigenvalues, in the particular case where all the eigenvalues of the population covariance matrix are equal to one. The graph shows excess dispersion of the sample eigenvalues. The formula for this plot comes from solving the Marčenko-Pastur equation for $H = \mathbf{1}_{[1,+\infty)}$.

can see that the smallest and the largest sample eigenvalues are biased away from one, and

that the bias decreases in γ. Therefore, a key concern in multivariate statistics is to find the asymptotically optimal bias correction for the eigenvalues of the sample covariance matrix. As this correction will tend to reduce the dispersion of the eigenvalues, it is often called a *shrinkage* formula.

Ledoit and Wolf [20] made some progress along this direction by finding the optimal *linear* shrinkage formula for the sample eigenvalues (projecting Σ_N on the two-dimensional subspace spanned by S_N and I). However, shrinking the eigenvalues is a highly nonlinear problem (as Figure 3 below will illustrate). Therefore, there is strong reason to believe that finding the optimal *nonlinear* shrinkage formula for the sample eigenvalues would lead to a covariance matrix estimator that further improves upon the Ledoit-Wolf estimator. Theorem 1.2 paves the way for such a development.

To see how, let us think of the problem of estimating Σ_N in general terms. In order to construct an estimator of Σ_N, we must in turn consider what the eigenvectors and the eigenvalues of this estimator should be. Let us consider the eigenvectors first. In the general case where we have no prior information about the orientation of the population eigenvectors, it is reasonable to require that the estimation procedure be invariant with respect to rotation by any p-dimensional orthogonal matrix W. If we rotate the variables by W, then we would ask our estimator to also rotate by the same orthogonal matrix W. The class of orthogonally invariant estimators of the covariance matrix is constituted of all the estimators that have the same eigenvectors as the sample covariance matrix (see [23], Lemma 5.3). Every rotation-invariant estimator of Σ_N is thus of the form:

$$U_N D_N U_N^*, \quad \text{where} \quad D_N = \mathsf{Diag}(d_1, \ldots, d_N) \text{ is diagonal,}$$

and where U_N is the matrix whose i^{th} column is the sample eigenvector u_i. This is the class that we consider.

Our objective is to find the matrix in this class that is closest to the population covariance matrix. In order to measure distance, we choose the Frobenius norm, defined as: $\|A\|_F = \sqrt{\mathsf{Tr}\,(AA^*)}$ for any matrix A. Thus we end up with the following optimization problem: $\min_{D_N \text{ diagonal}} \|U_N D_N U_N^* - \Sigma_N\|_F$. Elementary matrix algebra shows that its solution is:

$$\widetilde{D}_N = \mathsf{Diag}(\widetilde{d}_1, \ldots, \widetilde{d}_N) \quad \text{where} \quad \forall i = 1, \ldots, N \quad \widetilde{d}_i = u_i^* \, \Sigma_N \, u_i.$$

The interpretation of $\widetilde{d}_i$ is that it captures how the ith sample eigenvector u_i relates to the

12

population covariance matrix Σ_N as a whole.

While $U_N \widetilde{D}_N U_N^*$ does not constitute a *bona fide* estimator (because it depends on the unobservable Σ_N), new estimators that seek to improve upon the existing ones will need to get as close to $U_N \widetilde{D}_N U_N^*$ as possible. This is exactly the path that led Ledoit and Wolf [20] to their improved covariance matrix estimator. Therefore, it is important, in the interest of developing a new and improved estimator, to characterize the asymptotic behavior of $\widetilde{d}_i$ $(i = 1, \ldots, N)$. The key object that will enable us to achieve this goal is the nondecreasing function defined by:

$$\forall x \in \mathbb{R}, \quad \Delta_N(x) = \frac{1}{N} \sum_{i=1}^{N} \widetilde{d}_i \, \mathbf{1}_{[\lambda_i, +\infty)}(x) = \frac{1}{N} \sum_{i=1}^{N} u_i^* \Sigma_N u_i \times \mathbf{1}_{[\lambda_i, +\infty)}(x). \tag{1.10}$$

When all the sample eigenvalues are distinct, it is straightforward to recover the $\widetilde{d}_i$'s from Δ_N:

$$\forall i = 1, \ldots, N \qquad \widetilde{d}_i = \lim_{\varepsilon \to 0^+} \frac{\Delta_N(\lambda_i + \varepsilon) - \Delta_N(\lambda_i - \varepsilon)}{F_N(\lambda_i + \varepsilon) - F_N(\lambda_i - \varepsilon)}. \tag{1.11}$$

The asymptotic behavior of Δ_N can be deduced from Theorem 1.2 in the special case where $g(\tau) = \tau$: for all $x \in \mathbb{R}$ such that Δ_N continuous at x

$$\Delta_N(x) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \mathsf{Im} \left[\Theta_N^g(\xi + i\eta) \right] d\xi, \quad g(x) \equiv x. \tag{1.12}$$

We are now ready to state our third main result.

Theorem 1.4. *Assume that conditions* $(H_1) - (H_4)$ *hold true and let* Δ_N *be defined as in (1.10). There exists a nonrandom function* Δ *defined over* $\mathbb{R}$ *such that* $\Delta_N(x)$ *converges a.s. to* $\Delta(x)$ *for all* $x \in \mathbb{R} - \{0\}$. *If in addition* $\gamma \neq 1$, *then* Δ *can be expressed as:* $\forall x \in \mathbb{R}, \quad \Delta(x) = \int_{-\infty}^{x} \delta(\lambda) \, dF(\lambda)$, *where*

$$\forall \lambda \in \mathbb{R}, \qquad \delta(\lambda) = \begin{cases} \dfrac{\lambda}{|1 - \gamma^{-1} - \gamma^{-1}\lambda \, \breve{m}_F(\lambda)|^2} & \text{if } \lambda > 0 \\ \dfrac{\gamma}{(1 - \gamma) \, \breve{m}_{\underline{F}}(0)} & \text{if } \lambda = 0 \text{ and } \gamma < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.13}$$

By Equation (1.11) the asymptotic quantity that corresponds to $\widetilde{d}_i = u_i^* \Sigma_N u_i$ is $\delta(\lambda)$, provided that λ corresponds to λ_i. Therefore, the way to get closest to the population covariance matrix (according to the Frobenius norm) would be to divide each sample eigenvalue λ_i by the correction factor $|1 - \gamma^{-1} - \gamma^{-1}\lambda \, \breve{m}_F(\lambda_i)|^2$. This is what we call the optimal nonlinear shrinkage
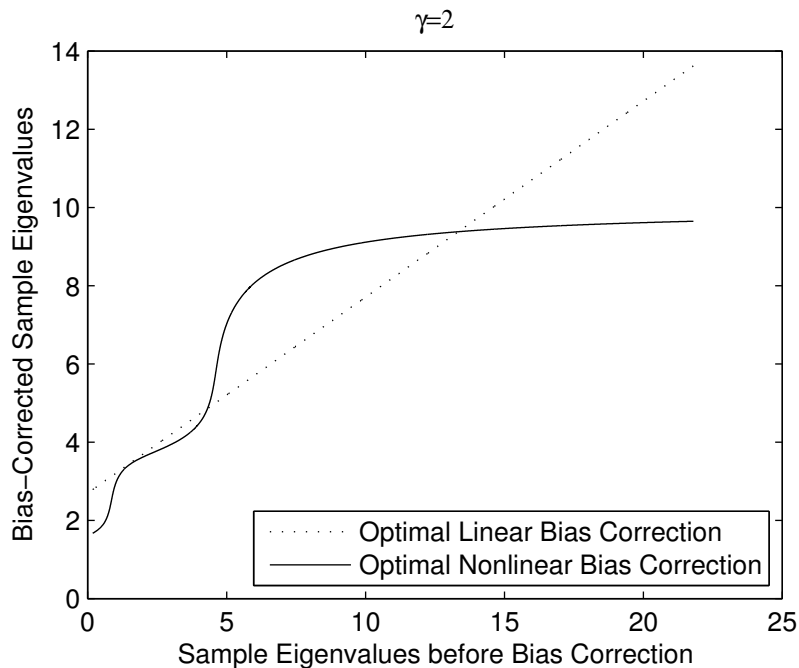
13

Figure 3: Comparison of the Optimal Linear vs. Nonlinear Bias Correction Formulæ. In this example, the distribution of population eigenvalues H places 20% mass at 1, 40% mass at 3 and 40% mass at 10. The solid line plots $\delta(\lambda)$ as a function of λ.

formula or asymptotically optimal bias correction. Figure 3 shows how much it differs from Ledoit and Wolf's [20] optimal linear shrinkage formula. In addition, when $\gamma < 1$, the sample eigenvalues equal to zero need to be replaced by $\delta(0) = \gamma/[(1 - \gamma) \, \breve{m}_{\underline{F}}(0)]$.

In a statistical context of estimation, $\breve{m}_F(\lambda_i)$ and $\breve{m}_{\underline{F}}(0)$ are not known, so they need to be replaced by $\breve{m}_{\widehat{F}}(\lambda_i)$ and $\breve{m}_{\underline{\widehat{F}}}(0)$ respectively, where $\widehat{F}$ is some estimator of the limiting p.d.f. of sample eigenvalues. Research is currently underway to prove that a covariance matrix estimator constructed in this manner has desirable properties under large-dimensional asymptotics.

Monte-Carlo simulations indicate that applying this bias correction is highly beneficial, even in small samples. We ran 10,000 simulations based on the distribution of population eigenvalues H that places 20% mass at 1, 40% mass at 3 and 40% mass at 10. We kept γ constant at 2 while increasing the number of variables from 5 to 100. For each set of simulations, we computed the Percentage Relative Improvement in Average Loss (PRIAL). The PRIAL of

14

an estimator M of Σ_N is defined as

$$PRIAL(M) = 100 \times \left[1 - \frac{\mathbb{E} \left\| M - U_N \widetilde{D}_N U_N^* \right\|_F^2}{\mathbb{E} \left\| S_N - U_N \widetilde{D}_N U_N^* \right\|_F^2} \right].$$

By construction, the PRIAL of the sample covariance matrix S_N (resp. of $U_N \widetilde{D}_N U_N^*$) is 0% (resp. 100%), meaning no improvement (resp. meaning maximum attainable improvement). For each of the 10,000 Monte-Carlo simulations, we consider $\widetilde{S}_N$, which is the matrix obtained from the sample covariance matrix by keeping its eigenvectors and dividing its ith eigenvalue by the correction factor $|1 - \gamma^{-1} - \gamma^{-1}\lambda_i \, \breve{m}_F(\lambda_i)|^2$. The expected loss $\mathbb{E} \left\| \widetilde{S}_N - U_N \widetilde{D}_N U_N^* \right\|_F^2$ is estimated by computing its average across the 10,000 Monte-Carlo simulations. Figure 4 plots the PRIAL obtained in this way, that is by applying the optimal nonlinear shrinkage formula to the sample eigenvalues. We can see that, even with a modest sample size like $p = 40$, we already get 95% of the maximum possible improvement.

A similar formula can be obtained for the purpose of estimating the *inverse* of the population covariance matrix. To this aim, we set $g(\tau) = 1/\tau$ in Equation (1.5) and define

$$\Psi_N(x) := N^{-1} \sum_{i=1}^{N} u_i^* \Sigma_N^{-1} u_i \times \mathbf{1}_{[\lambda_i, +\infty)}(x), \, \forall x \in \mathbb{R}.$$

Theorem 1.5. *Assume that conditions* $(H_1) - (H_4)$ *are satisfied. There exists a nonrandom function Ψ defined over $\mathbb{R}$, such that $\Psi_N(x)$ converges a.s. to $\Psi(x)$ for all $x \in \mathbb{R} - \{0\}$. If in addition $\gamma \neq 1$, then Ψ can be expressed as:* $\forall x \in \mathbb{R}, \quad \Psi(x) = \int_{-\infty}^{x} \psi(\lambda) \, dF(\lambda)$, *where*

$$\forall \lambda \in \mathbb{R} \quad \psi(\lambda) = \begin{cases} \dfrac{1 - \gamma^{-1} - 2\gamma^{-1}\lambda \, \mathsf{Re}\left[\breve{m}_F(\lambda)\right]}{\lambda} & \text{if } \lambda > 0 \\ \dfrac{1}{1 - \gamma} \breve{m}_H(0) - \breve{m}_{\underline{F}}(0) & \text{if } \lambda = 0 \text{ and } \gamma < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.14}$$

Therefore, the way to get closest to the inverse of the population covariance matrix (according to the Frobenius norm) would be to multiply the inverse of each sample eigenvalue λ_i^{-1} by the correction factor $1 - \gamma^{-1} - 2\gamma^{-1}\lambda_i \, \mathsf{Re}[\breve{m}_F(\lambda_i)]$. This represents the optimal nonlinear shrinkage formula (or asymptotically optimal bias correction) for the purpose of estimating the inverse covariance matrix. Again, in a statistical context of estimation, the unknown $\breve{m}_F(\lambda_i)$ needs to be replaced by $\breve{m}_{\widehat{F}}(\lambda_i)$, where $\widehat{F}$ is some estimator of the limiting p.d.f. of sample eigenvalues. This question is investigated in some work under progress.
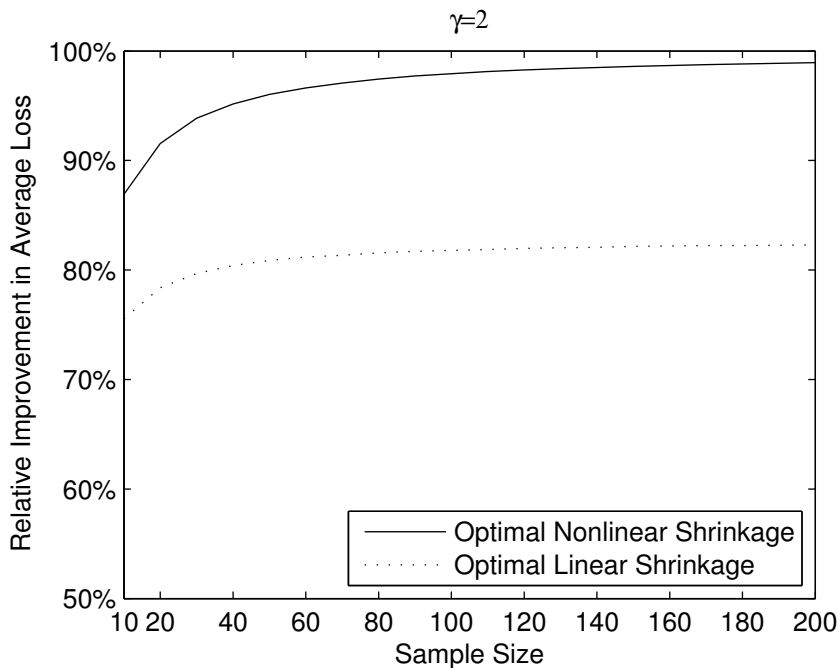
Figure 4: Percentage Relative Improvement in Average Loss (PRIAL) from applying the optimal nonlinear shrinkage formula to the sample eigenvalues. The solid line shows the PRIAL obtained by dividing the ith sample eigenvalue by the correction factor $|1-\gamma^{-1}-\gamma^{-1}\lambda_i \, \breve{m}_F(\lambda_i)|^2$, as a function of sample size. The dotted line shows the PRIAL of the Ledoit-Wolf [20] linear shrinkage estimator. For each sample size we ran 10,000 Monte-Carlo simulations. Like in Figure 3, we used $\gamma = 2$ and the distribution of population eigenvalues H placing 20% mass at 1, 40% mass at 3 and 40% mass at 10.

The rest of the paper is organized as follows. Section 2 contains the proof of Theorem 1.2. Section 3 contains the proof of Theorem 1.3. Section 4 is devoted to the proofs of Theorems 1.4 and 1.5.

2 Proof of Theorem 1.2

First we need to adapt a Lemma from Bai and Silverstein [4].

Lemma 2.1. *Let* $Y = (y_1, \ldots, y_N)$ *be a random vector with i.i.d. entries satisfying:*

$$\mathbb{E}y_1 = 0, \quad \mathbb{E}|y_1|^2 = 1, \quad \mathbb{E}|y_1|^{12} \leq B,$$

16

where the constant B does not depend on N. Let also A be a given $N \times N$ matrix. Then there exists a constant $K > 0$ independent of N, A and Y such that:

$$\mathbb{E} \, |Y A Y^* - \mathsf{Tr}(A)|^6 \leq K \|A\|^6 N^3.$$

Proof of Lemma 2.1 The proof of Lemma 2.1 directly follows from that of Lemma 3.1 in Silverstein and Bai (1995). Therein the assumption that $\mathbb{E}|y_1|^{12} \leq B$ is replaced with the assumption that $|y_1| \leq \ln N$. One can easily check that all their arguments carry through if one assumes that the twelfth moment of y_1 is uniformly bounded. $\square$

Next, we need to introduce some notation. We set $R_N(z) = (S_N - zI)^{-1}$ and define $\Theta_N^{(k)}(z) = N^{-1}\mathsf{Tr}[R_N(z)\Sigma^k]$ for all $z \in \mathbb{C}^+$ and integer k. Thus, $\Theta_N^{(k)} = \Theta_N^g$ if we take $g(\tau) = \tau^k$, $\forall \tau \in \mathbb{R}$. In particular, $\Theta_N^{(0)} = m_{F_N}$. To avoid confusion, the dependency of most of the variables on N will occasionally be dropped from the notation. All convergence statements will be as $N \to \infty$. Conditions $(H_1) - (H_4)$ are assumed to hold throughout.

Lemma 2.2. *One has that* $\forall z \in \mathbb{C}^+$, $\quad \Theta_N^{(1)}(z) \xrightarrow{\text{a.s.}} \Theta^{(1)}(z)$ *where:*

$$\Theta^{(1)}(z) = \frac{\gamma^2}{\gamma - 1 - z \, m_F(z)} - \gamma.$$

Proof of Lemma 2.2 In the first part of the proof, we show that

$$1 + z m_{F_N}(z) = \frac{p}{N} - \frac{1}{N} \sum_{k=1}^{p} \frac{1}{1 + (N/p)\Theta_N^{(1)}(z)} + o(1).$$

Using the a.s. convergence of the Stieltjes transform $m_{F_N}(z)$, it is then easy to deduce the equation satisfied by $\Theta^{(1)}$ in Lemma 2.2. Our proof closely follows some of the ideas of [27] and [4]. Therein the convergence of the Stieltjes transform $m_{F_N}(z)$ is investigated.

Let us define $C_k = p^{-1/2}\sqrt{\Sigma}\, X_k$, where X_k is the kth column of X. Then $S_N = \sum_{k=1}^{p} C_k C_k^*$. Using the identity $S_N - zI + zI = \sum_{k=1}^{p} C_k C_k^*$, one deduces that

$$\frac{1}{N}\mathsf{Tr}(I + zR_N(z)) = \frac{1}{N} \sum_{k=1}^{p} C_k^* R_N(z) C_k. \tag{2.1}$$

Define now for any integer $1 \leq k \leq p$

$$R_N^{(k)}(z) := (S_N - C_k C_k^* - zI)^{-1}.$$

By the resolvent identity $R_N(z) - R_N^{(k)}(z) = -R_N(z)C_kC_k^*R_N^{(k)}(z)$, we deduce that

$$C_k^*R_N(z)C_k - C_k^*R_N^{(k)}(z)C_k = -C_k^*R_N(z)C_kC_k^*R_N^{(k)}(z)C_k,$$

which finally gives that

$$C_k^*R_N(z)C_k = \frac{1}{1 + C_k^*R_N^{(k)}(z)C_k}C_k^*R_N^{(k)}(z)C_k.$$

Plugging the latter formula into (2.1), one can write that

$$1 + zm_{F_N}(z) = \frac{p}{N} - \frac{1}{N}\sum_{k=1}^{p}\frac{1}{1 + C_k^*R_N^{(k)}(z)C_k}. \tag{2.2}$$

We will now use the fact that $R_N^{(k)}$ and C_k are independent random matrices to estimate the asymptotic behavior of the last sum in (2.2). Using Lemma 2.1, we deduce that

$$\max_{k\in\{1,\ldots,p\}}\left|C_k^*R_N^{(k)}C_k - \frac{1}{p}\mathsf{Tr}\left(R_N^{(k)}\Sigma\right)\right| \overset{\text{a.s.}}{\to} 0, \tag{2.3}$$

as $N \to \infty$. Furthermore, using Lemma 2.6 in Silverstein and Bai (1995), one also has that

$$\frac{1}{p}\left|\mathsf{Tr}\left[\left(R_N - R_N^{(k)}\right)\Sigma\right]\right| \leq \frac{||\Sigma||}{py}. \tag{2.4}$$

Thus using (2.4), (2.3) and (2.2), one can write that

$$1 + zm_{F_N}(z) = \frac{p}{N} - \frac{1}{N}\sum_{k=1}^{p}\frac{1}{1 + (N/p)\Theta_N^{(1)}(z)} + \delta_N, \tag{2.5}$$

where the error term δ_N is given by $\delta_N = \delta_N^1 + \delta_N^2$ with

$$\delta_N^1 = \frac{1}{N}\sum_{k=1}^{p}\frac{\frac{1}{p}\mathsf{Tr}\left((R_N - R_N^{(k)})\Sigma\right)}{(1 + \frac{1}{p}\mathsf{Tr}(R_N\Sigma))(1 + \frac{1}{p}\mathsf{Tr}(R_N^{(k)}\Sigma))}$$

and

$$\delta_N^2 = -\frac{1}{N}\sum_{k=1}^{p}\frac{C_k^*R_N^{(k)}C_k - \mathsf{Tr}(R_N^{(k)}\Sigma)}{(1 + \frac{1}{p}\mathsf{Tr}(R_N^{(k)}\Sigma))(1 + \frac{1}{p}C_k^*R_N^{(k)}C_k)}.$$

We will now use (2.4) and (2.3) to show that δ_N a.s. converges to 0 as $N \to \infty$. It is known that F_N converges a.s. to the distribution F given by the Marčenko-Pastur equation (and has no subsequence vaguely convergent to 0). It is proven in Silverstein and Bai (1995) that under

18

these assumptions, there exists $m > 0$ such that $\inf_N F_N([-m, m]) > 0$. In particular, there exists $\delta > 0$ such that

$$\inf_N \text{Im} \left[\int \frac{1}{\lambda - z} dF_N(\lambda) \right] \geq \int \frac{y}{2\lambda^2 + 2x^2 + y^2} dF_N(\lambda) \geq \delta.$$

From this, we deduce that

$$\left| 1 + \frac{1}{p} \text{Tr}(\Sigma R_N) \right| \geq \text{Im} \left[\frac{1}{p} \text{Tr}(\Sigma R_N) \right] \geq \frac{h_1}{\gamma} \delta.$$

Using (2.4) we also get that

$$\left| 1 + \frac{1}{p} \text{Tr} \left(\Sigma R_N^{(k)} \right) \right| \geq \text{Im} \left[\frac{1}{p} \text{Tr} \left(\Sigma R_N^{(k)} \right) \right] \geq \frac{h_1}{2\gamma} \delta.$$

We first consider δ_N^1. Thus one has that

$$|\delta_N^1| \leq \frac{2||\Sigma||\gamma^2}{N y h_1^2 \delta^2} = O(1/N). \tag{2.6}$$

We now turn to δ_N^2. Using the a.s. convergence (2.3), it is not hard to deduce that

$$\delta_N^2 \to 0, \text{ a.s.}$$

This completes the proof of Lemma 2.2. □

Lemma 2.3. *For every* $k = 1, 2, \ldots$ *the limit* $\lim_{N \to \infty} \Theta_N^{(k)}(z) := \Theta^{(k)}(z)$ *exists and satisfies the recursion equation*

$$\forall z \in \mathbb{C}^+, \quad \Theta^{(k+1)}(z) = \left[z \Theta^{(k)}(z) + \int_{-\infty}^{+\infty} \tau^k dH(\tau) \right] \times \left[1 + \frac{1}{\gamma} \Theta^{(1)}(z) \right]. \tag{2.7}$$

Proof of Lemma 2.3 The proof is inductive, so we assume that formula (2.7) holds for any integer smaller than or equal to q for some given integer q. We start from the formula

$$\text{Tr} \left(\Sigma^q + z \Sigma^q R_N(z) \right) = \text{Tr} \left(\Sigma^q R_N(z) S_N \right) = \sum_{k=1}^p C_k^* \Sigma^q R_N(z) C_k.$$

Using once more the resolvent identity, one gets that

$$C_k^* \Sigma^q R_N(z) C_k = \frac{C_k^* \Sigma^q R_N^{(k)}(z) C_k}{1 + C_k^* R_N^{(k)}(z) C_k},$$

19

which yields that

$$\frac{1}{N}\mathsf{Tr}\left(\Sigma^q + z\Sigma^q R_N(z)\right) = \sum_{k=1}^{p} \frac{C_k^* \Sigma^q R_N^{(k)}(z) C_k}{1 + C_k^* R_N^{(k)}(z) C_k}. \tag{2.8}$$

It is now an easy consequence of the arguments developed in the case where $q = 0$ to check that

$$\max_{k\in\{1,\dots,p\}} \left| C_k^* R_N^{(k)}(z) C_k - \mathsf{Tr}\left(\Sigma R_N(z)\right) \right| + \left| C_k^* \Sigma^q R_N^{(k)}(z) C_k - \mathsf{Tr}\left(\Sigma^{q+1} R_N(z)\right) \right|$$

converges a.s. to zero. Using the recursion assumption that $\lim_{N\to\infty}\Theta_N^{(k)}(z)$ exists, $\forall k \le q$, one can deduce that $\lim_{N\to\infty}\Theta_N^{(q+1)}(z)$ exists and that the limit $\Theta^{(q+1)}(z)$ satisfies

$$\left[z\Theta^{(q)}(z) + \int_{-\infty}^{+\infty} \tau^q dH(\tau)\right] \times \left[1 + \frac{1}{\gamma}\Theta^{(1)}(z)\right] = \Theta^{(q+1)}(z).$$

This finishes the proof of Lemma 2.3. □

Lemma 2.4. *Theorem 1.2 holds when the function g is a polynomial.*

Proof of Lemma 2.4 Given the linearity of the problem, it is sufficient to prove that Theorem 1.2 holds when the function g is of the form: $\forall \tau \in \mathbb{R}, \quad g(\tau) = \tau^k$, for any nonnegative integer k. In the case where $k = 0$, this is a direct consequence of Theorem 1.1 in [27].

The existence of a function $\Theta^{(k)}$ defined on $\mathbb{C}^+$ such that $\Theta_N^{(k)}(z) \xrightarrow{\text{a.s.}} \Theta^{(k)}(z)$ for all $z \in \mathbb{C}^+$ is established by Lemma 2.2 for $k = 1$ and by Lemma 2.3 for $k = 2, 3, \dots$ Therefore, all that remains to be shown is that Equation (1.5) holds for $k = 1, 2, \dots$

We will first show it for $k = 1$. From the original Marčenko-Pastur equation we know that:

$$1 + z m_F(z) = \int_{-\infty}^{+\infty} \frac{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right]}{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right] - z} dH(\tau). \tag{2.9}$$

From Lemma 2.2 we know that:

$$\Theta^{(1)}(z) = \frac{\gamma^2}{\gamma - 1 - z\, m_F(z)} - \gamma = \frac{1 + z\, m_F(z)}{1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)},$$

yielding that

$$1 + z m_F(z) = \frac{\Theta^{(1)}(z)}{1 + \gamma^{-1}\Theta^{(1)}(z)}. \tag{2.10}$$

Combining Equations (2.9) and (2.10) yields:

$$\int_{-\infty}^{+\infty} \frac{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right]}{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right] - z} dH(\tau) = \frac{\Theta^{(1)}(z)}{1 + \gamma^{-1}\Theta^{(1)}(z)}. \tag{2.11}$$

From Lemma 2.2, we also know that:

$$1 + \gamma^{-1}\Theta^{(1)}(z) = \frac{1}{1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)}. \tag{2.12}$$

Putting together Equations (2.11) and (2.12) yields the simplification:

$$\Theta^{(1)}(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)\right] - z}\,\tau\,dH(\tau),$$

which establishes that Equation (1.5) holds when $g(\tau) = \tau$, $\forall \tau \in \mathbb{R}$.

We now show by induction that Equation (1.5) holds when $g(\tau) = \tau^k$ for $k = 2, 3, \ldots$ Assume that we have proven it for $k - 1$. Thus the recursion hypothesis is that:

$$\Theta^{(k-1)}(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)\right] - z}\,\tau^{k-1}\,dH(\tau). \tag{2.13}$$

From Lemma 2.3 we know that:

$$\Theta^{(k)}(z) = \left[z\Theta^{(k-1)}(z) + \int_{-\infty}^{+\infty} \tau^{k-1}dH(\tau)\right] \times \left[1 + \frac{1}{\gamma}\Theta^{(1)}(z)\right]. \tag{2.14}$$

Combining Equations (2.13) and (2.14) yields:

$$
\begin{aligned}
\frac{\Theta^{(k)}(z)}{1 + \frac{1}{\gamma}\Theta^{(1)}(z)} &= z\Theta^{(k-1)}(z) + \int_{-\infty}^{+\infty} \tau^{k-1}dH(\tau) \\
&= \int_{-\infty}^{+\infty} \left\{\frac{z}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)\right] - z} + 1\right\} \tau^{k-1}\,dH(\tau) \\
&= \int_{-\infty}^{+\infty} \frac{1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)\right] - z}\,\tau^k\,dH(\tau).
\end{aligned} \tag{2.15}
$$

Putting together Equations (2.12) and (2.15) yields the simplification:

$$\Theta^{(k)}(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)\right] - z}\,\tau^k\,dH(\tau),$$

which proves that the desired assertion holds for k. Therefore, by induction, it holds for all $k = 1, 2, 3, \ldots$ This completes the proof of Lemma 2.4. $\square$

Lemma 2.5. *Theorem 1.2 holds for any function g that is continuous on $[h_1, h_2]$.*

Proof of Lemma 2.5 We shall deduce this from Lemma 2.4. Let g be any function that is continuous on $[h_1, h_2]$. By the Weierstrass approximation theorem, there exists a sequence of polynomials that converges to g uniformly on $[h_1, h_2]$. By Lemma 2.4, Theorem 1.2 holds for every polynomial in the sequence. Therefore it also holds for the limit g. $\square$

We are now ready to prove Theorem 1.2. We shall prove it by induction on the number k of points of discontinuity of the function g on the interval $[h_1, h_2]$. The fact that it holds for $k = 0$ has been established by Lemma 2.5. Let us assume that it holds for some k. Then consider any bounded function g which has $k + 1$ points of discontinuity on $[h_1, h_2]$. Let ν be one of these $k + 1$ points of discontinuity. Construct the function: $\forall x \in [h_1, h_2], \quad \rho(x) = g(x) \times (x - \nu)$. The function ρ has k points of discontinuity on $[h_1, h_2]$: all the ones that g has, except ν. Therefore, by the recursion hypothesis, $\Theta^\rho_N(z) = N^{-1}\mathsf{Tr}\left[(S_N - zI)^{-1}\rho(\Sigma_N)\right]$ converges a.s. to

$$\Theta^\rho(z) = \int_{-\infty}^{+\infty} \frac{1}{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right] - z}\, \rho(\tau)\, dH(\tau) \tag{2.16}$$

for all $z \in \mathbb{C}^+$. It is easy to adapt the arguments developed in the proof of Lemma 2.3 to show that $\lim_{N \to \infty} \Theta^g_N(z)$ exists (as g is bounded) and is equal to:

$$\Theta^g(z) = \frac{\Theta^\rho(z) - \left[1 + \gamma^{-1}\Theta^{(1)}(z)\right] \int_{-\infty}^{+\infty} g(\tau)dH(\tau)}{z\left[1 + \gamma^{-1}\Theta^{(1)}(z)\right] - \nu} \tag{2.17}$$

for all $z \in \mathbb{C}^+$. Plugging Equation (2.16) into Equation (2.17) yields:

$$\Theta^g(z) = \frac{\int_{-\infty}^{+\infty} \left\{ \frac{\tau - \nu}{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right] - z} - \left[1 + \gamma^{-1}\Theta^{(1)}(z)\right] \right\} g(\tau)\, dH(\tau)}{z\left[1 + \gamma^{-1}\Theta^{(1)}(z)\right] - \nu}.$$

Using Equation (2.12) we get:

$$
\begin{aligned}
\Theta^g(z) &= \frac{\int_{-\infty}^{+\infty} \left\{ \frac{\tau - \nu}{\tau[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)] - z} - \frac{1}{1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)} \right\} g(\tau)\, dH(\tau)}{\frac{z}{1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)} - \nu} \\[2mm]
&= \frac{\int_{-\infty}^{+\infty} \frac{z - \nu[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)]}{\{\tau[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)] - z\} \times [1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)]} g(\tau)\, dH(\tau)}{\frac{z - \nu[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)]}{1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)}} \\[2mm]
&= \int_{-\infty}^{+\infty} \frac{1}{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right] - z}\, g(\tau)\, dH(\tau),
\end{aligned}
$$

which means that Equation (1.5) holds for g. Therefore, by induction, Theorem 1.2 holds for any bounded function g with a finite number of discontinuities on $[h_1, h_2]$. $\square$

3 Proof of Theorem 1.3

At this stage, we need to establish two Lemmas that will be of general use for deriving implications from Theorem 1.2.

Lemma 3.1. *Let g denote a (real-valued) bounded function defined on $[h_1, h_2]$ with finitely many points of discontinuity. Consider the function Ω_N^g defined by:*

$$\forall x \in \mathbb{R}, \ \Omega_N^g(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{[\lambda_i, +\infty)}(x) \sum_{j=1}^{N} |u_i^* v_j|^2 \times g(\tau_j).$$

Then there exists a nonrandom function Ω^g defined on $\mathbb{R}$ such that $\Omega_N^g(x) \overset{a.s.}{\to} \Omega^g(x)$ at all points of continuity of Ω^g. Furthermore,

$$\Omega^g(x) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \text{Im}\left[\Theta^g(\lambda + i\eta) \right] d\lambda \tag{3.1}$$

for all x where Ω^g is continuous.

Proof of Lemma 3.1 The Stieltjes transform of Ω_N^g is the function Θ_N^g defined by Equation (1.3). From Theorem 1.2, we know that there exists a nonrandom function Θ^g defined over $\mathbb{C}^+$ such that $\Theta_N^g(z) \overset{a.s.}{\to} \Theta^g(z)$ for all $z \in \mathbb{C}^+$. Therefore, Silverstein and Bai's [4] Equation (2.5) implies that: $\lim_{N \to \infty} \Omega_N(x) \equiv \Omega^g(x)$ exists for all x where Ω^g is continuous. Furthermore, the Stieltjes transform of Ω^g is Θ^g. Then Equation (3.1) is simply the inversion formula for the Stieltjes transform. $\square$

Lemma 3.2. *Under the assumptions of Lemma 3.1, if $\gamma > 1$ then for all $(x_1, x_2) \in \mathbb{R}^2$:*

$$\Omega^g(x_2) - \Omega^g(x_1) = \frac{1}{\pi} \int_{x_1}^{x_2} \lim_{\eta \to 0^+} \text{Im}\left[\Theta^g(\lambda + i\eta) \right] d\lambda. \tag{3.2}$$

If $\gamma < 1$ then Equation (3.2) holds for all $(x_1, x_2) \in \mathbb{R}^2$ such that $x_1 x_2 > 0$.

Proof of Lemma 3.2 One can first note that $\lim_{z \in \mathbb{C}^+ \to x} \text{Im}\left[\Theta^g(z) \right] \equiv \text{Im}\left[\Theta^g(x) \right]$ exists for all $x \in \mathbb{R}$ (resp. all $x \in \mathbb{R} - \{0\}$) in the case where $\gamma > 1$ (resp. $\gamma < 1$). This is obvious if $x \in \text{Supp}(F)$. In the case where $x \notin \text{Supp}(F)$, then it can be deduced from Theorem 4.1 in [11] that $\dfrac{x}{1 - \gamma^{-1}(1 + x \breve{m}_F(x))} \notin \text{Supp}(H)$, which ensures the desired result. Now Θ^g is the Stieltjes transform of Ω^g. Therefore, Silverstein and Choi's [11] Theorem 2.1 implies that:

$$\Omega^g \text{ is differentiable at } x \text{ and its derivative is: } \frac{1}{\pi} \text{Im}\left[\Theta^g(x) \right]$$

for all $x \in \mathbb{R}$ (resp. all $x \in \mathbb{R} - \{0\}$) in the case where $\gamma > 1$ (resp. $\gamma < 1$). When we integrate, we get Equation (3.2). $\square$

23

We are now ready to proceed with the proof of Theorem 1.3. Let $\tau \in \mathbb{R}$ be given and take $g = \mathbf{1}_{(-\infty,\tau)}$. Then we have:

$$\forall z \in \mathbb{C}^+, \quad \Theta_N^{\mathbf{1}_{(-\infty,\tau)}}(z) = \frac{1}{N}\sum_{i=1}^N \frac{1}{\lambda_i - z}\sum_{j=1}^N |u_i^* v_j|^2 \times \mathbf{1}_{(-\infty,\tau)}.$$

Since the function $g = \mathbf{1}_{(-\infty,\tau)}$ has a single point of discontinuity (at τ), Theorem 1.2 implies that $\forall z \in \mathbb{C}^+, \quad \Theta_N^{\mathbf{1}_{(-\infty,\tau)}}(z) \xrightarrow{\text{a.s.}} \Theta^{\mathbf{1}_{(-\infty,\tau)}}(z)$, where:

$$\forall z \in \mathbb{C}^+, \quad \Theta^{\mathbf{1}_{(-\infty,\tau)}}(z) = \int_{-\infty}^{\tau} \frac{1}{t\left[1 - \gamma^{-1} - \gamma^{-1}z\,m_F(z)\right] - z}\,dH(t). \tag{3.3}$$

Remember from Equation (1.8) that:

$$\Phi_N(\lambda, \tau) = \lim_{\eta \to 0^+} \frac{1}{\pi}\int_{-\infty}^{\lambda} \text{Im}\left[\Theta_N^{\mathbf{1}_{(-\infty,\tau)}}(l + i\eta)\right] dl.$$

Therefore, by Lemma 3.1, $\lim_{N\to\infty}\Phi_N(\lambda,\tau)$ exists and is equal to:

$$\Phi(\lambda,\tau) = \lim_{\eta \to 0^+} \frac{1}{\pi}\int_{-\infty}^{\lambda} \text{Im}\left[\Theta^{\mathbf{1}_{(-\infty,\tau)}}(l + i\eta)\right] dl, \tag{3.4}$$

for every $(\lambda,\tau) \in \mathbb{R}^2$ where Φ is continuous. We first evaluate $\Phi(\lambda,\tau)$ in the case where $\gamma > 1$, so that the limiting e.d.f. of sample eigenvalues F is continuously differentiable on all of $\mathbb{R}$. Plugging (3.3) into (3.4) yields:

$$\begin{aligned}
\Phi(\lambda,\tau) &= \lim_{\eta \to 0^+} \frac{1}{\pi}\int_{-\infty}^{\lambda} \text{Im}\left\{\int_{-\infty}^{\tau} \frac{1}{t\left[a(l,\eta) + ib(l,\eta)\right] - l - i\eta}\,dH(t)\right\} dl \\
&= \frac{1}{\pi}\int_{-\infty}^{\lambda}\int_{-\infty}^{\tau}\lim_{\eta \to 0^+}\text{Im}\left\{\frac{1}{t\left[a(l,\eta) + ib(l,\eta)\right] - l - i\eta}\right\} dH(t)\,dl, \tag{3.5}
\end{aligned}$$

where $a(l,\eta) + ib(l,\eta) = 1 - \gamma^{-1} - \gamma^{-1}(l + i\eta)\,m_F(l + i\eta)$. The last equality follows from Lemma 3.2. Notice that:

$$\text{Im}\left\{\frac{1}{t\left[a(l,\eta) + ib(l,\eta)\right] - l - i\eta}\right\} = \frac{\eta - b(l,\eta)t}{\left[a(l,\eta)t - l\right]^2 + \left[b(l,\eta)t - \eta\right]^2}.$$

Taking the limit as $\eta \to 0^+$, we get:

$$a(l,\eta) \longrightarrow a = \text{Re}\left[1 - \frac{1}{\gamma} - \frac{l\breve{m}_F(l)}{\gamma}\right],\; b(l,\eta) \longrightarrow b = \text{Im}\left[1 - \frac{1}{\gamma} - \frac{l\breve{m}_F(l)}{\gamma}\right].$$

The inversion formula for the Stieltjes transform implies: $\forall l \in \mathbb{R}$, $F'(l) = \frac{1}{\pi}\mathsf{Im}\,[\check{m}_F(l)]$, therefore $b = -\pi\gamma^{-1}lF'(l)$. Thus we have:

$$\lim_{\eta \to 0^+} \mathsf{Im}\left\{\frac{1}{t\,[a(l,\eta) + ib(l,\eta)] - l - i\eta}\right\} = \frac{\pi\gamma^{-1}lt}{(at-l)^2 + b^2t^2} \times F'(l). \tag{3.6}$$

Plugging Equation (3.6) back into Equation (3.5) yields that:

$$\Phi(\lambda,\tau) = \int_{-\infty}^{\lambda}\int_{-\infty}^{\tau} \frac{\gamma^{-1}lt}{(at-l)^2 + b^2t^2}\, dH(t)\, dF(l),$$

which was to be proven. This completes the proof of Theorem 1.3 in the case where $\gamma > 1$.

In the case where $\gamma < 1$, much of the arguments remain the same, except for an added degree of complexity due to the fact that the limiting e.d.f. of sample eigenvalues F has a discontinuity of size $1 - \gamma$ at zero. This is handled by using the following three Lemmas.

Lemma 3.3. *If $\gamma \neq 1$, F is constant over the interval $\left(0, (1 - \frac{1}{\sqrt{\gamma}})^2 h_1\right)$.*

Proof of Lemma 3.3 If H placed all its weight on h_1, then we could solve the Marčenko-Pastur equation explicitly for F, and the infimum of the support of the limiting e.d.f. of nonzero sample eigenvalues would be equal to $(1 - \gamma^{-1/2})^2 \times h_1$. Since, by Assumption (H_4), H places all its weight on points greater than or equal to h_1, the infimum of the support of the limiting e.d.f. of nonzero sample eigenvalues has to be greater than or equal to $(1 - \gamma^{-1/2})^2 \times h_1$ (see Equation (1.9b) in Bai and Silverstein [7]). Therefore, F is constant over the open interval $\left(0, (1 - \gamma^{-1/2})^2 \times h_1\right)$. $\square$

Lemma 3.4. *Let $\kappa > 0$ be a given real number. Let μ be a complex holomorphic function defined on the set $\{z \in \mathbb{C}^+ : \mathsf{Re}[z] \in (-\kappa, \kappa)\}$. If $\mu(0) \in \mathbb{R}$ then:*

$$\lim_{\varepsilon \to 0^+}\left\{\lim_{\eta \to 0^+} \frac{1}{\pi}\int_{-\varepsilon}^{+\varepsilon} \mathsf{Im}\left[-\frac{\mu(\xi + i\eta)}{\xi + i\eta}\right] d\xi\right\} = \mu(0).$$

Proof of Lemma 3.4 For all ε in $(0, \kappa)$, we have:

$$\begin{aligned}
\lim_{\eta \to 0^+} \frac{1}{\pi}\int_{-\varepsilon}^{+\varepsilon} \mathsf{Im}\left[-\frac{1}{\xi + i\eta}\right] d\xi &= \lim_{\eta \to 0^+} \frac{1}{\pi}\int_{-\epsilon}^{+\epsilon} \frac{\eta}{\xi^2 + \eta^2} d\xi \\
&= \lim_{\eta \to 0^+} \frac{1}{\pi}\left[\arctan\left(\frac{\varepsilon}{\eta}\right) - \arctan\left(\frac{-\varepsilon}{\eta}\right)\right] \\
&= 1. \tag{3.7}
\end{aligned}$$

Since μ is continuously differentiable, there exist $\delta > 0$, $\beta > 0$ such that $|\mu'(z)| \leq \beta, \forall z, |z| \leq \delta$. Using Taylor's theorem, we get that $|\mu(z) - \mu(0)| \leq \beta|z|, \forall |z| \leq \delta$. Now we can perform the following decomposition:

$$\lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[-\frac{\mu(\xi + i\eta)}{\xi + i\eta} \right] d\xi \right\}$$

$$= \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[-\frac{\mu(\xi + i\eta) - \mu(0) + \mu(0)}{\xi + i\eta} \right] d\xi \right\}$$

$$= \mu(0) \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[-\frac{1}{\xi + i\eta} \right] d\xi \right\}$$

$$+ \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[-\frac{\mu(\xi + i\eta) - \mu(0)}{\xi + i\eta} \right] d\xi \right\}$$

$$= \mu(0) + \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[-\frac{\mu(\xi + i\eta) - \mu(0)}{\xi + i\eta} \right] d\xi \right\},$$

where the last equality follows from Equation (3.7). The second term vanishes because:

$$\left| \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[-\frac{\mu(\xi + i\eta) - \mu(0)}{\xi + i\eta} \right] d\xi \right\} \right|$$

$$\leq \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \left| \frac{\mu(\xi + i\eta) - \mu(0)}{\xi + i\eta} \right| d\xi \right\}$$

$$\leq \lim_{\varepsilon \to 0^+} \left\{ \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \beta \, d\xi \right\} = 0.$$

This yields Lemma 3.4. $\square$

Lemma 3.5. *Assume that $\gamma < 1$. Let g be a (real-valued) bounded function defined on $[h_1, h_2]$ with finitely many points of discontinuity. Then:*

$$\lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \int_{-\infty}^{+\infty} \mathsf{Im} \left\{ \frac{g(\tau)}{\tau \left[1 - \gamma^{-1} - \gamma^{-1}(\xi + i\eta) \, m_F(\xi + i\eta) \right] - \xi - i\eta} \right\} dH(\tau) d\xi$$

$$= \int_{-\infty}^{+\infty} \frac{g(\tau)}{1 + \breve{m}_{\underline{F}}(0)\tau} \, dH(\tau),$$

where $\underline{F} = \left(1 - \gamma^{-1} \right) \mathbf{1}_{[0, +\infty)} + \gamma^{-1} F$, *and* $\breve{m}_{\underline{F}}(0) = \lim_{z \in \mathbb{C}^+ \to 0} m_{\underline{F}}(z)$.

Proof of Lemma 3.5   One has that

$$\forall z \in \mathbb{C}^+, \quad 1 + zm_F(z) = \gamma + \gamma z m_{\underline{F}}(z), \tag{3.8}$$

$$\tau \left[1 - \gamma^{-1} + \gamma^{-1} z m_F(z) \right] - z = -z \left[1 + m_{\underline{F}}(z)\tau \right]. \tag{3.9}$$

26

Define:

$$\mu(z) = \int_{-\infty}^{+\infty} \frac{g(\tau)}{1 + m_{\underline{F}}(z)\tau}\, dH(\tau).$$

Equation (3.9) yields:

$$\lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \int_{-\infty}^{+\infty} \mathsf{Im} \left\{ \frac{g(\tau)}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}(\xi + i\eta)\, m_F(\xi + i\eta)\right] - \xi - i\eta} \right\} dH(\tau) d\xi$$

$$= \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left\{ -\frac{1}{\xi + i\eta} \int_{-\infty}^{+\infty} \frac{g(\tau)}{1 + m_{\underline{F}}(\xi + i\eta)\tau}\, dH(\tau) \right\} d\xi$$

$$= \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left\{ -\frac{\mu(\xi + i\eta)}{\xi + i\eta} \right\}$$

$$= \int_{-\infty}^{+\infty} \frac{g(\tau)}{1 + \breve{m}_{\underline{F}}(0)\tau}\, dH(\tau),$$

where the last equality follows from Lemma 3.4. $\square$

We are now ready to complete the proof of Theorem 1.3 for the case where $\gamma < 1$. The inversion formula for the Stieltjes transform implies that:

$$\lim_{\varepsilon \to 0^+} \left[\Phi(\varepsilon, \tau) - \Phi(-\varepsilon, \tau)\right]$$

$$= \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left[\Theta^{\mathbf{1}_{(-\infty,\tau)}}(\xi + i\eta) \right] d\xi$$

$$= \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left\{ \int_{-\infty}^{\tau} \frac{dH(t)}{t\left[1 - \gamma^{-1} - \gamma^{-1}(\xi + i\eta)m_F(\xi + i\eta)\right] - \xi - i\eta} \right\} d\xi$$

$$= \int_{-\infty}^{\tau} \frac{1}{1 + \breve{m}_{\underline{F}}(0)\, t}\, dH(t), \tag{3.10}$$

where the last equality follows from Lemma 3.5. By Lemma 3.3, we know that for λ in a neighborhood of zero: $F(\lambda) = (1 - \gamma)\mathbf{1}_{[0,+\infty)}(\lambda)$. From Equation (3.10) we know that for λ in a neighborhood of zero:

$$\Phi(\lambda, \tau) = \int_{-\infty}^{\lambda} \int_{-\infty}^{\tau} \frac{1}{1 + \breve{m}_{\underline{F}}(0)\, t}\, dH(t)\, d\mathbf{1}_{[0,+\infty)}(l).$$

Comparing the two expressions, we find that for λ in a neighborhood of zero:

$$\Phi(\lambda, \tau) = \int_{-\infty}^{\lambda} \int_{-\infty}^{\tau} \frac{1}{(1 - \gamma)\left[1 + \breve{m}_{\underline{F}}(0)\, t\right]}\, dH(t)\, dF(l).$$

Therefore, if we define φ as in (1.9), then we can see that for λ in a neighborhood of zero:

$$\Phi(\lambda, \tau) = \int_{-\infty}^{\lambda} \int_{-\infty}^{\tau} \varphi(l, t)\, dH(t)\, dF(l). \tag{3.11}$$

27

From this point onwards, the fact that Equation (3.11) holds for all $\lambda > 0$ can be established exactly like we did in the case where $\gamma > 1$. This completes the proof of Theorem 1.3. $\square$

4 Proofs of Theorems 1.4 and 1.5

4.1 Proof of Theorem 1.4

Lemma 2.2 shows that $\forall z \in \mathbb{C}^+$, $\Theta_N^{(1)}(z) \overset{\text{a.s.}}{\to} \Theta^{(1)}(z)$, where:

$$\forall z \in \mathbb{C}^+,\ \Theta^{(1)}(z) = \frac{\gamma}{1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)} - \gamma. \tag{4.1}$$

Remember from Equation (1.12) that:

$$\Delta_N(x) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \operatorname{Im}\left[\Theta_N^{(1)}(\lambda + i\eta)\right] d\lambda.$$

Therefore, by Lemma 3.1, $\lim_{N \to \infty} \Delta_N(x)$ exists and is equal to:

$$\Delta(x) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \operatorname{Im}\left[\Theta^{(1)}(\lambda + i\eta)\right] d\lambda \tag{4.2}$$

for every $x \in \mathbb{R}$ where Δ is continuous. We first evaluate $\Delta(x)$ in the case where $\gamma > 1$. Plugging Equation (4.1) into Equation (4.2) yields:

$$
\begin{aligned}
\Delta(x) &= \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \operatorname{Im}\left[\frac{\gamma}{1 - \gamma^{-1} - \gamma^{-1}(\lambda + i\eta) m_F(\lambda + i\eta)} - \gamma\right] d\lambda \\
&= \lim_{\eta \to 0^+} \int_{-\infty}^{x} \frac{\pi^{-1} \operatorname{Im}\left[(\lambda + i\eta) m_F(\lambda + i\eta)\right]}{\left|1 - \gamma^{-1} - \gamma^{-1}(\lambda + i\eta) m_F(\lambda + i\eta)\right|^2} d\lambda \\
&= \int_{-\infty}^{x} \lim_{\eta \to 0^+} \frac{\pi^{-1} \operatorname{Im}\left[(\lambda + i\eta) m_F(\lambda + i\eta)\right]}{\left|1 - \gamma^{-1} - \gamma^{-1}(\lambda + i\eta) m_F(\lambda + i\eta)\right|^2} d\lambda \\
&= \int_{-\infty}^{x} \frac{\pi^{-1} \operatorname{Im}\left[\lambda \breve{m}_F(\lambda)\right]}{\left|1 - \gamma^{-1} - \gamma^{-1}\lambda \breve{m}_F(\lambda)\right|^2} d\lambda \\
&= \int_{-\infty}^{x} \frac{\lambda F'(\lambda)}{\left|1 - \gamma^{-1} - \gamma^{-1}\lambda \breve{m}_F(\lambda)\right|^2} d\lambda \\
&= \int_{-\infty}^{x} \frac{\lambda}{\left|1 - \gamma^{-1} - \gamma^{-1}\lambda\, \breve{m}_F(\lambda)\right|^2} dF(\lambda),
\end{aligned}
\tag{4.3}
$$

where Equation (4.3) made use of Lemma 3.2. This completes the proof of Theorem 1.4 in the case where $\gamma > 1$.

In the case where $\gamma < 1$, much of the arguments remain the same. The inversion formula for the Stieltjes transform implies that:

$$\lim_{\varepsilon \to 0^+} [\Delta(\varepsilon) - \Delta(-\varepsilon)]$$

$$= \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im}\left[\Theta^{(1)}(\xi + i\eta) \right] d\xi$$

$$= \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im}\left\{ \int_{-\infty}^{+\infty} \frac{\tau \, dH(\tau)}{\tau\left[1 - \gamma^{-1} - \gamma^{-1}(\xi + i\eta) m_F(\xi + i\eta)\right] - \xi - i\eta} \right\} d\xi$$

$$= \int_{-\infty}^{+\infty} \frac{\tau}{1 + \widecheck{m}_{\underline{F}}(0)\,\tau} \, dH(\tau), \tag{4.4}$$

where the last equality follows from Lemma 3.5. Notice that for all $z \in \mathbb{C}^+$:

$$\int_{-\infty}^{+\infty} \frac{\tau}{1 + m_{\underline{F}}(z)\,\tau} \, dH(\tau) = \frac{1}{m_{\underline{F}}(z)} \int_{-\infty}^{+\infty} \frac{1 + m_{\underline{F}}(z)\,\tau - 1}{1 + m_{\underline{F}}(z)\,\tau} \, dH(\tau)$$

$$= \frac{1}{m_{\underline{F}}(z)} - \frac{1}{m_{\underline{F}}(z)} \int_{-\infty}^{+\infty} \frac{1}{1 + m_{\underline{F}}(z)\,\tau} \, dH(\tau). \tag{4.5}$$

Plugging Equation (3.9) into Equation (4.5) yields:

$$\int_{-\infty}^{+\infty} \frac{\tau}{1 + m_{\underline{F}}(z)\,\tau} \, dH(\tau)$$

$$= \frac{1}{m_{\underline{F}}(z)} + \frac{z}{m_{\underline{F}}(z)} \int_{-\infty}^{+\infty} \frac{1}{1 - \gamma^{-1} + \gamma^{-1} z m_F(z)} \, dH(\tau)$$

$$= \frac{1 + z\, m_F(z)}{m_{\underline{F}}(z)}, \tag{4.6}$$

where the last equality comes from the original Marčenko-Pastur equation. Plugging Equation (3.8) into Equation (4.6) yields:

$$\int_{-\infty}^{+\infty} \frac{\tau}{1 + m_{\underline{F}}(z)\,\tau} \, dH(\tau) = \gamma \frac{1 + z\, m_{\underline{F}}(z)}{m_{\underline{F}}(z)}.$$

Taking the limit as $z \in \mathbb{C}^+ \to 0$, we get:

$$\int_{-\infty}^{+\infty} \frac{\tau}{1 + \widecheck{m}_{\underline{F}}(0)\,\tau} \, dH(\tau) = \frac{\gamma}{\widecheck{m}_{\underline{F}}(0)}.$$

Plugging this result back into Equation (4.4) yields:

$$\lim_{\varepsilon \to 0^+} [\Delta(\varepsilon) - \Delta(-\varepsilon)] = \frac{\gamma}{\widecheck{m}_{\underline{F}}(0)}. \tag{4.7}$$

29

By Lemma 3.3, we know that for λ in a neighborhood of zero: $F(\lambda) = (1-\gamma)\mathbf{1}_{[0,+\infty)}(\lambda)$. From Equation (4.7) we know that for x in a neighborhood of zero:

$$\Delta(x) = \int_{-\infty}^{x} \frac{\gamma}{\breve{m}_{\underline{F}}(0)} \, d\mathbf{1}_{[0,+\infty)}(\lambda).$$

Comparing the two expressions, we find that for x in a neighborhood of zero:

$$\Delta(x) = \int_{-\infty}^{x} \frac{\gamma}{(1-\gamma)\,\breve{m}_{\underline{F}}(0)} \, dF(\lambda).$$

Therefore, if we define δ as in (1.13), then we can see that for x in a neighborhood of zero:

$$\Delta(x) = \int_{-\infty}^{x} \delta(\lambda)\, dF(\lambda). \tag{4.8}$$

From this point onwards, the fact that Equation (4.8) holds for all $x > 0$ can be established exactly like we did in the case where $\gamma > 1$. Thus the proof of Theorem 1.4 is complete. $\square$

4.2   Proof of Theorem 1.5

As

$$\forall x \in \mathbb{R}, \quad \Psi_N(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{[\lambda_i,+\infty)}(x) \sum_{j=1}^{N} \frac{|u_i^* v_j|^2}{\tau_j},$$

$$\forall z \in \mathbb{C}^+, \quad \Theta_N^{(-1)}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} \sum_{j=1}^{N} \frac{|u_i^* v_j|^2}{\tau_j},$$

and using the inversion formula for the Stieltjes transform, we obtain:

$$\forall x \in \mathbb{R},\ \Psi_N(x) = \lim_{\eta \to 0^+} \int_{-\infty}^{x} \mathsf{Im}\left[\Theta_N^{(-1)}(\lambda + i\eta)\right] d\lambda.$$

Since the function $g(\tau) = 1/\tau$ is continuous on $[h_1, h_2]$, Theorem 1.2 implies that $\forall z \in \mathbb{C}^+$, $\Theta_N^{(-1)}(z) \overset{\text{a.s.}}{\to} \Theta^{(-1)}(z)$, where:

$$\forall z \in \mathbb{C}^+, \quad \Theta^{(-1)}(z) = \int_{-\infty}^{+\infty} \frac{\tau^{-1}}{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_F(z)\right] - z} \, dH(\tau). \tag{4.9}$$

Therefore, by Lemma 3.1, $\lim_{N\to\infty} \Psi_N(x)$ exists and is equal to:

$$\Psi(x) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\infty}^{x} \mathsf{Im}\left[\Theta^{(-1)}(\lambda + i\eta)\right] d\lambda, \tag{4.10}$$

for every $x \in \mathbb{R}$ where Ψ is continuous. We first evaluate $\Psi(x)$ in the case where $\gamma > 1$, so that F is continuously differentiable on all of $\mathbb{R}$.

In the notation of Lemma 2.5, we set ν equal to zero so that $\forall \tau \in \mathbb{R}, \quad \rho(\tau) = g(\tau) \times \tau = 1$. Then Equation (2.17) implies that:

$$\forall z \in \mathbb{C}^+, \quad \Theta^{(-1)}(z) = \frac{m_F(z) - \left[1 + \gamma^{-1}\Theta^{(1)}(z)\right] \int_{-\infty}^{+\infty} \tau^{-1} dH(\tau)}{z \left[1 + \gamma^{-1}\Theta^{(1)}(z)\right]}.$$

Using Equation (2.12), we obtain:

$$\forall z \in \mathbb{C}^+, \quad \Theta^{(-1)}(z) = \frac{m_F(z)}{z} \left[1 - \gamma^{-1} - \gamma^{-1} z \, m_F(z)\right] - \frac{1}{z} \int_{-\infty}^{+\infty} \tau^{-1} dH(\tau). \qquad (4.11)$$

Thus for all $\lambda \in \mathbb{R}$:

$$\begin{aligned}
\lim_{\eta \to 0^+} \mathsf{Im}\left[\Theta^{-1}(\lambda + i\eta)\right] &= \frac{1}{\lambda} \mathsf{Im}\left\{\breve{m}_F(\lambda)\left[1 - \gamma^{-1} - \gamma^{-1}\lambda \breve{m}_F(\lambda)\right]\right\} \\
&= \frac{1}{\lambda}\left\{1 - \gamma^{-1} - 2\gamma^{-1}\lambda \, \mathsf{Re}\left[\breve{m}_F(\lambda)\right]\right\} \times \mathsf{Im}\left[\breve{m}_F(\lambda)\right] \\
&= \frac{1}{\lambda}\left\{1 - \gamma^{-1} - 2\gamma^{-1}\lambda \, \mathsf{Re}\left[\breve{m}_F(\lambda)\right]\right\} \times \pi F'(\lambda).
\end{aligned}$$

Plugging this result back into Equation (4.10) yields:

$$\begin{aligned}
\Psi(x) &= \frac{1}{\pi} \int_{-\infty}^{x} \lim_{\eta \to 0^+} \mathsf{Im}\left[\Theta^{(-1)}(\lambda + i\eta)\right] d\lambda \\
&= \int_{-\infty}^{x} \frac{1 - \gamma^{-1} - 2\gamma^{-1}\lambda \, \mathsf{Re}\left[\breve{m}_F(\lambda)\right]}{\lambda} \, dF(\lambda),
\end{aligned}$$

where we made use of Lemma 3.2. This completes the proof of Theorem 1.5 in the case where $\gamma > 1$.

We now turn to the case where $\gamma < 1$. Equation (4.10) implies that:

$$\lim_{\varepsilon \to 0^+} \left[\Psi(\varepsilon) - \Psi(-\varepsilon)\right] = \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im}\left[\Theta^{(-1)}(\xi + i\eta)\right] d\xi. \qquad (4.12)$$

Plugging Equation (3.8) into Equation (4.11) yields for all $z \in \mathbb{C}^+$:

$$\begin{aligned}
\Theta^{(-1)}(z) &= -m_F(z) \, m_{\underline{F}}(z) - \frac{1}{z} \int_{-\infty}^{+\infty} \frac{1}{\tau - 0} dH(\tau) \\
&= \frac{1}{z}\left[1 - \gamma - \gamma z m_{\underline{F}}(z)\right] m_{\underline{F}}(z) - \frac{1}{z}\breve{m}_H(0).
\end{aligned}$$

31

Plugging this result into Equation (4.12), we get:

$$\lim_{\varepsilon \to 0^+} [\Psi(\varepsilon) - \Psi(-\varepsilon)] = \lim_{\varepsilon \to 0^+} \lim_{\eta \to 0^+} \frac{1}{\pi} \int_{-\varepsilon}^{+\varepsilon} \mathsf{Im} \left\{ -\frac{\mu(\xi + i\eta)}{\xi + i\eta} \right\} d\xi,$$

where $\mu(z) = -[1 - \gamma - \gamma z m_{\underline{F}}(z)] m_{\underline{F}}(z) + \breve{m}_H(0)$. Therefore, by Lemma 3.4, we have:

$$\lim_{\varepsilon \to 0^+} [\Psi(\varepsilon) - \Psi(-\varepsilon)] = \mu(0) = -(1 - \gamma)\breve{m}_{\underline{F}}(0) + \breve{m}_H(0). \tag{4.13}$$

By Lemma 3.3, we know that for λ in a neighborhood of zero: $F(\lambda) = (1-\gamma)\mathbf{1}_{[0,+\infty)}(\lambda)$. From Equation (4.13) we know that for x in a neighborhood of zero:

$$\Psi(x) = \int_{-\infty}^{x} \left[-(1 - \gamma)\breve{m}_{\underline{F}}(0) + \breve{m}_H(0) \right] d\mathbf{1}_{[0,+\infty)}(\lambda).$$

Comparing the two expressions, we find that for x in a neighborhood of zero:

$$\Psi(x) = \int_{-\infty}^{x} \left[-\breve{m}_{\underline{F}}(0) + \frac{1}{1 - \gamma} \breve{m}_H(0) \right] dF(\lambda).$$

Therefore, if we define ψ as in (1.14), then we can see that for x in a neighborhood of zero:

$$\Psi(x) = \int_{-\infty}^{x} \psi(\lambda) \, dF(\lambda). \tag{4.14}$$

From this point onwards, the fact that Equation (4.14) holds for all $x > 0$ can be established exactly like we did in the case where $\gamma > 1$. Thus the proof of Theorem 1.5 is complete. $\square$

References

[1] BAI, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. ii, sample covariance matrices. *Ann. Probab. 21*, 649–672.

[2] BAI, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Stat. Sinica 9*, 611–677.

[3] BAI, Z. D., MIAO, B. Q., AND PAN, G., M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab. 35*, 1532–1572.

[4] BAI, Z. D. AND SILVERSTEIN, J. W. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal. 54*, 175–192.

[5] BAI, Z. D. AND SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab. 26*, 316–345.

[6] BAI, Z. D. AND SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Ann. Probab. 27*, 1536–1555.

[7] BAI, Z. D. AND SILVERSTEIN, J. W. (2004). Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab. 32*, 553–605.

[8] BAI, Z. D. AND SILVERSTEIN, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices.* Science Press, Beijing.

[9] BAI, Z. D. AND YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab. 21*, 1275–1294.

[10] BICKEL, P. J. AND LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Stat. 36, No 1*, 199–227.

[11] CHOI, S. I. AND SILVERSTEIN, J. W. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal. 54*, 295–309.

[12] COMBETTES, P. L. AND SILVERSTEIN, J. W. (1992). Signal detection via spectral theory of large dimensional random matrices. *IEEE Trans. Signal Process. 40*, 2100–2105.

[13] EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist. 36*, 2757–2790.

[14] GRENANDER, U. AND SILVERSTEIN, J. W. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math. 32*, 499–519.

[15] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist. 29*, 295–327.

[16] KRISHNAIAH, P. R. AND YIN, Y. Q. (1983). A limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal. 13*, 489–507.

[17] LEDOIT, O.AND WOLF, M. (2004). Honey, i shrunk the sample covariance matrix. *Journal of Portfolio Management 30*, 110–119.

[18] LEDOIT, O. (1995). Essays on risk and return in the stock market. Ph.D. thesis, Massachusetts Institute of Technology.

[19] LEDOIT, O. AND WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist. 30*, 1081–1102.

[20] LEDOIT, O. AND WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal. 88*, 365–411.

[21] MARČENKO, V. A. AND PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb. 1*, 457–486.

[22] PECHE, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Prob. Theor. Relat. Fields 143, no 3-4*, 481–516.

[23] PERLMAN, M. D. (2007). *Multivariate Statistical Analysis.* Online textbook published by the Department of Statistics of the University of Washington, Seattle.

[24] SILVERSTEIN, J. W. (1984). Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal. 15*, 295–324.

[25] SILVERSTEIN, J. W. (1989). On the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal. 30*, 1–16.

[26] SILVERSTEIN, J. W. (1990). Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab. 18*, 1174–1194.

[27] SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal. 55*, 331–339.

[28] WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab. 6*, 1–18.

[29] YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal. 20*, 50–68.