M3R Project in Mathematics

Artur Kotlicki
CID: 00695063

Supervisor: Prof. Rama Cont

Random matrix theory and estimation of high-dimensional covariance matrices

Imperial College London
Department of Mathematics

July 10, 2014

Declaration: I confirm that this submission is my own work. In it, I give references and citations whenever I refer to or use the published, or unpublished, work of others.

Signed: _____ , date: _____ .

Abstract

This projects aims to present significant results of random matrix theory in regards to the principal component analysis, including Wigner's semicircular law and Marčenko-Pastur law describing limiting distribution of large dimensional random matrices. The work bases on the large dimensional data assumptions, where both the number of variables and sample size tends to infinity, while their ratio tends to a finite limit.

Random matrix theory, over the past decade has been a fast growing area of mathematics, due to the advancements in technology and data collection methods. Treated as a tool to solve large dimensional problems, it has found its application in many research areas, such as signal processing, network security, image processing, genetic statistics, stock market analysis, and other finance or economic problems [1, p. 3].

In this project, key results enabling to establish a low dimensional factor model form a large noisy data will be stated, as well as a general way of proving them will be given. A significant portion of the proofs relies on the Stieltjes transform, a common tool used for studying the convergence of spectral distribution of large matrice, which is also discussed in this project. An algorithm suggested by Karoui [14] will be presented, giving a method of estimating the true population covariance method.

Empirical verification of main theorems is conducted, showing fast convergence rate in case of the Marčenko-Pastur law, and slower rate for the Wigner's semicircular law. Also, the established theory is applied to a real-life financial data, based on the S&P 500 index, for which 12 principal components have been identified when time horizon is equal to 10 years, and 10 principal components for data set over 5 years.

Contents

1 Introduction

For over 50 years the statistical properties of financial data sets have been extensively researched. However, only during the recent decade has it been possible to apply computer-intensive methods to analyse large data sets and seek for correlation between stock prices [5, p. 223]. It is currently a challenge to extract meaningful information from a *principal component analysis* done on highly dimensional data. The key interest of this project will be the eigenvalues of covariance matrices, which allow a low-dimensional approximation to the data to be obtained through a projection on the lower-dimensional subspace, in order to explain as much variance in the given data set as possible [14, p. 2758].

The underlying problem comes from the fact that today's statistics bases on data sets for which not only the sample size (represented by time index T, say) is large, but also the number of variables n is regarded as large. In this case the commonly used estimators, proven under the implicit assumption of asymptotic framework in which n is fixed, while T tends to infinity, are no longer justified [14, p. 2757]. In this project, the assumption of *large T, large n* is explored, where it will be assumed that the growth rate of both dimensions tends to a finite limit, so in other words, asymptotically, the ratio $y := n/T$ is constant. The key reason for the occurrence of the observed shift in the paradigm of statistical assumption, from the one where the number of variables n is fixed to the one in which n tends to infinity, is perhaps due to the significant advancement, over the past three or four decades, of computer science. Availability of data has rapidly increased, as well as technological limitations in regards to computational speeds and storage have been lifted. This change has given motivation for statisticians to develop new theory in regards to highly-dimensional datasets.

It has to be noted that in practice, however, there is some ambiguity regarding the choice of the setup, which determines whether classical limit theorems (for fixed n) are used or if large dimensional limit theorems (when $n \to \infty$) are applicable. For example, in some applications a sample size of $T = 100$ can be argued to be large enough to assume that $T \to \infty$, but it is not always clear whether the corresponding $n = 20$ variables should also be treated as n tending to infinity or as a fixed quantity in this case. For example, as argued by Huber [10], it is beneficial to study the case of increasing dimension together with the sample size in linear regression analysis. However, this problem should not be the focus of this project, in which the assumption of *large T, large n* is adopted.

Random matrix theory (RMT) has been developed as a tool providing special limiting theorems to deal with (practical) problems, in which classical limiting theorems fail. An example highlighting the case in which the classical limiting theorems are not suitable for use with high dimensional data is given in section 2.1. As random matrix theory has been useful in providing ways of dealing with large dimensional data analysis, it has been found in applications in many research areas, such as signal processing, network security, image processing, genetic statistics, stock market analysis, and other finance or economic problems [1, p. 3].

In this project a strong focus is being placed on the spectral distribution of large dimensional square matrices, studying key results such as the Wigner's Circular Law, sample covariance matrices, and the Marčenko-Pastur Law. A key tool in establishing these results will be the Stieltjes transform, discussed in section 1.3. Moreover, empirical verification of some of the results will be presented, as well as further extension of the results (such as the possibility of extracting information concerning the true population covariance matrix) will be suggested. The motivation for this project bases on the practical study conducted on the *S&P 500* index. Such approach will allow to link the random matrix theory to the stock market analysis, displaying powerful tools, which can be used by statisticians to extract meaningful information regarding the markets, stock indices and, in general, the economy, i.e. will allow for extraction of a low dimensional factor model from large and noisy dataset.

1.1 Structure of the Document

The document will be structured as follows:

- Section 1: Introduction and motivation of the study; preliminary definitions and theorems; and a discussion on Stieltjes transform – a key tool for establishing the further results in this project.
- Section 2: Theoretical discussion on spectral analysis of high dimensional random matrices, in which main results regarding the distribution of the eigenvalues of large matrices will be presented. The main application relates to the sample covariance matrices of highly dimensional datasets, with key theorems being the Wigner's semicircular law, and the Marčenko-Pastur law.
- Section 3: Empirical verification and discussion on the theory established in the previous section, based on the computer simulated data.
- Section 4: Further extension to the Marčenko-Pastur results, discussing a practical algorithm for retrieving information about the population covariance matrix based on the data.
- Section 5: Application of the established random matrix theory to the financial data, based on the S&P500 index.
- Section 6: Conclusion summarising findings of this project.

1.2 Preliminary Definitions and Theorems

In order to begin with the study of Stieltjes transform, it is useful to state the definition of Hermitian matrix, as well as give (without the proof[1]) the Lebesgue Dominated Convergence theorem.

Definition 1.2.1 (Hermitian matrix [23]) *A square matrix $\boldsymbol{A} = (a_{ij})$ is called Hermitian if it is self-adjoint, i.e. $\boldsymbol{A} = \bar{\boldsymbol{A}}^{\intercal}$, where $\bar{\boldsymbol{A}}^{\intercal}$ denotes the conjugate transpose. A square matrix being Hermitian is equivalent to the condition $a_{ij} = \bar{a}_{ji}$ holding.*

Theorem 1.2.1 (Lebesgue Dominated Convergence Theorem [20]) *Let $f_n : \mathbb{R} \to [-\infty, \infty]$ be Lebesgue measurable functions such that the pointwise limit $f(x) = \lim_{n\to\infty} f_n(x)$ exists, and assume that there exists an integrable function $g : \mathbb{R} \to [0, \infty]$ with $|f_n(x)| \leq g(x)$ for each $x \in \mathbb{R}$. Then f is integrable as is f_n for each n, and*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n \, d\mu = \int_{\mathbb{R}} \lim_{n\to\infty} f_n \, d\mu = \int_{\mathbb{R}} f \, d\mu.$$

1.3 Stieltjes Transform

One of the commonly used techniques in random matrix theory, and perhaps the key tool used to establish some of the main results in this project, is Stieltjes transform (also called Cauchy transform in the literature) of functions of bounded variation. It allows for a convenient, and yet very powerful, way of studying the convergence of spectral distribution of matrices (or operators), and can be though of, in a loose sense, as an equivalent of the characteristic function of a probability distribution used as a tool for the central limit theorems [14, p. 2763].

[1] Refer to Timoney [20] for the detailed discussion and the proof.

Definition 1.3.1 (Stieltjes transform [1, p. 514]) *Let $G(x)$ be a function of bounded variation on the real line, then its Stieltjes transform is defined by*

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

where $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}$, i.e. $\mathbb{C}^+$ is the set of complex numbers with strictly positive imaginary part.

It is possible to establish a one-to-one correspondence between the finite measures and their Stieltjes transforms, as shown in Theorem 1.3.1, below.

Theorem 1.3.1 (Inversion fromula [1, p. 514]) *For any continuity points $a < b$ of G, the following equation holds:*

$$G\{[a.b]\} = \lim_{\epsilon \downarrow 0} \frac{1}{\pi} \int_a^b \Im\, m_G(x + i\epsilon)\, dx.$$

Proof (of Theorem 1.3.1 [1, p. 514-515]): Firstly, note that it is possible to write

$$\begin{aligned}\frac{1}{\pi}\int_a^b \Im\, m_G(x+i\epsilon)\, \mathrm{d}x &= \frac{1}{\pi}\int_a^b \int \frac{\epsilon\, \mathrm{d}G(y)}{(x-y)^2+\epsilon^2} \\ &= \int \frac{1}{\pi}\left[\arctan(\epsilon^{-1}(b-y)) - \arctan(\epsilon^{-1}(a-y))\right]\, \mathrm{d}G(y).\end{aligned}$$

Then, letting $\epsilon \to 0$ gives that the right-hand side tends to $G[a, b]$ by applying the dominated convergence theorem (see Theorem 1.2.1). ∎

An important observation in the above proof has to be noted, namely

$$\Im\, m(z) = v \int \frac{\mathrm{d}G(x)}{(x-u)^2 + v^2}, \tag{1.3.1}$$

where $z = u + iv$ with $v > 0$. This can be used to establish the property of stieltjes transform of any distribution function F, which is given in Theorem 1.3.2.

Theorem 1.3.2 (Stieltjes transform of a distribution [1, p.517]) *For any distribution function F, its Stieltjes transform $m(z)$ satisfies*

$$|\Re\, m(x)| \le v^{-1/2} \sqrt{\Im\, m(z)}.$$

Proof (of Theorem 1.3.2 [1, p. 517]): Clearly

$$\begin{aligned}|\Re\, m(z)| &= \left| \int \frac{(x-u)\, \mathrm{d}F(x)}{(x-u)^2+v^2} \right| \\ &\le \int \frac{\mathrm{d}F(x)}{\sqrt{(x-u)^2+v^2}} \\ &\le \left(\int \frac{\mathrm{d}F(x)}{(x-u)^2+v^2} \right)^{1/2},\end{aligned}$$

and hence the result follows using the relation in (1.3.1). ∎

However, most importantly, a simple connection between the Stieltjes transform of the spectral distribution of a matrix and its eigenvalues exists [14, p. 2764]. It has to be noted that for a $m \times m$ matrix $\boldsymbol{A}$, and its corresponding spectral distribution Γ (see section 2), the Stieltjes transform is just

$$m_\Gamma(z) = \frac{1}{m}\text{trace}\left((\boldsymbol{A} - z\boldsymbol{I})^{-1}\right), \tag{1.3.2}$$

an equation that is extremely useful in further results in this project, [14, p. 2764]. Moreover, as given by Bai & Silverstein (see [1, p. 10]), applying the inverse formula from Theorem 1.3.1 to equation (1.3.2) gives

$$m_\Gamma(z) = \frac{1}{m}\sum_{k=1}^{m}\frac{1}{a_{kk} - z - \boldsymbol{\alpha}_k^\intercal(\boldsymbol{A}_k - z\boldsymbol{I})^{-1}\boldsymbol{\alpha}_k},$$

where $\boldsymbol{A}_k$ is the $(m-1)\times(m-1)$ matrix formed by removing the k^{th} row and column from $\boldsymbol{A}$, and $\boldsymbol{\alpha}_k$ is the k^{th} column vector of $\boldsymbol{A}$ with the k^{th} element removed.

Finally, the study in this project will be limited to compactly supported measures only, and hence for the main results, used in the investigation closely following Karoui (see [14]) work in section 4, it is important to mention five important properties of Stieltjes transforms of measures on $\mathbb{R}$. These are summarised in Theorem 1.3.3, below.

Theorem 1.3.3 (Important properties of Stieltjes transforms of measures on $\mathbb{R}$ [14, p. 2763], [8]) *Let $\mathbb{C}^+$ be the set of complex numbers with strictly positive imaginary part, then:*

1. *If G is a probability measure, $m_G(z) \in \mathbb{C}^+$ if $z \in \mathbb{C}^+$ and $\lim_{y\to\infty} -iy \times m_G(iy) = 1$.*
2. *If F and G are two measures, and if $m_F(z) = m_G(z)$, for all $z \in \mathbb{C}^+$, then $G = F$ almost everywhere.*
3. *If G_n is a sequence of probability measures and $m_{G_n}(z)$ has a (pointwise) limit $m(z)$ for all $z \in \mathbb{C}^+$, then there exists a probability measure G with Stieltjes transform $m_G = m$ if and only if $\lim_{y\to\infty} -iym(iy) = 1$. If it is the case, G_n converges weakly to G.*
4. *The same is true if the convergence happens only for an infinite sequence $\{z_i\}_{i=1}^{\infty}$ in $\mathbb{C}^+$ with a limit point in $\mathbb{C}^+$.*
5. *If x is a continuity point of the cumulative distribution function of G, then*

$$\frac{dG(x)}{dx} = \lim_{\epsilon\downarrow 0}\frac{1}{\pi}\Im\, m_G(x + i\epsilon).$$

Clearly, the point (5.) in Theorem 1.3.3 is equivalent to the inversion formula, given in Theorem 1.3.1. Moreover, the proof of the point (2.) in Theorem 1.3.3 follows trivially by the application of the inversion formula. The other points will be considered here as a fact, and their proof will not be given in this project; their proofs are given by Geronimo & Hill in [8, p. 54-58].

2 Spectral Analysis of High Dimensional Random Matrices

Let $\boldsymbol{A}$ be an $m \times m$ matrix with the corresponding eigenvalues λ_j, for $j = 1, 2, \ldots, m$, such that each eigenvalue is real. Associate the *empirical spectral distribution* of the matrix $\boldsymbol{A}$, denoted $F^{\boldsymbol{A}}$, with the vector of eigenvalues of the matrix $\boldsymbol{A}$ [14, p. 2762]. Define the following one-dimensional distribution function

$$F^{\boldsymbol{A}}(x) = \frac{1}{m}\#\{j \leq m : \lambda_j \leq x\}$$

to be the measure associated with the eigenvalues of the matrix $\boldsymbol{A}$, where $\#E$ denotes the cardinality of the set E [1, p. 4–5]. The above can be equivalently written as

$$F^{\boldsymbol{A}}(x) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}_{[\lambda_j \leq x]},$$

using the indicator function notation, where $\mathbb{I}_A$ is the indicator function of the event A [13, p. 6].

Hence, associate the corresponding measure

$$\mathrm{d}F^{\boldsymbol{A}}(x) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}_{[\lambda_j = x]},$$

which clearly has m point masses of equal weight.

For a given sequence of high dimensional[2] random matrices $\{\boldsymbol{A}_n\}$, the key focus is to investigate the convergence of the corresponding sequence of empirical spectral distributions $\{F^{\boldsymbol{A}_n}\}$ to the *limiting spectral distribution*, denoted by F [1, p. 5]. Note that, if $\lambda_j \to \infty$ for some $j = 1, 2, \ldots, m$, then the above limit distribution is defective in a sense that the total mass is not 1 [1, p. 5].

2.1 Sample Covaraince Matrices

In the field of multivariate statistical inference, the most important and extensively studied random matrix is the sample covariance matrix [1, p. 39]. For a given population of a fixed size, and number of taken samples tending to infinity, the sample covariance matrix provides a good and reliable approximate of the population covariance matrix [3, p. 1382].

That is, in a standard statistical setup, i.i.d. random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T \in \mathbb{R}^n$ are observed, where for $\boldsymbol{X}_t$, $t = 1, \ldots, T$, the corresponding covariance matrix is denoted by $\boldsymbol{\Sigma}$. The data matrix $\boldsymbol{X}$ is then defined as a $n \times T$ matrix containing the realisation of a random variable $\boldsymbol{X}_t$ in its t^{th} column, for $t = 1, \ldots, n$. Under the classical assumptions, value of n is fixed, while T is assumed to tend to infinity. In this context, the sample covariance matrix, defined below, is a good estimator of the population eigenvalues – that is the corresponding eigenvalues of the matrix $\boldsymbol{\Sigma}$ [14, p. 2758].

Let $\boldsymbol{X}_t \in \mathbb{R}^n$, for $t = 1, 2, \ldots, T$ be a set of vectors containing realisations taken from the underlying distribution $\boldsymbol{X}$ [13, p. 7]. Denote $\boldsymbol{X}_t = X_{it}$, for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, n$. Then the sample covariance matrix $\boldsymbol{S}$ is defined as

$$\boldsymbol{S} = \frac{1}{T-1}\sum_{t=1}^{T} \left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)^{\intercal}, \tag{2.1.1}$$

where the sample mean vector, $\bar{\boldsymbol{X}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_t$, is an estimator of the underlying population mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ [13, p. 7]. The covariance matrix defined in this way is an unbiased estimator of

[2] A sequence of matrices where the number of columns tends to infinity.

the population covariance matrix. Note that sometimes, to emphasise the sample size T, the sample covariance matrix is denoted by $\boldsymbol{S}_T$, instead of just $\boldsymbol{S}$.

For high dimensional sample covariance matrices it is possible to use T instead of $T-1$ in the denominator in the equation (2.1.1) without altering the result in a significant degree, since $T \approx T-1$ when T is large. Moreover, it is very common in spectral analysis of large dimensional random matrices to remove the terms $\bar{\boldsymbol{X}}$ for the equation (2.1.1), in order to calculate the sample covariance matrix as

$$\boldsymbol{S} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_t \, \boldsymbol{X}_t^{\intercal} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\intercal}. \tag{2.1.2}$$

The use of this simplified form of calculating the sample covariance matrix is justified by the fact that the matrix $\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^{\intercal}$ is of rank 1, and thus removal of the term $\bar{\boldsymbol{X}}$ does not affect the limiting spectral distribution[3] [1, p. 39].

Consider the case where X_{ij} are i.i.d. standard normal variables, and write

$$S_T = \left(\frac{1}{T}\sum_{k=1}^{T} X_{ik}X_{jk} \right)_{i,j=1}^{n},$$

which can be considered as a sample covariance matrix with T samples of a n-dimensional mean-zero random vector with population matrix equal to identity [1, p. 3]. Then, another important statistic, used in multivariate analysis, is

$$\mathcal{T}_T = \log(\det S_T) = \sum_{j=1}^{n} \log(\lambda_{T,j}),$$

where $\lambda_{T,j}$, for $j = 1, \ldots, n$, are the eigenvalues of S_T [1, p. 2-3]. Since for fixed n, $\lambda_{T,j} \to 1$ almost surely as $T \to \infty$, it follows that $\mathcal{T}_T \overset{a.s.}{\to} 0$ [1, p. 2-3]. Moreover, for any fixed n,

$$\sqrt{T/n}\,\mathcal{T}_T \overset{\mathcal{D}}{\to} N(0,2),$$

which can be seen by taking Taylor expansion on $\log(1+x)$, and which suggests that $\mathcal{T}_T$ is asymptotically normal given that $n = O(T)$ [1, p. 3]. However, using results on the limiting spectral distribution of $\{S_T\}$, considered further in this section, it is possible to show that if $n/T \to y \in (0,1)$ as $T \to \infty$, then

$$\sqrt{T/n}\,\mathcal{T}_T \sim \left(\frac{y-1}{y}\log(1-y) - 1 \right)\sqrt{Tn} \to -\infty, \quad a.s.,$$

as done by Bai & Silverstein in [1, p. 2-3]. This example illustrates the problem of using tests based on asymptotic normality assumption of $\mathcal{T}_T$, which will produce a large error if in fact $n/T \to y \in (0,1)$ when $T \to \infty$ [1, p. 3].

[3] The proof of this claim is given in Bai & Silverstein [1, p. 503], as an immediate consequence of theorem A.2.1 (see appendix A.2).

2.2 Wigner Matrices and Semicircular Law

Consider the case where $\{\boldsymbol{X}_t\}_{t=1}^{T}$ are i.i.d. samples taken from an n-dimensional multivariate normal population $N(\boldsymbol{\mu}, \boldsymbol{I})$, for some mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$. Let $\boldsymbol{S}$ be the sample covariance matrix, as defined in (2.1.1), then as T tends to infinity, $\boldsymbol{S} \to \boldsymbol{I}$ and $\sqrt{T}(\boldsymbol{S} - \boldsymbol{I}) \to \sqrt{n}\,\boldsymbol{W}_n$, where the entries above the main diagonal of $\sqrt{n}\,\boldsymbol{W}_n$ are i.i.d. standard normal $N(0,1)$ and the entries on the diagonal are i.i.d. normal $N(0,2)$ [1, p. 15]. Spectral analysis of matrices similar to $\sqrt{n}\,\boldsymbol{W}_n$, called the Wigner matrices (see definition 2.2.1 below), has been a major study in the field of random matrix theory.

Definition 2.2.1 (Wigner matrix [1, p. 15]) *Matrix $\sqrt{n}\,\boldsymbol{W}_n$, as defined above, is called the (standard) Wigner (or in some literature Gaussian) matrix. A Wigner matrix, in general, is a Hermitian random matrix with independent entries on and above the diagonal.*

Perhaps one of the key results established in regards to Wigner matrices is the *semicircular law*, given by Wigner in 1958. The result is stated as Theorem 2.2.1, below.

Theorem 2.2.1 (Wigner's Semicircular Law [1, p. 15]) *Let $\boldsymbol{A}$ be an $n \times n$ standard Wigner matrix, normalised by a factor of $n^{-1/2}$, then its expected empirical spectral distribution tends to the semicircular law G, where the density of G, denoted by $g(\cdot)$, is given by*

$$g(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & \text{if } |x| \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, the limiting density is a semicircle with radius 2, and note that it is non-random. To illustrate this, firstly note that if $\boldsymbol{A} = (A_{ij})$ is an $n \times n$ matrix with its elements being i.i.d. standard normal deviates, i.e.

$$A_{ij} \sim N(0,1), \quad i, j = 1, 2, \ldots, n,$$

where n is assumed to be large, then the matrix $\boldsymbol{H}_n$ such that

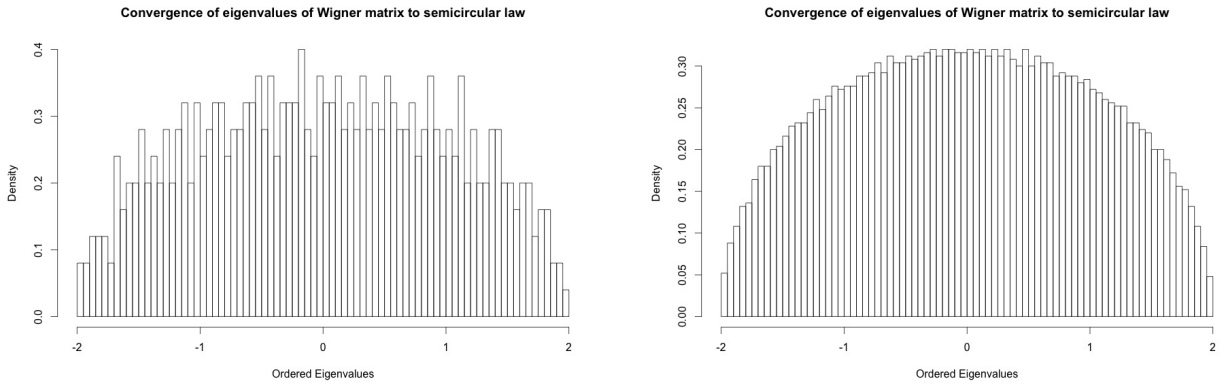
$$\boldsymbol{H}_n = \frac{1}{2}\left(\boldsymbol{A} + \boldsymbol{A}^{\intercal}\right),$$

has real entries and is symmetric ($\boldsymbol{H}_n = \boldsymbol{H}_n^{\intercal}$), and therefore is Hermitian [13, p. 4-5]. Moreover, by definition, matrix $\boldsymbol{H}_n$ has independent entries on and above the diagonal, so it is a Wigner matrix. Note that the ij^{th} entry of $\boldsymbol{H}_n = (H_{ij})$ is given by

$$H_{ij} \sim N(0, \sigma_{ij}), \quad \sigma_{ij} = \frac{1}{2}(1 + \delta_{ij}), \tag{2.2.1}$$

where δ_{ij} takes value of 1 if $i = j$ or 0 otherwise [13, p. 5], i.e. is Kronecker's delta. It is possible to easily simulate such matrix $\boldsymbol{H}_n$ in order to illustrate the convergence of the empirical spectral distribution to the semicircular law. Note that, in order for $\boldsymbol{H}_n$ to be standard Wigner matrix it has to be scaled by a factor of $\sqrt{2}$, which can be seen from (2.2.1) and from the fact that if $Z \sim N(0,1)$ then $cZ \sim N(0, c^2)$ for any constant $c \in \mathbb{R}$. The R code written to produce the resultant plots, given in Figure 2.2.1, uses the in-built `rnorm()` function to generate i.i.d. normal sample, while the eigenvalues of the simulated $\boldsymbol{H}_n$ matrix are obtained through the in-built `eigen()` function; the code is given in appendix B.1.

Figure 2.2.1 illustrates the fact that the (density) histogram of the ordered eigenvalues converges to a semicircle lying on the real axis in the range $[-2, 2]$. In the case when $n = 500$ the convergence is distinguishable but there is much irregularity and noise observed nonetheless. As n is increased 10-fold to 5000, the density histogram shows very clear pattern of a semicircle, empirically verifying the result firstly given by Wigner.



(a) Simulation with $n = 500$, bin size 100.

(b) Simulation with $n = 5000$, bin size 100.

Figure 2.2.1: Illustration of the convergence to Wigner's Semicircular Law, through histograms of the eigenvalues from a simulated $n \times n$ normalised Wigner matrix.

Further results following Wigner's semicircular law have been given in literature, where, for example, the convergence of $||F^{\boldsymbol{W}_n} - G|| \to 0$ has been shown to hold in probability [1, p. 15]. Bai & Silverstein have provided a general result that holds for non-i.i.d. case, shown in Theorem 2.2.2, below.

Theorem 2.2.2 (Generalisation of Wigner's Semicircular Law to non-i.i.d. case [1, p. 26]) *Let $\boldsymbol{W}_n = \frac{1}{\sqrt{n}}\boldsymbol{X}_n$ be a Wigner matrix, such that entries above or on the diagonal of $\boldsymbol{X}_n$ are independent but may be dependent on n, and are not necessarily identically distributed. Moreover, assume that all the entries of $\boldsymbol{X}_n = \left(x_{jk}^{(n)}\right)$ have zero mean and unit variance, and satisfy the condition that, for any constant $\eta > 0$,*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{j,k} \mathbb{E}\left(|x_{jk}^{(n)}|^2 \mathbb{I}_{\left[|x_{jk}^{(n)}| \geq \eta\sqrt{n}\right]} \right) = 0.$$

Then, the empirical spectral distribution of $\boldsymbol{W}_n$ converges to the semicircular law almost surely.

The proof of Theorem 2.2.2 through *moment convergence theorem* (MCT), as given by Bai & Silverstein in [1, p. 27-31], bases on 5 steps:

1. *Truncation*, where a new matrix $\widetilde{\boldsymbol{W}} = \frac{1}{\sqrt{n}} n \left(x_{ij}^{(n)} \mathbb{I}_{\left[|x_{ij}^{(n)}| \leq \eta_n \sqrt{n}\right]} \right)$ is formed for a selected sequence $\eta_n \downarrow 0$; then it can be shown that in order to prove convergence with probability one of $F^{\boldsymbol{W}_n}$ to the semicircular law, it is sufficient to show the convergence in probability of $F^{\widetilde{\boldsymbol{W}}_n}$ to the semicircular law.

2. *Diagonal elements removal*, where the matrix $\widehat{\boldsymbol{W}}_n$ is formed from the matrix $\widetilde{\boldsymbol{W}}_n$ by setting diagonal elements equal to 0; then it can be shown that $L^3\left(F^{\widetilde{\boldsymbol{W}}_n}, F^{\widehat{\boldsymbol{W}}_n}\right) \to 0$.

3. *Centralisation*, where it is shown that $L^3\left(F^{\widehat{\boldsymbol{W}}_n}, F^{\widehat{\boldsymbol{W}}_n - \mathbb{E}(\widehat{\boldsymbol{W}}_n)}\right) \to 0$.

4. *Rescaling*, where by writing $\widetilde{\boldsymbol{W}}_n = \frac{1}{\sqrt{n}}\widetilde{\boldsymbol{X}}_n$, for

$$\widetilde{\boldsymbol{X}}_n = \left(\frac{x_{ij}^{(n)} \mathbb{I}_{\left[|x_{ij}^{(n)}| \leq \eta_n \sqrt{n}\right]} - \mathbb{E}\left(x_{ij}^{(n)} \mathbb{I}_{\left[|x_{ij}^{(n)}| \leq \eta_n \sqrt{n}\right]} \right)}{\sigma_{ij}} (1 - \delta_{ij}) \right),$$

   it can be shown that $L^3\left(F^{\widetilde{\boldsymbol{W}}_n}, F^{\widehat{\boldsymbol{W}}_n - \mathbb{E}(\widehat{\boldsymbol{W}}_n)}\right) \to 0$, almost surely.

5. And finishing the proof using *Moment Convergence Theorem*[4], by which it is shown that $\mathbb{E}(\beta_k(\boldsymbol{W}_n))$ converges to the k^{th} moment β_k of the semicircular distribution, and that for each fixed k,

$$\sum_n \mathbb{E}\left|\beta_k(\boldsymbol{W}_n) - \mathbb{E}(\beta_k(\boldsymbol{W}_n))\right|^4 < \infty,$$

   where $\beta_k(\boldsymbol{W}_n)$ is the k^{th} moment of the empirical spectral distribution of $\boldsymbol{W}_n$, defined as $\beta_k(\boldsymbol{W}_n) = \beta_k(F^{\boldsymbol{W}_n}) = \int x^k \, \mathrm{d}F^{\boldsymbol{W}_n}(x)$.

Refer to Bai & Silverstein [1, p. 26-31] for explicit proofs for each of the aforementioned steps. Note that the proof through the Moment Convergence Theorem relies on the existence of moments and hence is not fully desired. As an alternative, following completion of steps 1-4, it is possible to complete the proof of Theorem 2.2.2 using Stieltjes transform of the semicircular law (refer to [1, p. 31-38] for the complete proof).

It should be noted that the Stieltjes transform for the semicircular law is given by [22, p. 7]

$$m(z) = -\frac{1}{2}(z - \sqrt{z^2 - 4}),$$

or, more generally, for a semicircular law that has been scaled[5] through a parameter σ^2, its Stieltjes transform is given by [**?**, p. 31-32]

$$m(z) = -\frac{1}{2\sigma^2}(z - \sqrt{z^2 - 4\sigma^2}),$$

as argued using the fact that by definition

$$m(z) = \frac{1}{2\sigma^2} \int_{-2\sigma}^{2\sigma} \frac{1}{x - z} \sqrt{4\sigma^2 - x^2} \, \mathrm{d}x,$$

which can be rewritten (letting $x = 2\sigma \cos y$) as

$$m(z) = -\frac{1}{4i\pi} \oint_{|\zeta|=1} \frac{(\zeta^2 - 1)^2}{\zeta^2(\sigma\zeta^2 + \sigma - z\zeta)} \mathrm{d}\zeta,$$

where $\zeta = e^{iy}$. Evaluating the residues, and then using the residue theorem to evaluate the integral yields the result (see [**?**, p. 31-32] for the extensive proof); where by convention the square root of a complex number is taken to be the one with positive imaginary part.

[4] See appendix A.6 for the statement of the Moment Convergence Theorem.

[5] The density of the scaled semicircular law is given by

$$g(x) = \begin{cases} \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}, & \text{if } |x| \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

2.3 Marčenko-Pastur Law

The Marčenko-Pastur law, perhaps the key result of random-matrix theory for use in multivariate analysis, can be regarded as an analogue of Wigner's semicircular law [13, p. 10]. It provides a non-random limiting distribution of the eigenvalues of a $n \times n$, say, sample covariance matrix when $n \to \infty$. Therefore, although the standard estimator $\boldsymbol{S}$, given in equation (2.1.1), of the sample covariance matrix provides a good approximation to the true $n \times n$, say, population covariance matrix $\boldsymbol{\Sigma}$, when n is fixed and finite, it is not true in the case when n is large, and not fixed (tends to infinity).

Let y be the dimension to sample size ratio index. Using the notation from section 2.1, y is defined as the ratio of n/T. Then, for $y \leq 1$, the *Marčenko-Pastur* (M-P) *law*, denoted by $F_y(x)$, has a density function given by

$$f_y(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)}, & \text{for } a \leq x \leq b, \\ 0, & \text{otherwise,} \end{cases} \tag{2.3.1}$$

where $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$; and for $y > 1$ it has a point mass $1 - 1/y$ at the origin [1, p. 40]. Here, σ^2 denotes the scale parameter, where the standard Marčenko-Pastur law is defined for $\sigma^2 = 1$ [1, p. 40].

It has been shown that for a wide class of sample covariance matrices, the corresponding empirical spectral distribution converges to asymptotically non-random result [14, p. 2763]. For the special case where the population covariance matrix, $\boldsymbol{\Sigma}$, is an identity matrix multiplied by some positive constant $\sigma^2 \in \mathbb{R}^+$, the Marčenko-Pastur law describes the limiting behaviour of the empirical spectral distribution. The more general result is then considered in section 2.5.

Theorem 2.3.1 (Convergence to Marčenko-Pastur law [1, p. 47]) *Let $\{x_{it}\}$, for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, n$, be i.i.d. complex random variables with variance σ^2. Assume that $n/T \to y \in (0, \infty)$. Then the empirical spectral distribution of sample covariance matrix, $F^{\boldsymbol{S}}$, tends to Marčenko-Pastur law, defined in (2.3.1), with probability one.*

Theorem 2.3.1 considers the limiting spectral distribution of the sample covariance matrix with the underlying variable coming from an i.i.d. population. This result can be also obtained using Theorem 2.5.1, stated by Karoui in [14], in a setting where the population eigenvalues are 1.

Yin [24, p. 50-67], has given a proof of the result similar to Theorem 2.3.1 (only for real complex variables with mean zero), which bases on a truncation technique and sophisticated combinatorial techniques, requiring an extensive discussion on results in graph theory. Bai & Silverstein [1], have then provided an extension to that result, denoted here as Theorem 2.3.1, which bases on two main steps: firstly doing truncation, centralisation and rescaling, and then working with moments of the Marčenko-Pastur law, using results from graph theory and combinatorics. For the extensive proof see Bai & Silverstein [1, p. 48–50].

That said, as in the case of the convergence to Wigner's semicircular law, (an extension of the) result shown in Theorem 2.3.1 can also be proven using Stieltjes transform. This is considered in section 2.3.2, which makes use of the truncation, centralisation and rescaling step discussed next.

2.3.1 Truncation, Centralisation, and Rescaling Technique

A similar idea of application of the truncation, centralisation and rescaling technique has already been mentioned in regards to proving Theorem 2.2.2. In the Marčenko-Pastur setting, it provides theoretical grounds for assumption that variables x_{it} are uniformly bounded with mean zero and variance 1. This technique has been stated by Bai & Silverstein [1, p. 48], and shall be closely reproduced in this section.

Firstly, recall that n and T are assumed to tend to infinity. Now, fix a positive number C, and define

$$\begin{aligned}
\widehat{x}_{it} &= x_{it}\,\mathbb{I}_{[|x_{it}|\leq C]},\\
\widetilde{x}_{it} &= \widehat{x}_{it} - \mathbb{E}(\widehat{x}_{11}),\\
\widehat{\boldsymbol{x}}_t &= (\widehat{x}_{t1},\ldots,\widehat{x}_{tn})',\\
\widetilde{\boldsymbol{x}}_t &= (\widetilde{x}_{t1},\ldots,\widetilde{x}_{tn})',\\
\widehat{\boldsymbol{S}}_T &= \frac{1}{T}\sum_{i=1}^{T}\widehat{\boldsymbol{x}}_i\widehat{\boldsymbol{x}}_i^{\intercal} = \frac{1}{T}\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^{\intercal},\\
\widetilde{\boldsymbol{S}}_T &= \frac{1}{T}\sum_{i=1}^{T}\widetilde{\boldsymbol{x}}_i\widetilde{\boldsymbol{x}}_i^{\intercal} = \frac{1}{T}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\intercal},
\end{aligned}$$

where notation $\boldsymbol{A}'$ means conjugate transpose of $\boldsymbol{A}$.

Denote the empirical spectral distributions of $\widehat{\boldsymbol{S}}_T$ and $\widetilde{\boldsymbol{S}}_T$ by $F^{\widehat{\boldsymbol{S}}_T}$ and $F^{\widetilde{\boldsymbol{S}}_T}$, respectively. Using Theorem A.1.1 and the strong law of large numbers, it follows that

$$\begin{aligned}
L^4(F^{\boldsymbol{S}}, F^{\widehat{\boldsymbol{S}}_T}) &\leq \left(\frac{2}{Tn}\sum_{i,t}(|x_{it}^2| + |\widehat{x}_{it}^2|)\right)\left(\frac{2}{Tn}\sum_{i,t}(|x_{it}-\widehat{x}_{it}|^2)\right)\\
&\leq \left(\frac{4}{Tn}\sum_{i,t}|x_{it}^2|\right)\left(\frac{1}{Tn}\sum_{i,t}\left(|x_{it}^2|\,\mathbb{I}_{[|x_{it}|>C]}\right)\right),
\end{aligned}$$

and so

$$L^4(F^{\boldsymbol{S}}, F^{\widehat{\boldsymbol{S}}_T}) \to 4\mathbb{E}\left(|x_{it}^2|\,\mathbb{I}_{[|x_{it}|>C]}\right), \quad \text{a.s.} \tag{2.3.2}$$

By choosing C large enough, the right-hand side of equation (2.3.2) can be made arbitrarily small.

Moreover, using Theorem A.2.1, it follows that

$$||F^{\widehat{\boldsymbol{S}}_T} - F^{\widetilde{\boldsymbol{S}}_T}|| \leq \frac{1}{n}\text{rank}\left[\mathbb{E}(\widehat{\boldsymbol{X}})\right] = \frac{1}{n}. \tag{2.3.3}$$

Let $\widetilde{\sigma}^2 = \mathbb{E}(|\widetilde{x}_{it}|^2) \to 1$, as $C \to \infty$. Using Theorem A.1.1, it follows that

$$L^4(F^{\widetilde{\boldsymbol{S}}_T}, F^{\widetilde{\sigma}^{-2}\widetilde{\boldsymbol{S}}_T}) \leq 2\left(\frac{1+\widetilde{\sigma}^2}{Tn\widetilde{\sigma}^2}\sum_{i,t}|\widetilde{x}_{it}|^2\right)\left(\frac{1-\widetilde{\sigma}^2}{Tn\widetilde{\sigma}^2}\sum_{i,t}|\widetilde{x}_{it}|^2\right),$$

and so

$$L^4(F^{\widetilde{\boldsymbol{S}}_T}, F^{\widetilde{\sigma}^{-2}\widetilde{\boldsymbol{S}}_T}) \to 2(1-\widetilde{\sigma}^4), \quad \text{a.s.} \tag{2.3.4}$$

Again, by choosing C large enough, the right-hand side of equation (2.3.4) can be made arbitrarily small, due to the fact that $\widetilde{\sigma}^2 \to 1$ as $C \to \infty$, and hence so does $\widetilde{\sigma}^4$.

By combining equations (2.3.2), (2.3.3) and (2.3.4), it has been shown that indeed variables x_{it} are uniformly bounded with mean zero and variance 1, which concludes this section.

2.3.2 Generalisation to Not Identically Distributed Case

It has been further possible, using the Stieltjes transform approach, to generalise the previous result, stated in Theorem 2.3.1, to the case in which the entries of $\boldsymbol{X}_t$ depend on t, and for each t they are independent, but not identically distributed [1, p. 51]. The result given by Bai & Silverstein is given in Theorem 2.3.2, using the established previously notation.

Theorem 2.3.2 (Convergence to Marčenko-Pastur law for not identically distributed random variables [1, p. 51]) *Let for each t the entries of $\boldsymbol{X}$ be independent complex variables with a common mean μ and variance σ^2. Assume that $n/T \to y \in (0, \infty)$, and that, for any $\eta > 0$,*

$$\frac{1}{\eta^2 nt} \sum_{j,k} \mathbb{E}\left(|x_{jk}^{(n)}|^2 \mathbb{I}_{\left[|x_{jk}^{(n)}| \geq \eta\sqrt{n}\right]} \right) \to 0. \tag{2.3.5}$$

Then, with probability one, $F^{\boldsymbol{S}}$ tends to Marčenko-Pastur law with ratio index y and scale index σ^2.

As already mentioned, there are different methods of proving Theorem 2.3.2, however, in this project an illustration of a sketch proof through an application of Stieltjes transforms to sample covariance matrices will be shown (in contrast to using the Moment Convergence Theorem).

Firstly, in order to proof[6] Theorem 2.3.2 using Stieltjes transform, the transform of the Marčenko-Pastur law has to be established; this result and a sketch proof is shown below, as Lemma 2.3.3.

Throughout the section, let $z = u + iv$ be such that $v > 0$, and denote the Stieltjes transform of the Marčenko-Pastur law by $m(z)$.

Lemma 2.3.3 (Stieltjes Transform of the Marčenko-Pastur Law [1, p. 52]) *Using previously established notation,*

$$m(z) = \frac{\sigma^2(1-y) - z + \sqrt{(z - \sigma^2 - y\sigma^2)^2 - 4y\sigma^4}}{2yz\sigma^2}. \tag{2.3.6}$$

Proof (of Lemma 2.3.3, sketch version [1, p. 52-53]): For $y < 1$, by definition

$$m(z) = \int_a^b \frac{1}{x - z} \frac{1}{2\pi x y \sigma^2} \sqrt{(b-x)(x-a)}\, \mathrm{d}x,$$

where $a = \sigma^2(1 - \sqrt{y})^2$ and $b = \sigma^2(1 + \sqrt{y}^2)$.

Now, let $x = \sigma^2(1 + y + 2\sqrt{y}\cos w$ and set $\zeta = e^{iw}$, which allows the Stieltjes transform of the Marčenko-Pastur law to be written as

$$\begin{aligned}
m(z) &= \int_0^\pi \frac{2}{\pi} \frac{1}{(1 + y + 2\sqrt{y}\cos w)(\sigma^2(1 + y + 2\sqrt{y}\cos w) - z)} \sin^2 w \, \mathrm{d}w \\
&= \frac{1}{\pi} \int_0^{2\pi} \frac{((e^{iw} - e^{-iw})/2i)^2}{(1 + y + \sqrt{y}(e^{iw} + e^{-iw}))(\sigma^2(1 + y + \sqrt{y}(e^{iw} + e^{-iw})) - z)} \mathrm{d}w \\
&= -\frac{1}{4i\pi} \oint_{|\zeta|=1} \frac{(\zeta - \zeta^{-1})^2}{\zeta(1 + y + \sqrt{y}(\zeta + \zeta^{-1}))(\sigma^2(1 + y + \sqrt{y}(\zeta + \zeta^{-1})) - z)} \mathrm{d}\zeta \\
&= -\frac{1}{4i\pi} \oint_{|\zeta|=1} \frac{(\zeta^2 - 1)^2}{\zeta((1+y)\zeta + \sqrt{y}(\zeta^2 + 1))(\sigma^2(1+y)\zeta + \sqrt{y}\sigma^2(\zeta^2 + 1) - z\zeta)} \mathrm{d}\zeta.
\end{aligned}$$

[6] All theorems and proofs given in this section reproduce very closely results given by Bai & Silverstein [1].

There are five simple poles, of the integrand function above, at

$$\begin{aligned}
\zeta_0 &= 0,\\
\zeta_1 &= \frac{-(1+y)+(1-y)}{2\sqrt{y}},\\
\zeta_2 &= \frac{-(1+y)-(1-y)}{2\sqrt{y}},\\
\zeta_3 &= \frac{-\sigma^2(1+y)+z+\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}}{2\sigma^2\sqrt{y}},\\
\zeta_4 &= \frac{-\sigma^2(1+y)+z-\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}}{2\sigma^2\sqrt{y}}.
\end{aligned}$$

It can be shown, after a calculation, that the residues at these five poles are given by

$$\frac{1}{y\sigma^2}, \quad \mp\frac{1-y}{yz}, \quad \text{and} \quad \pm\frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}.$$

Observe that $\zeta_3\zeta_4 = 1$. Also, by the definition of the square root of complex numbers, the real part and imaginary part of $\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}$ and $-\sigma^2(1+y)+z$ have the same signs and thus $|\zeta_3| > 1$, $|\zeta_4| < 1$. Similarly, $|\zeta_1| = |-\sqrt{y}| < 1$, and $|\zeta_2| = |-1/\sqrt{y}| > 1$. Cauchy integration, therefore, gives

$$\begin{aligned}
m(z) &= -\frac{1}{2}\left(\frac{1}{y\sigma^2}-\frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}-\frac{1-y}{yz}\right)\\
&= \frac{\sigma^2(1-y)-z+\sqrt{(z-\sigma^2-y\sigma^2)^2-4y\sigma^4}}{2yz\sigma^2},
\end{aligned}$$

consequently proving Lemma 2.3.3 for the case of $y < 1$.

When $y > 1$, as the Marčenko-Pastur law has also a point mass $1-\frac{1}{y}$ at zero, its Stieltjes transform, $m(z)$, is equal to the integral above plus $-(y-1)/yz$. Note that, now, $|\zeta_1| = |-\sqrt{y}| > 1$ and $|\zeta_2| = |-1/\sqrt{y}| < 1$, so the residue at ζ_2 should be counted into integral. After some elementary calculations, similar to the $y < 1$ case, it can be shown that indeed equation given by Lemma 2.3.3 holds when $y > 1$.

Finally, for the case $y = 1$, it can be argued that quation given by Lemma 2.3.3 holds by continuity in y. This, hence, concludes the proof. ∎

In order to prove Theorem 2.3.2, as done by Bai & Silverstein in [1, p. 52-58], it has to be noted that by using the truncating, centralising and rescaling approach (see section 2.3.1), it is possible to further assume that:

1. $|x_{it}| < \eta_T\sqrt{T}$, where a sequence $\eta_T \downarrow 0$ is selected so that the condition (2.3.5) holds true for η replaced by η_T.
2. $\mathbb{E}(x_{it}) = 0$ and $\mathrm{Var}(x_{it}) = 1$.

Proof (of Theorem 2.3.2, sketch version [1, p. 53-58]): For clarity of notation, denote $\boldsymbol{S}_T$ to be the sample covariance matrix of data $\boldsymbol{X}$, previously denoted as $\boldsymbol{S}$. Let the Stieltjes transform of the empirical spectral distribution of $\boldsymbol{S}_T$ be denoted by $m_T(z)$. As in (1.3.2), define

$$m_T(z) = \frac{1}{n}\mathrm{trace}\left((\boldsymbol{S}_T - z\boldsymbol{I})^{-1}\right),$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix.

Then, the proof is completed by three steps:

1. For any fixed $z \in \mathbb{C}^+$, $m_T(z) - \mathbb{E}\left(m_T(z)\right) \to 0$, a.s.
2. For any fixed $z \in \mathbb{C}^+$, $\mathbb{E}\left(m_T(z)\right) \to m(z)$, the Stieltjes transform of the Marčenko-Pastur law.
3. Except for a null set, $m_T(z) \to m(z)$ for every $z \in \mathbb{C}^+$.

Note that step 3 is being implied by the previous two steps, and hence its proof can be omitted.

Step 1: almost sure convergence of the random part: The aim of the first step is to prove that

$$m_T(z) - \mathbb{E}\left(m_T(z)\right) \to 0, \quad \text{a.s.} \tag{2.3.7}$$

Let $\mathbb{E}_k(\cdot)$ denote the conditional expectation given $\{\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_T\}$. Then, using the formula

$$(\boldsymbol{A} + \boldsymbol{\alpha}\boldsymbol{\beta}^\intercal)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}^\intercal\boldsymbol{A}^{-1}}{1 + \boldsymbol{\beta}^\intercal\boldsymbol{A}^{-1}\boldsymbol{\alpha}},$$

it follows that

$$\begin{aligned} m_T(z) - \mathbb{E}\left(m_T(z)\right) &= \frac{1}{n}\sum_{k=1}^{T}\left[\mathbb{E}_k\left(\operatorname{trace}\left((\boldsymbol{S}_T - z\boldsymbol{I})^{-1}\right)\right) - \mathbb{E}_{k-1}\left(\operatorname{trace}\left((\boldsymbol{S}_T - z\boldsymbol{I})^{-1}\right)\right)\right] \\ &= \frac{1}{n}\sum_{k=1}^{n}\gamma_k, \end{aligned}$$

where, by Theorem A.3.1,

$$\begin{aligned} \gamma_k &= (\mathbb{E}_k - \mathbb{E}_{k-1})\left[\operatorname{trace}\left((\boldsymbol{S}_T - z\boldsymbol{I})^{-1}\right) - \operatorname{trace}\left((\boldsymbol{S}_{Tk} - z\boldsymbol{I})^{-1}\right)\right] \\ &= -(\mathbb{E}_k - \mathbb{E}_{k-1})\frac{\boldsymbol{x}_k^\intercal\left(\boldsymbol{S}_{Tk} - z\boldsymbol{I}\right)^{-2}\boldsymbol{x}_k}{1 + \boldsymbol{x}_k^\intercal\left(\boldsymbol{S}_{Tk} - z\boldsymbol{I}\right)^{-1}\boldsymbol{x}_k}, \end{aligned}$$

and $\boldsymbol{S}_{Tk} = \boldsymbol{S}_T - \boldsymbol{x}_k\boldsymbol{x}_k^\intercal$. Furthermore, it should be noted that

$$\left|\frac{\boldsymbol{x}_k^\intercal\left(\boldsymbol{S}_{Tk} - z\boldsymbol{I}\right)^{-2}\boldsymbol{x}_k}{1 + \boldsymbol{x}_k^\intercal\left(\boldsymbol{S}_{Tk} - z\boldsymbol{I}\right)^{-1}\boldsymbol{x}_k}\right| \leq \frac{\boldsymbol{x}_k^\intercal((\boldsymbol{S}_{Tk} - u\boldsymbol{I})^2 + v^2\boldsymbol{I})^{-1}\boldsymbol{x}_k}{\Im(1 + \boldsymbol{x}_k^\intercal(\boldsymbol{S}_{Tk} - z\boldsymbol{I})^{-1}\boldsymbol{x}_k)} = \frac{1}{v}.$$

Since $\{\gamma_k\}$ forms a sequence of bounded martingale differences, by Theorem A.5.1 with $p = 4$, it follows that

$$\begin{aligned} \mathbb{E}|m_T(z) - \mathbb{E}\left(m_T(z)\right)|^4 &\leq \frac{K_4}{n^4}\mathbb{E}\left(\sum_{k=1}^{n}|\gamma_k|^2\right)^2 \\ &\leq \frac{4K_4T^2}{v^4n^4} = O(T^{-2}), \end{aligned}$$

which combined with Borel-Cantelli lemma, implies that (2.3.8) holds, completing the proof of the first step.

Step 2: mean convergence: The aim of the second step is to prove that

$$\mathbb{E}\left(m_T(z)\right) \to m(z), \tag{2.3.8}$$

with $m(z)$ as defined in (2.3.6), and $\sigma^2 = 1$.

It can be shown, by Theorem A.4.1, that

$$m_T(z) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\frac{1}{T}\boldsymbol{\alpha}_k'\bar{\boldsymbol{\alpha}}_k - z - \frac{1}{T^2}\boldsymbol{\alpha}_k'\boldsymbol{X}_k^\intercal\left(\frac{1}{T}\boldsymbol{X}_k\boldsymbol{X}_k^\intercal - z\boldsymbol{I}\right)^{-1}\boldsymbol{X}_k\bar{\boldsymbol{\alpha}}_k}, \tag{2.3.9}$$

where $\boldsymbol{I}$ is the corresponding $(n-1)\times(n-1)$ identity matrix, $\boldsymbol{X}_k$ is the matrix obtained from $\boldsymbol{X}$ by removing the k^{th} row, and $\boldsymbol{\alpha}_k'$ is the k^{th} row of $\boldsymbol{X}$.

Set

$$\epsilon_k = \frac{1}{T}\boldsymbol{\alpha}_k'\bar{\boldsymbol{\alpha}}_k - 1 - \frac{1}{T^2}\boldsymbol{\alpha}_k'\boldsymbol{X}_k^\intercal\left(\frac{1}{T}\boldsymbol{X}_k\boldsymbol{X}_k^\intercal - z\boldsymbol{I}\right)^{-1}\boldsymbol{X}_k\bar{\boldsymbol{\alpha}}_k + y_T + y_T z\,\mathbb{E}(m_T(z)),$$

where $y_T = n/T$. Then, using (2.3.9), it follows that

$$\mathbb{E}(m_T(z)) = \frac{1}{1 - z - y_T - y_T z\,\mathbb{E}(m_T(z))} + \delta_T, \tag{2.3.10}$$

where

$$\delta_T = -\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\frac{\epsilon_k}{(1 - z - y_T - y_T z\,\mathbb{E}(m_T(z)))(1 - z - y_T - y_T z\,\mathbb{E}(m_T(z)) + \epsilon_k)}\right).$$

Using equation (2.3.10) to solve for $\mathbb{E}(m_T(z))$, yields two solutions

$$m_1(z) = \frac{1}{2y_T z}\left(1 - z - y_T + y_T z\delta_T + \sqrt{(1 - z - y_T - y_T z\delta_T)^2 - 4y_T z}\right),$$
$$m_2(z) = \frac{1}{2y_T z}\left(1 - z - y_T + y_T z\delta_T - \sqrt{(1 - z - y_T - y_T z\delta_T)^2 - 4y_T z}\right),$$

By comparison with the Stieltjes transform of the Marčenko-Pastur law, given in (2.3.6), in order to prove the second step, it hence suffices to show that

$$\mathbb{E}(m_T(z)) = m_1(z), \tag{2.3.11}$$

and

$$\delta_T \to 0. \tag{2.3.12}$$

Starting with the proof of (2.3.11), observe that as $v \to \infty$, $\mathbb{E}(m_T(z)) \to 0$, and hence also $\delta_T \to 0$ by (2.3.10). This shows that $\mathbb{E}(m_T(z)) = m_1(z)$ for all z such that their imaginary part is large.

Now assume that $\mathbb{E}(m_T(z)) = m_1(z)$ does not hold for all $z \in \mathbb{C}^+$. Then, using continuity of m_1 and m_2 argument, there exists a $z_0 \in \mathbb{C}^+$ such that $m_1(z_0) = m_2(z_0)$, which implies that

$$(1 - z_0 - y_T + y_T z\delta_T)^2 - 4y_T z_0(1 + \delta_T(1 - z_0 - y_T)) = 0.$$

Therefore

$$\mathbb{E}(m_T(z)) = m_1(z_0) = \frac{1 - z_0 - y_T + y_T z_0\delta_T}{2y_T z_0}.$$

Substitution of the solution δ_T of equation (2.3.10) into the above identity, gives

$$\mathbb{E}(m_T(z_0)) = \frac{1 - z_0 + y_T}{y_T z_0} + \frac{1}{y_T + z_0 - 1 + y_T z_0 \, \mathbb{E}(m_T(z_0))}. \tag{2.3.13}$$

Note that for any Stieltjes transform $m(z)$ of probability F defined on positive real axis, $\mathbb{R}^+$, and for any positive y, the following result holds

$$\Im(y + z - 1 + yzm(z)) = \Im\left(z - 1 + \int_0^\infty \frac{yx \, \mathrm{d}F(x)}{x - z}\right) = v\left(1 + \int_0^\infty \frac{yx \, \mathrm{d}F(x)}{(x-u)^2 + v^2}\right) > 0. \tag{2.3.14}$$

Based on (2.3.14), it follows that the imaginary part of the second term in (2.3.13) is negative. For $y_T \leq 1$, it can be seen that $\Im(1 - z_0 - y_T)/(y_T z_0) < 0$, and hence it can be concluded that $\Im \, \mathbb{E}(m_T(z_0)) < 0$, which is a contradiction, as the imaginary part of the Stieltjes transform is positive. Hence, this argument proves, by contradiction, (2.3.11) for the case when $y_T \leq 1$.

In order to prove the statement (2.3.11) for the case when $y_T > 1$, note that by (2.3.13) and (2.3.14), it follows that

$$y_T + z_0 - 1 + y_T z_0 \, \mathbb{E}(m_T(z_0)) = \sqrt{y_T z_0}. \tag{2.3.15}$$

Now, let $\widetilde{m}_T(z)$ be the Stieltjes transform[7] of the matrix $\frac{1}{T}\boldsymbol{X}^\intercal \boldsymbol{X}$. Then, the relation between m_T and $\widetilde{m}_T$, noting that $\frac{1}{T}\boldsymbol{X}^\intercal \boldsymbol{X}$ and $\boldsymbol{S}_T = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^\intercal$ have the same set of nonzero eigenvalues, is given by

$$s_T(z) = y_T^{-1} \widetilde{m}_T(z) - \frac{1 - y_T^{-1}}{z},$$

which holds both when $y_T > 1$ and $y_T \leq 1$. Hence, using the above relation, it follows that

$$y_T - 1 + y_T z_0 \, \mathbb{E}(m_T(z_0)) = z_0 \, \mathbb{E}(\widetilde{m}_T(z_0)),$$

which substituted into (2.3.15) gives

$$1 + \mathbb{E}(\widetilde{m}_T(z_0)) = \frac{\sqrt{y}}{\sqrt{z_0}},$$

leading to a contradiction that the imaginary part of left-hand side is positive, while the imaginary part of the right-hand size is negative.

This concludes the proof of (2.3.11).

Turning to the proof of (2.3.12), rewrite

$$\begin{aligned}
\delta_T = &- \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\mathbb{E}(\epsilon_k)}{(1 - z - y_T - y_T z \, \mathbb{E}(m_T(z)))^2}\right) \\
&+ \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\frac{\epsilon_k^2}{(1 - z - y_T - y_T z \, \mathbb{E}(m_T(z)))^2(1 - z - y_T - y_T z \, \mathbb{E}(m_T(z)) + \epsilon_k)}\right) \\
= &\, J_1 + J_2.
\end{aligned}$$

Firstly, by the assumption made that $\mathbb{E}(x_{it}) = 0$ and $\mathrm{Var}(x_{it}) = 1$, it should be noted that it is possible to show (see [1, p. 57]) that

$$|\mathbb{E}(\epsilon_k)| \leq \frac{1}{T} + \frac{|z| \, y_T}{Tv} \to 0,$$

[7]Note, in section 2.5, the notation from $\widetilde{m}(z)$ is changed to $v(z)$.

implying that $J_1 \to 0$.

Moreover, it is possible to show that $J_2 \to 0$. Notice that, since

$$\begin{aligned}
&\Im(1 - z - y_T - y_T z\,\mathbb{E}(m_T(z)) + \epsilon_k) \\
&= \Im\left(\frac{1}{T}\boldsymbol{\alpha}_k'\bar{\boldsymbol{\alpha}}_k - z - \frac{1}{T^2}\boldsymbol{\alpha}_k'\boldsymbol{X}_k^\intercal\left(\frac{1}{T}\boldsymbol{X}_k\boldsymbol{X}_k^\intercal - z\boldsymbol{I}\right)^{-1}\boldsymbol{X}_k\bar{\boldsymbol{\alpha}}_k\right) \\
&= -v\left(1 + \frac{1}{T^2}\boldsymbol{\alpha}_k'\boldsymbol{X}_k^\intercal\left[\left(\frac{1}{T}\boldsymbol{X}_k\boldsymbol{X}_k^\intercal - u\boldsymbol{I}\right)^2 + v^2\boldsymbol{I}\right]^{-1}\boldsymbol{X}_k\bar{\boldsymbol{\alpha}}_k\right) < -v,
\end{aligned}$$

combining this with (2.3.14), gives

$$\begin{aligned}
|J_2| &\le \frac{1}{nT^3}\sum_{k=1}^{n}\mathbb{E}|\epsilon_k|^2 \\
&= \frac{1}{nT^3}\sum_{k=1}^{n}\left[\mathbb{E}|\epsilon_k - \widetilde{\mathbb{E}}(\epsilon_k)|^2 + \mathbb{E}|\widetilde{\mathbb{E}}(\epsilon_k) - \mathbb{E}(\epsilon_k)|^2 + (\mathbb{E}(\epsilon_k))^2\right],
\end{aligned}$$

where $\mathbb{E}(\cdot)$ denotes the conditional expectation given $\{\boldsymbol{\alpha}_j, j = 1, \ldots, k-1, k+1, \ldots, n\}$. Now, recall that

$$|\mathbb{E}(\epsilon_k)| \le \frac{1}{T} + \frac{|z|\,y}{Tv} \to 0.$$

Moreover, it is possible to show (see [1, p. 58]) that

$$\frac{1}{T^2}\widetilde{\mathbb{E}}|\epsilon_k' - \widetilde{\mathbb{E}}(\epsilon_k)|^2 \le \frac{\eta_T^2}{v^2} + \frac{2}{Tv^2},$$

for the sequence η_T defined previously, as well as (see [1, p. 58]) that

$$\mathbb{E}|\widetilde{\mathbb{E}}(\epsilon_k) - \mathbb{E}(\epsilon_k)|^2 \le \frac{|z|^2y^2}{Tv^2} \to 0.$$

Combining the three estimation above completes the proof of the mean convergence of the Stieltjes transform of the empirical spectral distribution of $\boldsymbol{S}_n$.

Consequently, Theorem 2.3.2 is proved by the method of Stieltjes transforms. ∎

2.4 Limits of Eigenvalues of a Large Dimensional Sample Covariance Matrix

The limiting behaviour of the spectral distribution of a sample covariance matrix was studied in the section 2.3. The almost sure convergence of the empirical spectral distribution has been shown for a sample covariance matrix, given that random variable $\boldsymbol{X}_{11}$ has a finite second moment[8] (variance). However, given more rigorous constraints on the moments, it is possible to establish almost sure bounds on the smallest and largest eigenvalues. This is shown in the following two theorems.

Theorem 2.4.1 (Bai-Yin Theorem [2, p. 1276]) *Let $\{\boldsymbol{X}_{it}\}$, for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, n$, be i.i.d. random variables with zero mean and unit variance, and define for $\boldsymbol{X} = (X_{it})$ the matrix*

$$\boldsymbol{S} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\intercal}.$$

Then, if $\mathbb{E}\,|X_{11}|^4 < \infty$, as $T \to \infty$, $n \to \infty$, and $n/T \to y \in (0, 1)$, the following inequality holds almost surely:

$$-2\sqrt{y} \leq \liminf \lambda_{\min}(\boldsymbol{S} - (1+y)\boldsymbol{I}) \leq \limsup \lambda_{\max}(\boldsymbol{S} - (1+y)\boldsymbol{I}) \leq 2\sqrt{y},$$

where $\boldsymbol{I}$ is the identity matrix, and $\lambda(\cdot)$ denotes the eigenvalues of the expression in $(\cdot)$.

An immediate result based on the Theorem 2.4.1 is as follows.

Theorem 2.4.2 (Extension to Bai-Yin Theorem [2, p. 1276]) *Under the conditions of Theorem 2.4.1, as $T \to \infty$, $n \to \infty$, and $n/T \to y \in (0, 1)$, the following inequalities holds almost surely:*

$$\lim \lambda_{\min} = (1 - \sqrt{y})^2$$

$$\lim \lambda_{\max} = (1 + \sqrt{y})^2$$

The proof of Theorem 2.4.1 (and hence of its immediate result stated in Theorem 2.4.2) is given by Bai & Yin (see [2]). To establish the result few intermediate lemmas are required, which will not be stated in this project.

In particular, Theorem 2.4.2 implies that the estimator for the largest (or, also, the smallest) eigenvalue is not consistent. Hence, under the assumption of *large T, large n* asymptotics, there are some fundamental differences in the multivariate statistics behaviour, example of which can be seen through the above theorem [14, p. 2758].

In the case when the population covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{I}$, i.e. is the identity matrix, implying i.i.d. random variables, and both T and n tend to infinity, the largest sample eigenvalue is biased, which in some cases can be in a very significant degree, as for example if the ratio $n/T \to y$, where y is approximately 1, then in the limit $\lambda_{\max} \approx 4$, while the true largest eigenvalue is just 1. This is a drastic 4-fold difference.

Therefore, it is important to seek methods for accounting or correcting this bias, in order to be able to successfully conduct data analysis. Karoui in [14] has used random matrix theory, building upon Marčenko-Pastur result, to aid statisticians in retrieving meaningful information under the *large T, large n* asymptotics framework; the proposed approach will be studied in this project in section 4. The key element of which is to provide an accurate estimate of the population spectral distribution, that is a probability measure responsible for the characterisation of the population eigenvalues.

[8] Since $\boldsymbol{X}_{it}$, for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, n$, are i.i.d. random variables, if one of them has a finite variance then all of them do.

2.5 Marčenko-Pastur Equation

Let $H^{\boldsymbol{\Sigma}}$ denote the spectral distribution of the $m \times m$ population covariance matrix, $\boldsymbol{\Sigma}$. Analogically to the empirical spectral distribution, define the *population spectral distribution* measure as

$$\mathrm{d}H^{\boldsymbol{\Sigma}}(x) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}_{[\lambda_j = x]},$$

using the indicator function notation, where $\mathbb{I}_A$ is the indicator function of the event A [14, p. 2761].

For example, if $\boldsymbol{\Sigma}$ is the identity matrix of size $n \times n$, then clearly the corresponding eigenvalues are all equal to 1, and thus the population spectral distribution in this case is a point mass at point 1 [14, p. 2762]. This is exactly the discussed previously special case, in which the limiting behaviour of the empirical spectral distribution is described by theorem 2.3.1.

Now, assume that a $n \times T$ data matrix is given, called $\boldsymbol{X}$. Following the work of Karoui (see [14, p. 2764]) calculate the sample covariance matrix $\boldsymbol{S}$ by $\boldsymbol{S} = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\intercal}$, as given in the equation (2.1.2). By the results established previously, without loss of generality it can be assumed that $\boldsymbol{X}$ has mean 0. Denote by $m_{F^{\boldsymbol{S}}}$ the Stieltjes transform of the spectral distribution, $F^{\boldsymbol{S}}$, of the matrix $\boldsymbol{S}$. Moreover, define a function

$$v_{F^{\boldsymbol{S}}}(z) = \left(1 - \frac{n}{T}\right)\frac{-1}{z} + \frac{T}{n} m_{F^{\boldsymbol{S}}}(z),$$

to be the Stieltjes transform of the spectral distribution of $\frac{1}{T}\boldsymbol{X}^{\intercal}\boldsymbol{X}$.

Then, the well known Marčenko-Pastur equation provides a remarkable way of linking the limiting spectral distribution, F, to the limiting behaviour of the population spectral distribution, H. This result is shown below as Theorem 2.5.1, using the established above notation. It is important to note that Silverstein (see [19, p. 331-338]) has provided a proof of this result requiring only for two moments, however, the further results based on Karoui (see [14]) call for all four moments.

Theorem 2.5.1 (The Marčenko-Pasur equation [**?**, p. 2764]) *Let $\boldsymbol{X}$ be a given data matrix, with $\boldsymbol{X}^{\intercal} = \boldsymbol{Y}\,\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{\Sigma}$ is a $n \times n$ positive definite matrix, and $\boldsymbol{Y} \equiv \{\boldsymbol{Y}_{ti}\}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, n$ is an $T \times n$ matrix containing i.i.d. (real or complex) entries, with $\mathbb{E}(\boldsymbol{Y}_{ti}) = 0$, $\mathbb{E}(|\boldsymbol{Y}_{ti}|^2) = 1$ and $\mathbb{E}(|\boldsymbol{Y}_{ti}|^4) < \infty$. Let $\boldsymbol{\Sigma}$ be the population covariance matrix, and assume that its spectral distribution $H^{\boldsymbol{\Sigma}}$ converges weakly to a limit denoted by H. Then for $T \to \infty$, $n \to \infty$, and $n/T \to y \in (0, \infty)$:*

1. *$v_{F^{\boldsymbol{S}}}(z) \overset{a.s.}{\to} v_\infty(z)$, where v_∞ is a deterministic function.*
2. *$v_\infty(z)$ satisfies the equation*

$$-\frac{1}{v_\infty(z)} = z - y\int \frac{\lambda\,\mathrm{d}H(\lambda)}{1 + \lambda v_\infty(z)} \quad \forall z \in \mathbb{C}^+ \tag{2.5.1}$$

3. *Equation (2.5.1) has exactly one solution which is the Stieltjes transform of a measure.*

Theorem 2.5.1 implies that asymptotically the spectral distribution of the sample covariance matrix $\boldsymbol{S}$ is nonrandom, and its characterisation is given through the relation, given in equation (2.5.1), with the true population spectral distribution [14, p. 2765]. Moreover, for the special case, in which the population eigenvectors are equal to 1, links the result of Theorem 2.5.1 directly to the studied previously Marčenko-Pastur law.

Proof of Theorem 2.5.1 has been given by Silverstein in [19], and strongly bases on techniques and results given by Bai & Silverstein [1]. That said, many results leading to establishing this theorem have been given in this project, in previous sections.

Theorem 2.5.1 has been basis for Karoui's work [14], which gives an algorithm to extract meaningful information about the population spectral distribution, H, in a practical setting; this idea is studied further in section 4. It has to be noted that the aforementioned algorithm has been proved by Karoui [14] to be consistent – a result that follows as a consequence of the theorem stated next.

Theorem 2.5.2 (Karoui [14, p. 2779-2780]) *Under the setup of Theorem 2.5.1, suppose that* $H^{\mathbf{\Sigma}} \to H$*, and* $n/T \to y \in (0, \infty)$*. Also, assume that the spectra of the* $\{\mathbf{\Sigma}_n\}$*'s are uniformly bounded. Let* $J_1, J_2, \ldots$ *be a sequence of integers that tend to infinity. Let* $z_0 \in \mathbb{C}^+$ *and* $r \in \mathbb{R}^+$ *be such that* $B(z_0, r) \subset \mathbb{C}^+$*. Let* $z_1, z_2, \ldots$ *be a sequence of complex variables with a limit point, all contained in* $B(z_0, r)$*. Let* $\hat{H}_n$ *be the solution of*

$$\hat{H}_n = \arg\min_{\tilde{H}} \max_{j \leq J_n} \left| \frac{1}{v_F \boldsymbol{s}(z_j)} + z_j - \frac{n}{T} \int \frac{\lambda \, d\tilde{H}(\lambda)}{1 + \lambda v_F \boldsymbol{s}(z_j)} \right|,$$

where $\tilde{H}$ *is a probability measure. Then* $\hat{H} \to H$*, almost surely.*

The proof of Theorem 2.5.2 has been given by Karoui [14, p. 2780-2785] and will not be reproduced in this project. It bases on some intermediate lemmas, with some results following from the Theorem 2.5.1.

3 Emperical Study of Spectral Analyis of Simulated Data

Theoretical results established in section 2 equip statisticians with very powerful large dimensional asymptotics results in regards to the spectral analysis of sample covariance matrices or Wigner matrices. It has to be noted, however, that, for example, although the semicircle law enables application to uncorrelated random graphs, practical applications used in description of complex systems such as the Internet, metabolic pathways, networks of power stations, or scientific collaborations, which often have inherited correlation, may not follow this law [6]. There has been done much research in the literature, over the recent decade, in practical applications of the theoretical results shown in this project, which extend the aforementioned results to fit the specific data under the study. In this particular project financial data is considered, for which the results from section 2 can be easily extended (see section 4), as well as used for the comparison with the actual observation in order to find meaningful information (see section 5).

In this chapter the focus is placed on verifying the results from section 2 based on the computer simulated data from two distributions: the normal distribution and the Cauchy distribution, where in the first part only the normal distribution is used to illustrate the convergence of the spectral distribution of sample covariance matrix to the semicircular law, and in the second part both distributions are used to verify the Marčenko-Pastur results.

The normal (or Gaussian) distribution with mean parameter μ and standard deviation σ, denoted as $N(\mu, \sigma^2)$, with a probability density function given by

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma > 0,$$

is perhaps one of the most commonly used distributions, with a large range of applications to practical data sets in many fields of science [16]. However, most importantly, the first four moments are finite (in fact, all moments are finite), which is sufficient for the application of the theoretical results established in this project[9].

In a contrast to this, the Cauchy distribution with the location parameter a and the scale parameter b, with a probability density given by

$$\tilde{f}(x|a,b) = \frac{1}{\pi\, b[1 + ((x-a)/b)^2]}, \quad a \in \mathbb{R},\ b > 0,$$

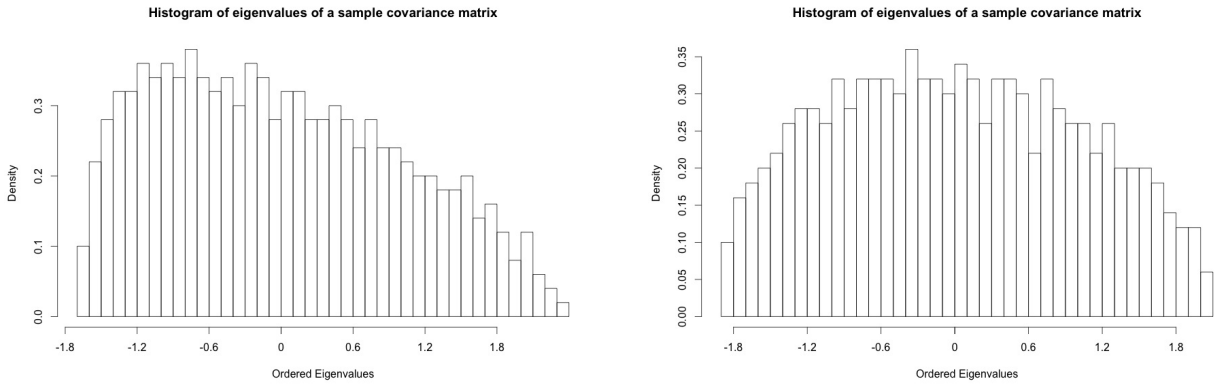
does not have any finite moment. Therefore, this practical choice of the two distributions enables to investigate and illustrate the contrast in achieved results regarding the spectral analysis of the simulated data, since the theoretical results are not applicable to distributions without finite moments of any order, i.e. to the Cauchy distribution.

3.1 Empirical Verification of Wigner's Semicircular Law

In section 2.2 an example of direct simulation of a Wigner matrix using the method of generating a sample of i.i.d. normal observations and creating a scaled symmetric matrix has already been given. A strong convergence to the theoretical results has been already observed for the case where the size of the matrix was 5000×5000, whereas relatively good, yet with some observable noise, result was achieved in the case when the matrix was only of size 500×500.

Here, Wigner's semicircular law is revisited, where under study is the empirical spectral distribution of sample covariance matrix based on a multivariate normal distribution with mean vector of n zeroes

[9] Note that four moments are mentioned, as the result given by Karoui [14] (see section 4) call for it; that said, the empirical verification of this result will not be pursued in this project.



(a) Simulation with $n = 500$, $T = 5000$, bin size 50. (b) Simulation with $n = 500$, $T = 50000$, bin size 50.

Figure 3.1.1: Study of the convergence to Wigner's Semicircular Law, through histograms of the eigenvalues of sample covariance matrix of simulated multivariate normal data.

and the population covariance matrix $\mathbf{\Sigma} = \boldsymbol{I}$, the $n \times n$ identity matrix. The value of n will be fixed to 500 in order to allow for the comparison with the result achieved in section 2.2, while keeping the value of n relatively small in order to allow for better code performance when sample size T is chosen large. During the simulation different values of T where chosen, where the results for $T = 5000$ and $T = 50000$ are presented[10].

Based on the figure 3.1.1, it can be seen that relatively large value of T is required (for $n = 500$, $T = 50000$ the data matrix is about 200 Mb in size), for the empirical spectral distribution to start to converge to semicircular law. In general, it has been observed that for results with $T < 20000$ the histogram is generally skewed, whereas increasing the sample size affects the overall shape of the histogram. Varying n on the other hand controls the amount of noise observed in the plot, with larger n resulting in smoother looking eigenvalue distribution and smaller local variance in ordered eigenvalues frequency.

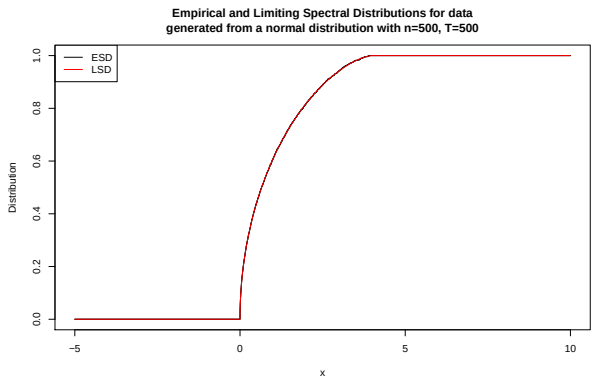
In conclusion, the convergence rate to the semicircular in the above experiment can be considered as slow. Moreover, in general, the theory requires data that not only follows the rigorous restrictions about the population distribution structure (covariance matrix), stated in Theorem 2.2.1, but also, based on the conducted study, needs to be drawn from a large sample size. That said, any observed significant deviations away from the typical $[-2, 2]$ range for the eigenvalues can highlight a statistically significant eigenvalue, and hence help to determine the eigenvector that holds true information about the population distribution – a key factor of the principal component analysis.

3.2 Empirical Verification of Marčenko-Pastur Results

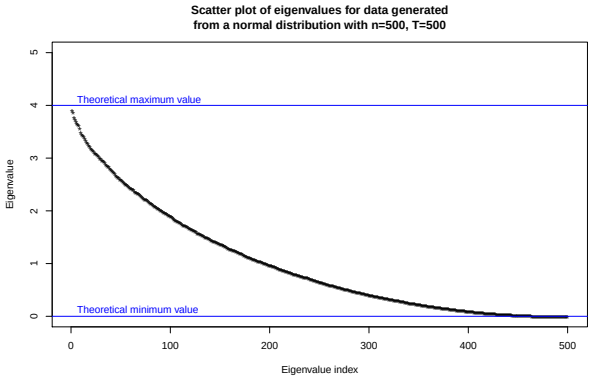
In this section the limiting spectral distribution, distributed according to Marčenko-Pastur law, with density as given in (2.3.1), will be compared to the attained empirical limiting spectral distribution of sample covariance matrix calculated from the simulated data on a computer. The generated data comes from multivariate normal distribution with zero mean and population covariance matrix equal to the identity matrix, as well as the i.i.d. (standard) Cauchy random variables with location parameter $a = 0$, and scale parameter $b = 1$.

Note that, using results of Theorem 2.4.2, it is then possible to establish the theoretical bounds for the resultant largest and smallest eigenvalue. Note that Marčenko-Pastur results only hold in

[10]The code used for the simulation is given in appendix B.2.

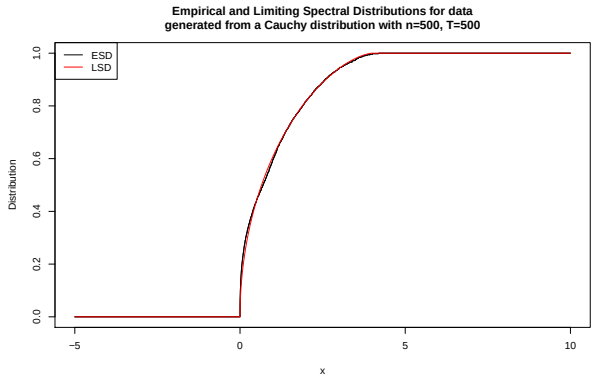


(a) Comparison between ESD and LSD.

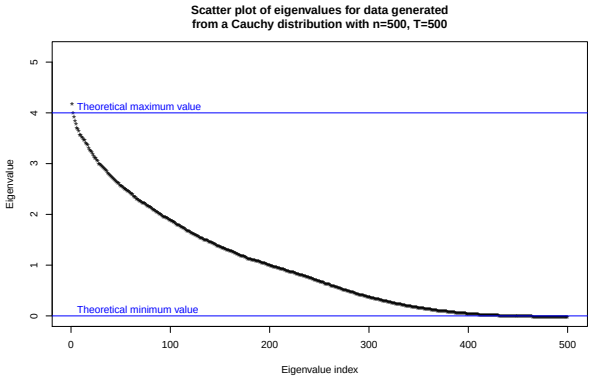


(b) Scatter plot of sample covaraiance matrix eigenvalues with theoretical bounds.

Figure 3.2.1: Empirical verification of Marčenko-Pastur theortical results for simulated data from *normal* distribution with $n = 500$, $T = 500$, $y = 1$.



(a) Comparison between ESD and LSD.

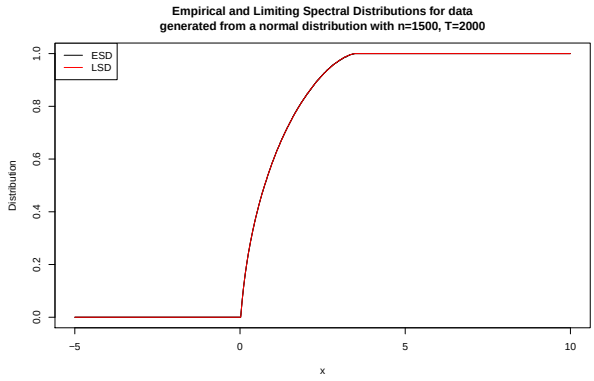


(b) Scatter plot of sample covaraiance matrix eigenvalues with theoretical bounds.

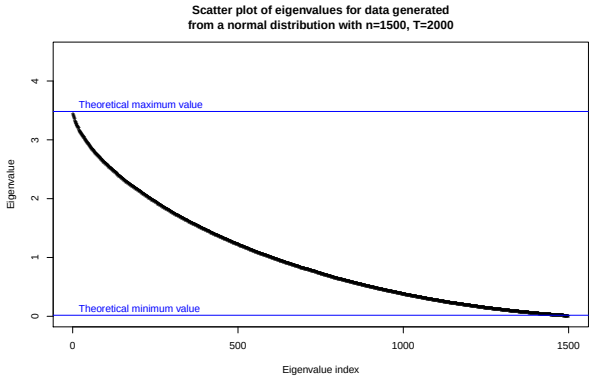
Figure 3.2.2: Empirical verification of Marčenko-Pastur theortical results for simulated data from *Cauchy* distribution with $n = 500$, $T = 500$, $y = 1$.

asymptotics, and hence some possible deviation may result in practice; that said, investigation in this section will look at the sample sizes T and dimension sizes n varying in the range of 500 to 2000, which can be considered as large. Moreover, as already mentioned, the Cauchy distribution, due to lack of moments, does not verify the conditions for the theoretical results to hold, and, in general, it is not expected that the observed empirical values will match the theoretical ones in this case. The following investigation aims to verify firstly what values of T, n can be considered as "large", and secondly to compare the results when applied to the Cauchy distribution.

The calculation of sample covariance matrix, empirical and limiting spectral distributions, as well as limiting bounds on the eigenvalues follow from the definitions and techniques established in section 2. The written R code is given in appendix B.3.

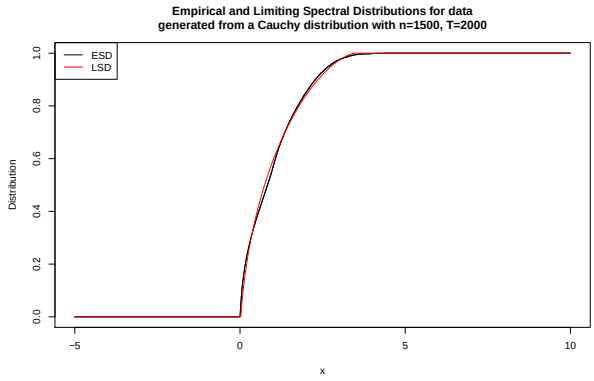


(a) Comparison between ESD and LSD.

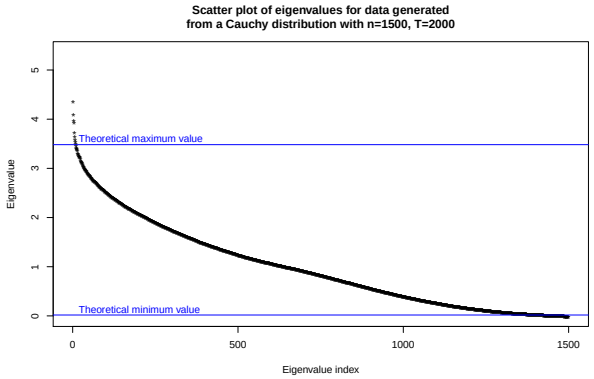


(b) Scatter plot of sample covaraiance matrix eigenvalues with theoretical bounds.

Figure 3.2.3: Empirical verification of Marčenko-Pastur theortical results for simulated data from *normal* distribution with $n = 1500$, $T = 2000$, $y = 0.75$.



(a) Comparison between ESD and LSD.



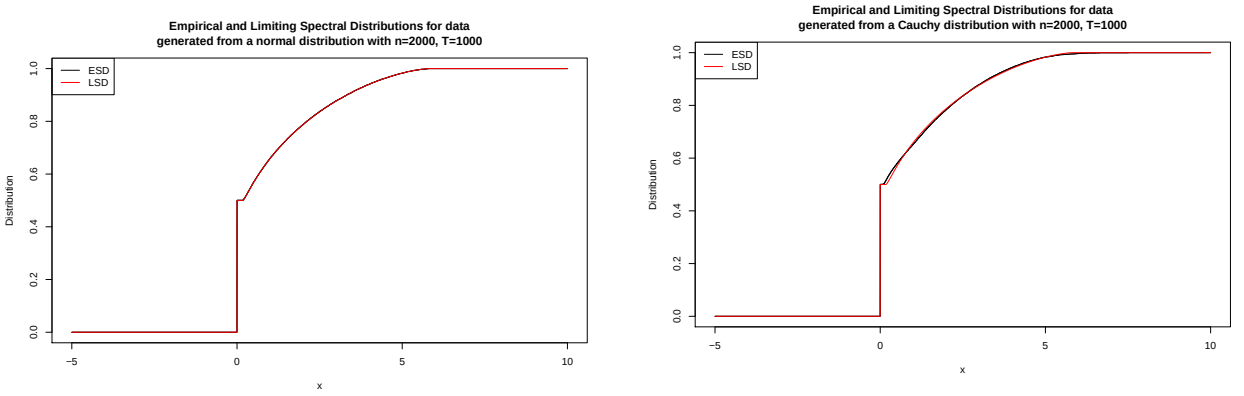
(b) Scatter plot of sample covaraiance matrix eigenvalues with theoretical bounds.

Figure 3.2.4: Empirical verification of Marčenko-Pastur theortical results for simulated data from *Cauchy* distribution with $n = 1500$, $T = 2000$, $y = 0.75$.

Note that, it is also possible to work with the correlation matrix, achieving the same results in the case of the normal distribution, and rescaled results for the Cauchy distribution (noting that, despite rescaling, the conclusion drawn are the same). This is due to the fact that the (standard) Marčenko-Pastur result considers data with unit variance, and hence computing the correlation matrix instead of sample covariance serves as a way of normalising each time series by its standard deviation.

The corresponding resultant plots of the simulation investigation are given in figures 3.2.1 - 3.2.5. Note that other dimensions for T and n have been considered, however, in general the resultant plots were very similar, and hence, for ease of readability, only few graphs that help to illustrate the main points of the conclusion of this investigation are presented.

Firstly, for the normal distribution with data set size $n = 500$ and $T = 500$, and so $y := n/T = 1$, the empirical limiting distribution and the theoretical spectral limiting distribution closely coincide, which suggests that it is possible to attain a relatively quick convergence of the limiting spectral



(a) Comparison between ESD and LSD.

(b) Scatter plot of sample covaraiance matrix eigenvalues with theoretical bounds.

Figure 3.2.5: Empirical verification of Marčenko-Pastur theortical results for simulated data from *normal* and *Cauchy* distributions with $n = 2000$, $T = 1000$, $y = 2$.

distribution, making the theory applicable even to relatively small data set sizes (< 500). This is perhaps not surprising, as Karoui claims that the algorithm, which bases on Marčenko-Pastur theory (see 4), is already attaining good results with data sizes of 30 or more [14, p. 2770]. Moreover, the theoretical maximum value of the sample covariance matrix eigenvalues is very slightly below the theoretical maximum, while the theoretical minimum is approximately attained.

Similar analysis, in the case of the normal distribution, follows from figure 3.2.3, where a larger data set with $n = 1500$ and $T = 2000$, and so $y := n/T = 0.75$ is used. The eigenvalues form a very clear pattern, and their distribution closely follows he Marčenko-Pastur law, with the eigenvalues attaining (in approximation), but not exceeding, the theoretical bounds.

On the other hand, for the Cauchy distribution, some of the eigenvalues (in general, roughly 10%) do not lie within the theoretical bounds, with values both above and below, as can be observed on figures 3.2.2b and 3.2.4b. However, there is no clear connection between the exact proportions below and above the theoretical limits and the dataset size; depending on different simulated values a slightly different pattern is observed. Similarly, in the case of the empirical spectral distribution, observed in figures 3.2.2a and 3.2.4a, although it deviates from the Marčenko-Pastur law, there is no evident pattern linking this behaviour and the dimensions size. That said, for larger sizes, in general larger disparity between the empirical spectral distribution and the Marčenko-Pastur law can be spotted.

Finally, in the case when $y > 1$, Theorem 2.4.2 is no longer applicable, and hence the investigation on the bounds of eigenvalues is omitted. Based on figure 3.2.5 it can be seen that a similar to the stated above conclusion can be drawn – there is some deviation from the Marčenko-Pastur law in the case of the Cauchy distribution, whereas the normal distribution follows it precisely. During this investigation, for different values of n and T within the rage $[500, 2000]$ such that $y > 1$, the produced plots gave similar results, where it was also observed that in general the smaller data size the less accurate the empirical spectral distribution in regards to the Marčenko-Pastur law, which seems to be intuitive, due to the asymptotic nature of the theory.

As a final remark, it should be noted that the number of variables n has a larger effect on the deviation of the empirical spectral distributions away from the theoretical one than the sample size T, where it has been possible, in the case of the normal distribution, to obtain very close fit to he Marčenko-Pastur law for the dataset of size $n = 2000$, $T = 500$, while it was much harder in the case when $n = 500$, $T = 2000$.

4 Retrieving Information on Limiting Behaviour of Population Spectral Distribution

Results in section 2 have established a fact that provided the limit of the population spectral distribution, H, it is possible to obtain the limiting spectral distribution, F. However, in the practical application of data analysis, no information is given about the population spectral distribution. The actual goal of such analysis is to retrieve the information about the population spectral distribution, based on the fact that the empirical spectral distribution is readily computed [14, p. 2763].

Therefore, the key issue is to estimate the population eigenvalues, $\lambda_1, \ldots, \lambda_n$, using the obtained eigenvalues of a sample covariance matrix, $l_1, \ldots, l_n$, through the relationship described by the equation (2.5.1). The difficulty, however, lies within the fact that the discussed relationship between F and H is entangled, and thus requires a careful consideration when seeking to extract a useful information [14, p. 2765].

Karoui has proposed a method that "inverts" the relation between F and H, in order to provide an algorithm that allows to calculate an estimate of the population spectral distribution, $\hat{H}^{\boldsymbol{\Sigma}}$ [14, p. 2763]. The provided strategy bases on three key points [14, p. 2765]:

1. Firstly, measure H has to be estimated from the Marčenko-Pastur equation, given in equation (2.5.1).

2. Provided that an estimator, $\hat{H}$, of the measure H is found, the next step is to estimate λ_i as the i^{th} quantile of the estimated distribution.

3. Finally, it has to be noted that since fixed distribution asymptotics are under consideration, the estimate of H also serves as the estimate of $H^{\boldsymbol{S}}$, so in other words $\hat{H}^{\boldsymbol{S}} = \hat{H}$.

It follows then that the main issue is step 1, in which estimation of H is required, based only on the availability of $F^{\boldsymbol{S}}$. As pointed out by Karoui [14, p. 2766], since the eigenvalues of $\boldsymbol{S}$ can be computed, it is also possible to evaluate $v_{F^{\boldsymbol{S}}}(z)$ for any choice of the variable z. Thus, the proposed approach is to evaluate the set of values of $v_{F^{\boldsymbol{S}}}$ at the grid formed by points $\{z_j\}_{j=1}^{J_T}$, for which (2.5.1) holds (approximately) [14, p. 2766]. Karoui argues that the most suitable estimate of $\hat{H}$ in this setup is then a value for which (2.5.1) is satisfied by $\{v_{F^{\boldsymbol{S}}}(z_j)\}_{j=1}^{J_T}$ in the largest degree, so hence proposing

$$\hat{H} = \underset{\tilde{H}}{\arg\min}\, L\left(\left\{\frac{1}{v_{F^{\boldsymbol{S}}}(z_j)} + z_j - \frac{n}{T}\int \frac{\lambda\, \mathrm{d}\tilde{H}(\lambda)}{1+\lambda v_{F^{\boldsymbol{S}}}(z_j)}\right\}_{j=1}^{J_T}\right),$$

where the optimisation is conducted over probability measures $\tilde{H}$, and L is a loss function of choice (see section 4.1). This indeed gives a method for *inversion* of (2.5.1), in which using $F^{\boldsymbol{S}}$ as an estimate of F allows to obtain an estimate H.

4.1 Algorithm Finding the Estimate of Population Spectral Distribution

In this section, a practical discussion on implementation of the algorithm proposed by Karoui [14], referred to in section 4, is given. Note that the work given in this section is closely following the study done by Karoui [14].

Firstly, note that the measure $\mathrm{d}H$ can be approximated by weighted sum of $K \in \mathbb{N}$ point masses:

$$\mathrm{d}H(x) \approx \sum_{k=1}^{K} w_k \delta_{t_k}(x),$$

where δ_{t_k} is a point mass of 1 at t_k, $\{t_k\}_{k=1}^K$ is a selected grid of points, and $\{w_k\}_{k=1}^K$ are weights such that

$$\sum_{k=1}^{K} w_k = 1 \quad \text{and} \quad w_k \leq 0.$$

Using this approach, the integral in (2.5.1) can be then approximated simply as

$$\int \frac{\lambda \, dH(\lambda)}{1+\lambda v} \approx \sum_{k=1}^{K} w_k \frac{t_k}{1+t_k v},$$

which turns the problem of finding the measure H into an optimisation problem in which set of weights $\{w_k\}_{k=1}^K$ is selected for which

$$-\frac{1}{v_{F^{\boldsymbol{s}}}(z_j)} \approx z_j - \frac{n}{T} \sum_{k=1}^{K} w_k \frac{t_k}{1+t_k v_{F^{\boldsymbol{s}}}(z_j)}, \tag{4.1.1}$$

holds approximately for all $j = 1, \ldots, J_T$, by noting the assumption made that $v_{F^{\boldsymbol{s}}}(z_j) \approx v_\infty(z_j)$.

Rewriting the problem in terms of (4.1.1) enables for large simplification, where the only unknown quantities are the weights $\{w_k\}_{k=1}^K$. As shown by Karoui [14], this presents a relatively simple convex optimisation problem; this is argued as follows.

For $j = 1, \ldots, J_T$, call the approximation errors made, that are in general complex numbers, to be

$$e_j = \frac{1}{v_{F^{\boldsymbol{s}}}(z_j)} + z_j - \frac{n}{T} \sum_{k=1}^{K} w_k \frac{t_k}{1+t_k v_{F^{\boldsymbol{s}}}(z_j)},$$

where e_j occur due to both discretisation of the measure and the use of $v_{F^{\boldsymbol{s}}}(z_j)$ instead of $v_\infty(z_j)$. Then the inversion of the Marčenko-Pastur equation can be turned into optimisation problem based on e_j's provided that a suitable loss function L is chosen, which brings many advantages such as the guarantee of fast algorithms (for example, see MOSEK optimisation package for Matlab, [17]). Karoui has listed three main choices, whereas the consistency of the algorithm is only proven by Karoui for the first one:

1. L_∞ version: search for weights $\{w_k\}_{k=1}^K$ in order to minimise

$$\max_{j=1,\ldots,J_T} \max\{|\Re\, e_j|, |\Im\, e_j|\}.$$

2. L_2 version: search for weights $\{w_k\}_{k=1}^K$ in order to minimise $\sum_{j=1}^{J_T} |e_j|$.

3. L_2-squared version: search for weights $\{w_k\}_{k=1}^K$ in order to minimise $\sum_{j=1}^{J_T} |e_j|^2$.

In this project, details about the resultant linear programming problem will be omitted. For further information in regards to the algorithm see the original work by Karoui [14], including the discussion on how to select the grid points $\{t_k\}_{k=1}^K$. Finally, as already mentioned, Karoui has used Theorem 2.5.2 to prove the consistency of the algorithm using the L_∞ choice for the loss function (see [14, p. 2785-2786] for the proof and further details).

5 Correlation of Financial Stocks

For the particular problem considered in this project, let $\boldsymbol{r}(t)$ denote a column vector containing in its i^{th} column the log-returns of a particular stock $r_i(t)$, for $i = 1, 2, \ldots, n$, at time $t = 1, 2, \ldots, T$. Then,

$$\boldsymbol{S} = \frac{1}{T-1} \sum_{t=1}^{T} \left(\boldsymbol{r}(t) - \bar{\boldsymbol{r}}\right) \left(\boldsymbol{r}(t) - \bar{\boldsymbol{r}}\right)^{\intercal},$$

where $\bar{\boldsymbol{r}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{r}(t)$, is the sample covariance matrix for the particular dataset containing log-returns of a stock market index, exactly as defined previously.

In this project, data from the *Standard & Poor 500* (S&P 500) index will be considered. This index, designed as a leading indicator of U.S. equities, aims to reflect the risk and return characteristics of the large capitalisation market [11]. Created in 1957, it contains 500 stocks that are regarded to be widely held, where over 70% of all U.S. equity is being tracked by it [12].

It has been possible to identify $n = 336$ stocks for which daily adjusted price, from which log-returns are obtained, is available over the horizon of $T = 2517$ days ≈ 10 years. The log-returns at date t are given by the formula $r_i(t) = \log\left(\frac{S^i(t+1)}{S^i(t)}\right)$, where $S^i(t)$ denotes the current price of the stock i at time t. Call the $n \times T$ matrix with this data as $\boldsymbol{X}$. Part of the matrix $\boldsymbol{X}$ is being displayed below, as a figure 5.0.1.

```
                A           AA          AAPL           ABT           ACE           ACN
[1,] -0.023770219 -0.04115807 -0.001140034 -0.063723478 -0.003188342 -0.050430854
[2,] -0.010362787  0.05185336  0.047159286 -0.011236073  0.019874186  0.017746111
[3,]  0.001156738 -0.01069529 -0.003997825  0.011856613  0.009347943 -0.005880448
[4,] -0.008125407 -0.01626052 -0.006943203  0.006184312 -0.008900815 -0.011202753
[5,]  0.025317808  0.02961185  0.002014837  0.033346527  0.011555684  0.031313713
```

Figure 5.0.1: Partial print-out of the data matrix $\boldsymbol{X}$.

In addition to the sample covariance matrix, $\boldsymbol{S}$, of $\boldsymbol{X}$, it is possible to calculate the correlation matrix $\boldsymbol{C} = (C_{ij})$ for $i, j = 1, \ldots, n$ with entries in $[-1, 1]$, through the formula

$$C_{ij} = \frac{Cov(S_i, S_j)}{\sigma_{S_i} \sigma_{S_j}},$$

where $Cov(S_i, S_j)$ is the covariance between stock S_i and S_j available in the ij^{th} entry of $\boldsymbol{S}$, while σ_{S_i} and σ_{S_i} is the standard deviation of stock S_i and S_j, respectively, calculated as the square root of the variance of the stock. This results in a $n \times n$ symmetric square matrix with entries in the range $[-1, 1]$ and 1's on the diagonal. Part of the matrix $\boldsymbol{C}$ is being displayed below, as a figure 5.0.2.

Similarly to the work done in section 3, it is possible to compute the eigenvalues and corresponding eigenvectors (the *principal components*), and compare them to the theoretical results in order to identify the *significant factors*. However, firstly a word on the stylised statistical properties of asset returns has to be mentioned, in order to justify the use of theoretical assumption of finite relevant moments of the log-returns.

```
             A         AA       AAPL        ABT        ACE        ACN
A    1.0000000 0.3494677 0.4853224 0.4965646 0.4132226 0.3320724
AA   0.3494677 1.0000000 0.4336627 0.4099646 0.4402923 0.3651333
AAPL 0.4853224 0.4336627 1.0000000 0.5416422 0.5456800 0.4068648
ABT  0.4965646 0.4099646 0.5416422 1.0000000 0.5028373 0.3705115
ACE  0.4132226 0.4402923 0.5456800 0.5028373 1.0000000 0.4219563
ACN  0.3320724 0.3651333 0.4068648 0.3705115 0.4219563 1.0000000
```

Figure 5.0.2: Partial print-out of the correlation matrix $\boldsymbol{C}$.

5.1 Stylised Statistical Properties of Asset Returns

The established theoretical results in section 2 base on distributions with at least the first two moments being finite. Moreover, the practical application of these results (see section 4) calls of the finite fourth moment. Since the distribution of financial data cannot be specified exactly, one needs to be careful when applying, say, Theorem 2.3.1 and Theorem 2.4.2, or solving the linear programming problem in order to attain an estimate of the population spectral distribution (see section 4). Based on the empirical evidence, as discussed in section 3, it can be seen that in an approximation the main results are also very vaguely applicable for distributions that do not have any finite moments, such as the Cauchy distribution. Of course, the asymptotic convergence in that case is not guaranteed, and the obtained results should be treated with extreme caution.

In the literature there have been many empirical studies done, from which it has been observed that financial time series exhibit common statistical properties, known as *stylised empirical facts* [5, p. 224]. As given in many sources (see for example [5], [9]), in general, the assumption of the finite fourth moment of the log-returns of S&P 500 index is justified by empirical study. In this project, however, this assumption will not be further put under the study and will be considered to be a fact.

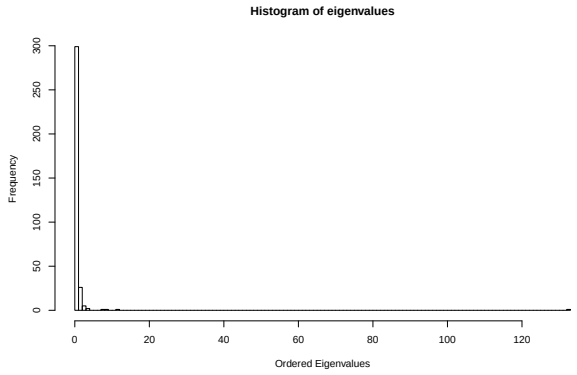
With that (reasonable) assumption made, it is now possible to move to the actual principal component analysis of the S&P 500 data set.

5.2 Principal Component Analysis of the S&P 500 Index

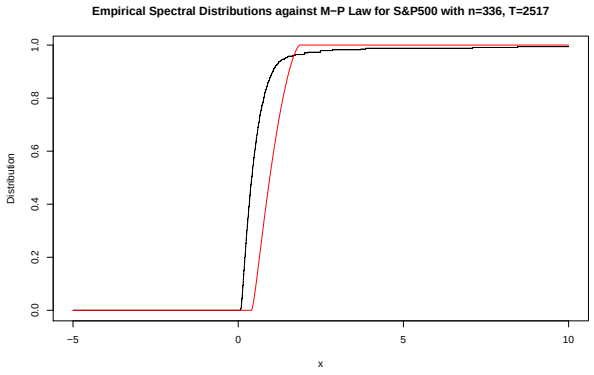
Based on the calculations and application of discussed in this project random matrix theory, it is possible to draw conclusions in regards to the principal component analysis of the S&P 500 index. The R code written for the investigation in this section is available in appendix B.4.

Firstly turning to the histogram of the eigenvalues, shown in figure 5.2.1a, an already clear patter is visible. There is one large eigenvalue of size 132, several eigenvalues that are larger than 1, while the significant proportion (roughly 90%) has values below 1, but all are positive. Moreover, since it is expected that there is inherit correlation between the stocks, the distributional condition of Marčenko-Pastur law does not hold, and so there is divergence of the observed empirical spectral distribution from the Marčenko-Pastur law, as observed in figure 5.2.1b.

A further analysis, using the theoretical bounds established through Theorem 2.4.2, shows that in fact it is possible to distinguish 12 significant eigenvalues that have values beyond the theoretical maximum, and hence can be regarded as non-random and containing information about the correlation structure. It has to be noted that, although there are also eigenvalues beyond the theoretical minimum, their deviation is not significant in size and hence will be ignored. See figure 5.2.2a for the graphical representation of the attained eigenvalues. Having determined the 12 principal components it is now possible to reach conclusions regarding the index and its underlying stocks.

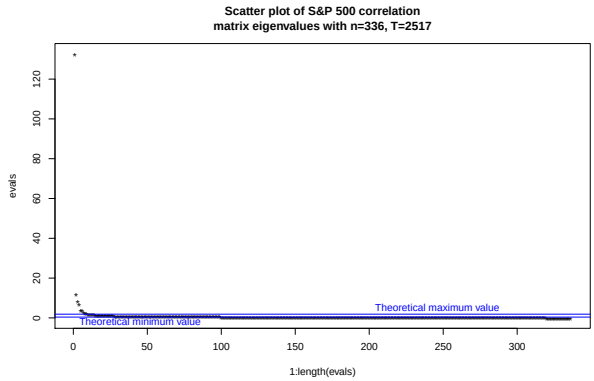


(a) Histogram of the eigenvalues of the correlation matrix $\boldsymbol{C}$.

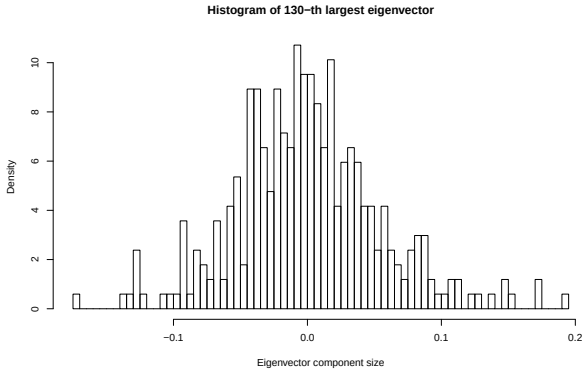


(b) Empirical spectral distribution of S&P 500 index against Marčenko-Pastur result.

Figure 5.2.1: Graphical analysis of the S&P spectral distribution.



(a) Scatter plot determining significant eigenvalues.

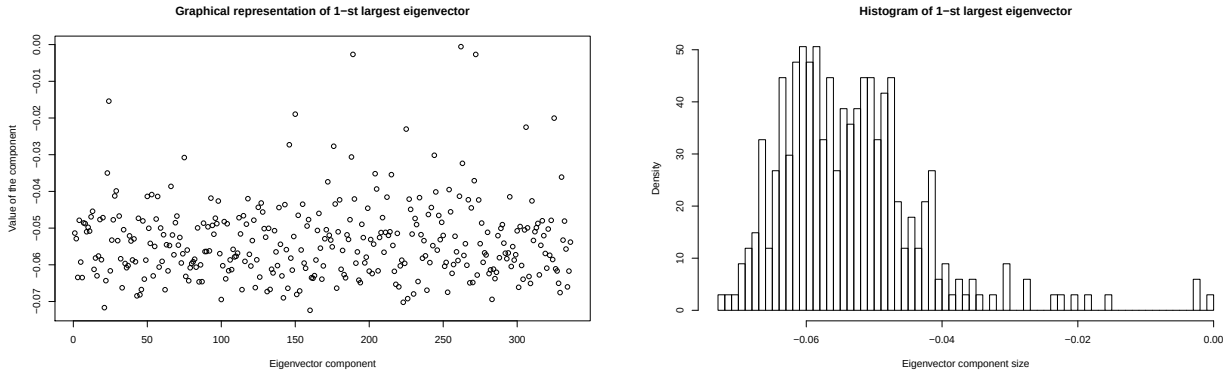


(b) Histogram of an example non-significant eigenvector.

Figure 5.2.2: Graphical analysis of the S&P 500 index.

Refer by the k^{th} largest eigenvector to the eigenvector corresponding to the k^{th} largest eigenvalue. If a significant component (eigenvectors corresponding to a large, non-random eigenvalue) has only a subset of elements different from zero, it can be seen to represent a specific market scenario [7, p. 536]. In the first largest eigenvector, all components carry the same sign. This can be interpreted that there exists a general trend in which stock prices in the index rise and fall together. Moreover, it can be concluded that the largest eigenvalue identifies the entire market [7, p. 536]. Moreover, the magnitude of each component, in general, is similar with most values ranging between $[-0.04, -0.07]$, to which there are some outliers with a smaller magnitude, such as the Kroger Co. (KR) stock, with the component value of only -0.00056. This information can be represented graphically, as done in figure 5.2.3.

Similar analysis of the significant components has been conducted extensively in literature (see for example [15]). It is possible to identify specific, for example, eigenvectors relating to the sectors of the index, or having a specific relation to the geographical location of the stocks. Hence, a detailed study of the first 12 largest eigenvalues, and their elements relating to each particular stock, can allow for a detailed overview of the index behaviour and provide important implications relating to different factors of the stocks, such as the geography, particular industry, market capitalisation, etc. Moreover,



(a) Scatter plot determining significant eigenvalues.

(b) Histogram of eigenvector.

Figure 5.2.3: Graphical analysis of the eigenvector corresponding to the largest eigenvalue.

it is possible to take different time periods (altering possibly the size of T), and compare the effects of economical events (such as economical crisis) on the particular groups of industries related by the significant eigenvectors (for example, see [18]). However, the main focus of this project is placed on the mathematical aspect and the random matrix theory, and therefore this study will not be further conducted here.

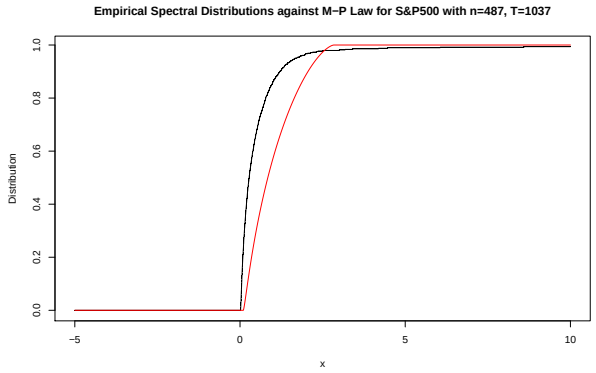
Finally, it should be noted that the eigenvectors for which eigenvalues are regarded as random (i.e. eigenvalues that lie within the theoretical Marčenko-Pastur bounds) contain, in general, components that are normally distributed with mean 0. This has been verified through empirical study in this section, as can be seen on an example of the 130^{th} largest eigenvector, whose elements are represented as a histogram in figure 5.2.2b. Note that other non-significant eigenvectors presented very similar pattern.

5.3 Repeating the Analysis of the S&P 500 Index on Different Time Interval

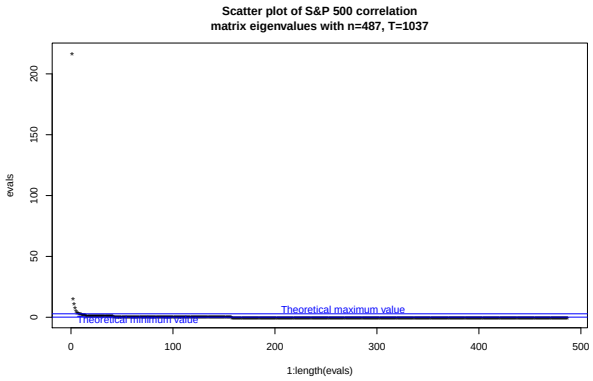
It can be argued that the considered time scope for the S&P 500 data of 10 years seemed unrealistically long. Moreover, the large value of T meant that not all stocks could have been used, resulting the the value of n significantly below 500. In this section the investigation is repeated in analogical way (with obvious changes done to the computer code), with the only difference of using $n = 487$ stocks over the horizon of $T = 1037$ days ≈ 5 years. Note that shorter time horizon enabled to load data for larger number of stocks.

Despite the changes, similar results are observed. These are summarised in graphical form in figure 5.3.1. That said, there have been only 10 principal components found, which lie beyond the theoretical maximum value given by Theorem 2.4.2. The discussion on specific largest eigenvectors will be omitted in this case, as very similar conclusion to the one in section 5.2 follows.

As previously, there have been some eigenvalues also below the theoretical minimum observed, but their deviation is not statistically significant and hence they will not be considered as non-random informative components.



(a) Empirical spectral distribution of S&P 500 index against Marčenko-Pastur result.



(b) Scatter plot determining significant factors.

Figure 5.3.1: Graphical summary of results found for the S&P 500 index over 5 years time horizon.

6 Conclusion

Results from random matrix theory give a powerful tool in dealing with the *large* T*, large* n asymptotic. In this project, main theory regarding the spectral distribution of large-dimensional random matrices, with focus on the sample covariance matrix, has been given. As the consistency of the large dimensional limit theorems has been proven, it is not surprising that their empirical verification through simulations has given desired results. Based on the study, it has been shown that the asymptotic properties of the Marčenko-Pastur result have been attained even for relatively small dimension sizes. That said, the observed through simulation convergence to the Wigner's semicircular law has been relatively poor, suggesting that very large data sets indeed need to be considered in order to attain reasonable asymptotic results.

Under the assumption of finite fourth moment of the the S&P 500 index, a practical investigation on the asset returns has been made, which concerned the estimation of the covariance and factor identification for financial data. For the data set looking at 10 years time horizon, it has been possible to identify 12 principal components, where each corresponding eigenvector has presented some structural pattern. In the shorter, more realistic time period of 5 years, only 10 principal components (out of 487 components) have been selected. In this project a short example of the first largest principal component has been given, omitting the exact pattern identification and its relation to the market scenario. A further extension to this project could include precise study of each 12 principal components, drawing conclusions relating the behaviour of the stocks in regards to factors such as sectors, geographical implications, or effects of different significant economical events.

Finally, in section 4, an algorithm has been proposed allowing to retrieve information for a large dimensional dataset regarding the population covariance matrix, dealing with the problem of inherent noise and variability of the time-series. Application of this algorithm to, for example, financial data set, could bring insight on the actual correlation between stocks or markets. It hence is possible to further extend the investigation in this project by implementing the algorithm to a real dataset (such as the S&P 500 index) in order to retrieve valuable information.

References

[1] Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer; Second Edition, 2010.

[2] Zhidong Bai and Y. Q. Yin. Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix. *The Annals of Probability*, 21(3):1275–1294, 1993.

[3] Jinho Baik and Jack W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[4] W. K. Chen. *Applied Graph Theory*. North-Holland series in applied mathematics and mechanics. Elsevier Science, 2012.

[5] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[6] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek. Spectra of "real-world" graphs: Beyond the semicircle law. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 64(2):026704, 2001.

[7] S.M. Focardi and F.J. Fabozzi. *The Mathematics of Financial Modeling and Investment Management*. Frank J. Fabozzi Series. Wiley, 2004.

[8] Jeffrey S. Geronimo and Theodore P. Hill. Necessary and sufficient condition that the limit of stieltjes transforms is a stieltjes transform. *Journal of Approximation Theory*, 121:54–60, 2003.

[9] Floyd B. Hanson and J. J. Westman. Jump-Diffusion Stock Return Models in Finance: Stochastic Process Density with Uniform-Jump Amplitude, howpublished = `http://www3.nd.edu/~mtns/papers/19046_4.pdf`, note = Online, accessed: 24 March 2013.

[10] Peter J. Huber. Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics*, 1(5):799–821, September 1973.

[11] Investopedia. Standard & Poor 500 Index. `http://www.investopedia.com/terms/s/sp500.asp`. Online, accessed: 10 January 2013.

[12] InvestorWords. S&P 500. `http://www.investorwords.com/4378/SP_500.html`. Online, accessed: 10 January 2013.

[13] Alan J. Izenman. Introduction to Random-Matrix Theory. `http://www.stat.osu.edu/~dmsl/Introduction_to_Random_Matrix_Theory.pdf`. Online, accessed: 14 January 2013.

[14] Noureddine El Karoui. Spectrum Estimation for Large Dimensional Covariance Matrices Using Random Matrix Theory. *The Annals of Statistics*, 36(6):2757–2790, 2008.

[15] Soo Yong Kim, Hee Yunm Choi, Tarik bin Mohd, Asraf Hanafi bin Mohammad Haizad, et al. Correlation structure analysis: China and the US stock markets. *APEC Youth Scientist Journal*, 3(3):25–36, 2011.

[16] K. Krishnamoorthy. *Handbook of Statistical Distributions with Applications*. Hoboken: Taylor & Francis, 2006.

[17] Mosek ApS. MOSEK Optimization Toolbox. Available at http://www.http://mosek.com/, 2006.

[18] Ashadun Nobi, Seong Eun Maeng, Gyeong Gyun Ha, and Jae Woo Lee. Random matrix theory and cross-correlations in global financial indices and local stock market indices. http://arxiv.org/abs/1302.6305. Online, accessed: 20 March 2013.

[19] Jack W. Silverstein. Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995.

[20] Richard M. Timoney. The University of Dublin, Mathematics module MA2224. University Lecture Notes, http://www.maths.tcd.ie/~richardt/MA2224/MA2224-ch4.pdf, 2014.

[21] Y. L. Tong. *Inequalities in Statistics and Probability: Proceedings of the Symposium on Inequalities in Statistics and Probability, October 27-30, 1982, Lincoln, Nebraska*. IMS Lecture Notes. Institute of Mathematical Statistics, 1984.

[22] Laura Veith. Proof of Wigner's Semicircular Law by the Stieltjes Transform Method. Master's thesis, The University of Washington, June 2013. http://www.maths.tcd.ie/~richardt/MA2224/MA2224-ch4.pdf. Online, accessed: 22 February 2013.

[23] Eric W. Weisstein. Hermitian Matrix. http://mathworld.wolfram.com/HermitianMatrix.html. From MathWorld–A Wolfram Web Resource. Online, accessed: 18 January 2013.

[24] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis*, 20(1):50–68, 1986.

Appendices

A Supplementary Theorems

A.1 Perturbation Inequality Theorem

Theorem A.1.1 (Bai & Silverstein, Perturbation Inequality Theorem [1, p. 502]) *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $p \times n$ matrices and the empirical spectral distributions of $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{A}^\intercal$ and $\bar{\boldsymbol{S}} = \boldsymbol{B}\boldsymbol{B}^\intercal$ be denoted by $F^{\boldsymbol{S}}$ and $F^{\bar{\boldsymbol{S}}}$, respectively. Then,*

$$L^4\left(F^{\boldsymbol{S}}, F^{\bar{\boldsymbol{S}}}\right) \leq \frac{2}{p^2}\left(\operatorname{trace}(\boldsymbol{A}\boldsymbol{A}^\intercal + \boldsymbol{B}\boldsymbol{B}^\intercal)\right)\left(\operatorname{trace}[(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})^\intercal]\right).$$

Proof of Theorem A.1.1 is given in Bai & Silverstein, pages 502-503 (see [1, p. 502-503]).

A.2 Rank Inequality Theorem

Theorem A.2.1 (Bai & Silverstein, Rank Inequality Theorem [1, p. 503]) *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $p \times n$ complex matrices. Then,*

$$\left|\left|F^{\boldsymbol{A}\boldsymbol{A}^\intercal} - F^{\boldsymbol{B}\boldsymbol{B}^\intercal}\right|\right| \leq \frac{1}{p}\operatorname{rank}(\boldsymbol{A}-\boldsymbol{B}).$$

More generally, if $\boldsymbol{F}$ and $\boldsymbol{D}$ are Hermitian matrices of orders $p \times p$ and $n \times n$, respectively, then it follows that

$$\left|\left|F^{\boldsymbol{F}+\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^\intercal} - F^{\boldsymbol{F}+\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^\intercal}\right|\right| \leq \frac{1}{p}\operatorname{rank}(\boldsymbol{A}-\boldsymbol{B}).$$

Proof of Theorem A.2.1 is given in Bai & Silverstein, pages 504-505 (see [1, p. 504-505]).

A.3 Difference of Traces of a Matrix and Its Major Submatrices Theorem

In order to present theorem regarding difference of traces of a matrix and its major submatrices, firstly consider, for completeness, the following definition, given below.

Definition A.3.1 (Major Submatrix [4, p. 40]) *For an arbitrary matrix $\boldsymbol{A}$ of order $p \times q$ and of rank p, a major submatrix of $\boldsymbol{A}$ is a nonsingular submatrix of order p.*

Moreover, consider a second definition, given below.

Definition A.3.2 (k^{th} Major Submatrix [1, p. 470]) *For a $n \times n$ matrix $\boldsymbol{A}$, matrix $\boldsymbol{A}_k$, called a major submatrix of order $(n-1)$, is the matrix resulting from removing the k^{th} row and column from $\boldsymbol{A}$.*

Hence, now the major theorem can be stated.

Theorem A.3.1 (Bai & Silverstein, Difference of Traces of a Matrix and Its Major Submatrices Theorem [1, p. 472]) *If the $n \times n$ matrix $\boldsymbol{A}$ and $\boldsymbol{A}_k$, the k^{th} major submatrix of $\boldsymbol{A}$ of order $(n-1)$, are both nonsingular and symmetric, then*

$$\operatorname{trace}(\boldsymbol{A}^{-1}) - \operatorname{trace}(\boldsymbol{A}_k^{-1}) = \frac{1 + \boldsymbol{\alpha}_k' \boldsymbol{A}_k^{-2} \boldsymbol{\alpha}_k}{\alpha_{kk} - \boldsymbol{\alpha}_k' \boldsymbol{A}_k^{-1} \boldsymbol{\alpha}_k},$$

where $\boldsymbol{\alpha}_k$ is the vector obtained from the k^{th} row of $\boldsymbol{A}$ with the k^{th} element removed. Clearly, if $\boldsymbol{A}$ is Hermitian, then $\boldsymbol{\alpha}_k'$ can be replaced by $\boldsymbol{\alpha}_k^\intercal$.

Proof of Theorem A.3.1 is given in Bai & Silverstein, pages 471-472 (see [1, p. 471-472]).

A.4 Trace of an Inverse Matrix Theorem

Following the definitions and notation established in appendix A.3, the following result holds, stated below.

Theorem A.4.1 (Trace of an Inverse Matrix [1, p. 470]) *If both $\boldsymbol{A}$ and $\boldsymbol{A}_k$, $k = 1, 2, \ldots, n$, are nonsingular, and if $\boldsymbol{A}^{-1}$ is written as $\boldsymbol{A}^{-1} = \left[a^{kl}\right]$, then*

$$a^{kk} = \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \boldsymbol{A}_k^{-1} \boldsymbol{\beta}_k},$$

and hence

$$\operatorname{trace}(\boldsymbol{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \boldsymbol{A}_k^{-1} \boldsymbol{\beta}_k},$$

where a_{kk} is the k^{th} diagonal entry of $\boldsymbol{A}$, $\boldsymbol{A}_k$ and $\boldsymbol{\alpha}_k$ as defined previously, and $\boldsymbol{\beta}_k$ is the vector from the k^{th} column of $\boldsymbol{A}$ with the k^{th} element removed.

Proof of Theorem A.4.1 is given in Bai & Silverstein, page 470 (see [1, p. 470]).

A.5 Extended Burkholder Inequality

Theorem A.5.1 (Extended Burkholder Inequality [1, p. 32]) *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p > 1$,*

$$\mathbb{E}\left|\sum X_k\right|^p \leq K_p\, \mathbb{E}\left(\sum |X_k|^2\right)^{p/2},$$

where K_p is a known constant [21, p. 78].

Proof of Theorem A.5.1 is given in Bai & Silverstein, pages 33-34 (see [1, p. 33-34]).

A.6 Moment Convergence Theorem

Let $\{F_n\}$ be a sequence of distribution functions, such that moments of all orders are finite. Denote the k^{th} moment of a distribution F_n by

$$\beta_{n,k} = \beta_k(F_n) := \int x^k \, \mathrm{d}F_n(x),$$

then following *Moment Convergence Theorem* holds, given below [1, p. 507].

Lemma A.6.1 (Moment Convergence Theorem, Unique Limit [1, p. 507]) *A sequence of distribution functions $\{F_n\}$ converges weakly to a limit if the following conditions are satisfied:*

1. *Each F_n has finite moments of all orders.*
2. *For each fixed integer $k \geq 0$, $\beta_{n,k}$ converges to a finite limit β_k as $n \to \infty$.*
3. *If two right-continuous nondecreasing function F and G have the same moment sequence $\{\beta_k\}$, then $F = G + const.$*

Proof of Lemma A.6.1 requires some intermediate lemmas, and is given in Bai & Silverstein, pages 507-514 (see [1, p. 507-514]).

B Written computer code

B.1 Wigner's Semicircular Law with "Direct" Wigner Matrix Simulation

```
library("MASS")

# Set size of the matrix
n <- 500
# Generate matrix with iid N(0,1) entries
H <- matrix(rnorm(n^2,0,1),n,n)
# Produce standard Wigner matrix
H <- (H+t(H))/2/sqrt(n)*sqrt(2)
# Calculate eigenvalues
H.eigen <- eigen(H)
H.evals <- H.eigen$values
# Produce plot
hist(H.evals,breaks=100,prob=T,
     xlab="Ordered Eigenvalues",
     main="Convergence of eigenvalues of Wigner matrix to semicircular law"
     )
```

B.2 Wigner's Semicircular Law and Sample Covariance Matrix

```
library("MASS")

# Generate data
n <- 500
t <- 5000
mu <- seq(0,0,length.out=n)
S <- diag(n)
X <- mvrnorm(t,mu,S)
X <- t(X)
# Calculate covariance matrix
S.hat <- (X%*%t(X))/t
# Subtract I, scale
H <- (S.hat-diag(n))*sqrt(t/n)
# Calculate eigenvalues
H.eigen <- eigen(H)
H.evals <- H.eigen$values
# Produce plot
hist(H.evals,breaks=50,prob=T,
     xlab="Ordered Eigenvalues",
     main="Histogram of eigenvalues of a sample covariance matrix",
     xaxt='n'
    )
axis(side=1, at=c(seq(-1.8,1.8,0.6),0), labels=c(seq(-1.8,1.8,0.6),0))
```

B.3 Empirical Verification of Marčenko-Pastur Results

```
#install.packages("RMTstat")

library("MASS")
library("RMTstat")

## Function simulating data from specified distribution based on given parameters
# Required function arguments:
# (i)   distr: sampling distribution; allowed values - "norm", "cauchy"
# (ii)  n: number of degrees of freedom (sample size)
# (iii) T: numer of dimensions (variables)
# (iv)  para: parameters of sampling distributions in a list
MPsim <- function(distr,n,t,para){

  ## Generate data
  if(distr=="norm"){
    case <- "normal"
    mu <- para[[1]]
    S <- para[[2]]
    X <- mvrnorm(t,mu,S)
  }
  else if(distr=="cauchy"){
    case <- "Cauchy"
    mu <- para[[1]]
    scale <- para[[2]]
    X <- matrix(rcauchy(n*t,mu,scale),t,n)
  }
  else (return(0))

  ## Find means
  mu.hat <- apply(X,2,function(x){mean(x,na.rm=TRUE)})

  ## Find covariance
  X.no.mu <- apply(X,1,function(x){x-mu.hat})
  S.hat <- X.no.mu%*%t(X.no.mu)/(t-1)

  ## Find correlation matrix
  rho.hat <- matrix(NA,n,n)
  for(i in 1:n){
    j <- 0
    while(j<i){
      j <- j+1
      rho.hat[i,j] <- S.hat[i,j]/sqrt(S.hat[i,i]*S.hat[j,j])
      rho.hat[j,i] <- rho.hat[i,j]
    }
  }
```

```
  ## Find eigenvalues
  evals <- eigen(rho.hat)$values

  ## ESD vs LSD
  # ESD
  ESD <- function(x){
    return(sum(evals <= x)/n)
  }
  resolution <- 1e4
  points <- seq(-5,10,length.out=resolution)
  ESDpoints <- sapply(points,ESD)

  #LSD
  MPlaw <- pmp(points,t,n)
  MPlaw[MPlaw>1] <- 1

  # Plots
  title <- paste("Empirical and Limiting Spectral Distributions for data\n",
                 "generated from a ",case," distribution with n=",n,", T=",
                 t,sep="")
  plot(points,ESDpoints,main=title,xlab="x",ylab="Distribution",'l')
  lines(points,ESDpoints)
  lines(points,MPlaw,col="red")
  legend('topleft',c('ESD','LSD'),lty=c(1,1),col=c("black","red"))

  ## Add max and min evals
  if(n/t<=1){
    title <- paste("Scatter plot of eigenvalues for data generated\n","from a ",
                   case," distribution with n=",n,", T=",t,sep="")
    plot(1:length(evals),evals,ylim=c(0,max(evals,(1+sqrt(n/t))^2)+1),
         main=title,xlab="Eigenvalue index",ylab="Eigenvalue",pch="*")
    abline(h=(1-sqrt(n/t))^2,col="blue")
    text(0,(1+sqrt(n/t))^2+0.1,"Theoretical maximum value",pos=4,col="blue")
    abline(h=(1+sqrt(n/t))^2,col="blue")
    text(0,(1-sqrt(n/t))^2+0.1,"Theoretical minimum value",pos=4,col="blue")
  }
  else {
    print(n/t)
  }
}

## Generate simulations
n <- seq(500,2000,500)
t <- seq(500,2000,500)
nt <- expand.grid(n,t)

# Save plots
save.path <- '~/Documents/Imperial College London/M3R/R/imgs/sim.pdf'
```

```
pdf(file=save.path,width=10,height=6.5)

for(i in 1:dim(nt)[1]){
  sizes <- nt[i,]
  n.c <- as.numeric(sizes[1])
  t.c <- as.numeric(sizes[2])
  print(n.c)
  print(t.c)
  MPsim("norm",n.c,t.c,list(numeric(n.c),diag(n.c)))
  MPsim("cauchy",n.c,t.c,c(0,1))
}

# Clean-up
dev.off()
```

B.4 Study of Financial Stocks

```
require(quantmod)
library("RMTstat")

# Save plots
save.path <- '~/Documents/Imperial College London/M3R/R/imgs/sp500.pdf'
pdf(file=save.path,width=10,height=6.5)

#### Load data ####
## Load stocks
# Get S&P500 symbol list
#snp500.tickers <- read.table("~/R/data/sp500-symbol-list.txt",quote="\"")
#tickers <- as.matrix(snp500.tickers)
# Load data into new environment
#data <- new.env()
#getSymbols(tickers,src='yahoo',from='2003-01-01',to='2013-01-01',
#           env=data,auto.assign=T)

## Load the dataset instead of fetching information each time
load("~/Documents/Imperial College London/M3R/R/data/sp500downloaded.RData")

## Load stocks' adjusted price to a matrix
stocks.adjusted <- as.matrix(do.call(cbind,eapply(data,Ad)))
colnames(stocks.adjusted) <- tickers[1:dim(stocks.adjusted)[2]]

#### Principal Component Analysis ####
## Calculate log-returns, r
r <- apply(stocks.adjusted,2,ROC)
# Remove NA, set it to 0
r[!is.finite(r)] <- 0
```

```
## Calculate means vector for r, mu.hat
mu.hat <- apply(r,2,function(x){mean(x,na.rm=TRUE)})

## Calculate covariance matrix for r, C.hat
t <- dim(r)[1]
n <- dim(r)[2]
y <- n/t

r.no.mu <- apply(r,1,function(x){x-mu.hat})
S.hat <- r.no.mu%*%t(r.no.mu)/(t-1)

## Find correlation matrix
rho.hat <- matrix(NA,n,n)
for(i in 1:n){
  j <- 0
  while(j<i){
    j <- j+1
    rho.hat[i,j] <- S.hat[i,j]/sqrt(S.hat[i,i]*S.hat[j,j])
    rho.hat[j,i] <- rho.hat[i,j]
  }
}

## Find evals and evecs of rho.hat
rho.hat.eigen <- eigen(rho.hat)
evals <- rho.hat.eigen$values
evecs <- rho.hat.eigen$vectors

# ESD
ESD <- function(x){
  return(sum(evals <= x)/n)
}
resolution <- 1e4
points <- seq(-5,10,length.out=resolution)
ESDpoints <- sapply(points,ESD)

# LSD (M-P)
MPlaw <- pmp(points,t,n)
MPlaw[MPlaw>1] <- 1

## Plots
# Eigenvalues histogram
hist(rho.hat.eigen$values,breaks=100,prob=F,
     xlab="Ordered Eigenvalues",main="Histogram of eigenvalues")

# ESD vs M-P
title <- paste("Empirical Spectral Distributions against M-P Law",
               " for S&P500 with n=",n,", T=",t,sep="")
plot(points,ESDpoints,main=title,xlab="x",ylab="Distribution",pch=".")
```

```
lines(points,ESDpoints)
lines(points,MPlaw,col="red")


title <- paste("Scatter plot of S&P 500 correlation\n",
               "matrix eigenvalues with n=",n,", T=",t,sep="")
plot(1:length(evals),evals,main=title,pch="*")
abline(h=(1-sqrt(n/t))^2,col="blue")
text(200,(1+sqrt(n/t))^2+3,"Theoretical maximum value",pos=4,col="blue")
abline(h=(1+sqrt(n/t))^2,col="blue")
text(0,(1-sqrt(n/t))^2-3,"Theoretical minimum value",pos=4,col="blue")



## Study significant factors
k <- sum(evals>(1+sqrt(n/t))^2)
for(i in 1:(k)){
  if(i>3){q <- "th"}
  else if(i==1){q <- "st"}
  else if(i==2){q <- "nd"}
  else {q <- "rd"}
  title <- paste("Graphical representation of ",i,"-",q," largest eigenvector",
                 sep="")
  plot(evecs[,i],xlab="Eigenvector component",ylab="Value of the component",
       main=title)
}
title <- paste("Histogram of 1-st largest eigenvector",sep="")
hist(evecs[,1],prob=T,breaks=100,xlab="Eigenvector component size",main=title)

## Not significant factors
title <- paste("Histogram of 130-th largest eigenvector",sep="")
hist(evecs[,130],prob=T,breaks=100,xlab="Eigenvector component size",main=title)

# Clean-up
dev.off()
```