Ledoit, Olivier; Wolf, Michael

Working Paper

Numerical implementation of the QuEST function

Working Paper, No. 215

Provided in Cooperation with:
Department of Economics, University of Zurich

Suggested Citation: Ledoit, Olivier; Wolf, Michael (2017) : Numerical implementation of the QuEST function, Working Paper, No. 215, University of Zurich, Department of Economics, Zurich, https://doi.org/10.5167/uzh-120492

This Version is available at:
https://hdl.handle.net/10419/162418

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 215

Numerical Implementation of the QuEST Function

Olivier Ledoit and Michael Wolf

Revised version, January 2017

Numerical Implementation of the QuEST Function*

Olivier Ledoit
Department of Economics
University of Zurich
CH-8032 Zurich, Switzerland
olivier.ledoit@econ.uzh.ch

Michael Wolf
Department of Economics
University of Zurich
CH-8032 Zurich, Switzerland
michael.wolf@econ.uzh.ch

First version: January 2016
This version: January 2107

Abstract

This paper deals with certain estimation problems involving the covariance matrix in large dimensions. Due to the breakdown of finite-dimensional asymptotic theory when the dimension is not negligible with respect to the sample size, it is necessary to resort to an alternative framework known as large-dimensional asymptotics. Recently, Ledoit and Wolf (2015) have proposed an estimator of the eigenvalues of the population covariance matrix that is consistent according to a mean-square criterion under large-dimensional asymptotics. It requires numerical inversion of a multivariate nonrandom function which they call the QuEST function. The present paper explains how to numerically implement the QuEST function in practice through a series of six successive steps. It also provides an algorithm to compute the Jacobian analytically, which is necessary for numerical inversion by a nonlinear optimizer. Monte Carlo simulations document the effectiveness of the code.

KEY WORDS: Large-dimensional asymptotics, numerical optimization, random matrix theory, spectrum estimation.

*The authors wish to thank Edgar Dobriban (Stanford University), Jonathan Fletcher (University of Strathclyde), Matan Gavish (The Hebrew University of Jerusalem), Na Huang (London School of Economics and Political Science), Wen Jun (National University of Singapore), Tatsuya Kubokawa (University of Tokyo), Clifford Lam (London School of Economics and Political Science), Artyom Lepold (University of Kaiserslautern), Stefano Monni (American University of Beirut), Nestor Parolya (University of Hannover), Simon Welsing (Technical University of Munich), and Zhao Zhao (Huazhong University of Science and Technology) for testing early versions of this code. Any errors are ours.

1 Introduction

Many data sets in econometrics, biostatistics and electrical engineering, among a host of other fields, contain large numbers of related variables. The estimation of the covariance matrix poses challenging statistical problems when the dimension is not small relative to sample size. Approximations that are valid under traditional asymptotics, that is, when the dimension remains fixed while the sample size goes to infinity, perform poorly. This is why attention has turned to *large-dimensional asymptotics* where the dimension and the sample size go to infinity together, with their ratio converging to a finite, nonzero limit called the *concentration (ratio)*.

Under large-dimensional asymptotics, the sample eigenvalues are not consistent estimators of the population eigenvalues. A new estimator for the population eigenvalues under large-dimensional asymptotics was recently introduced by Ledoit and Wolf (2015). It hinges critically on a multivariate nonrandom function called the QuEST function. This acronym stands for *Quantized Eigenvalues Sampling Transform*. Ledoit and Wolf (2015) provide the mathematical definition of the QuEST function, but do not provide any details about numerical implementation. The problem of numerical implementation is non-trivial, due to the complexity of the definition of the QuEST function. A direct application of this method is the optimal estimation of the covariance matrix in the class of rotation-equivariant estimators introduced by Stein (1975, 1986) under various loss functions; see Ledoit and Wolf (2014b).

This paper explains how to numerically implement the QuEST function accurately and efficiently. In addition, given that the estimation of the population eigenvalues requires numerically inverting the QuEST function using a nonlinear optimizer, we also give the Jacobian analytically.

Section 2 reviews the literature on this subject. Section 3 gives the definition of the problem that will be solved numerically. Sections 4–9 describe in detail the six steps needed to implement the QuEST function numerically, delineating all the mathematical results that are needed along the way. Section 10 provides extensive Monte Carlo simulations. Section 11 concludes.

2 Literature Review

2.1 Estimation of the Population Covariance Matrix Eigenvalues

El Karoui (2008) proposed a way to estimate the empirical c.d.f. of population eigenvalues

under large-dimensional asymptotics using a different approach than the QuEST function. However, the code executing his algorithm was not made available to other researchers in the field, and those who tried to replicate it themselves did not enjoy much success. The state of affairs is aptly summarized by Li et al. (2013):

> Actually, the general approach in El Karoui (2008) has several implementation issues that seem to be responsible for its relatively low performance as attested by the very simple nature of provided simulation results.

There are three reasons why the same criticisms cannot be levelled against the QuEST function: first, a Matlab code implementing the QuEST function has already been used independently by Welsing (2015), Ito and Kubokawa (2015), Huang and Fryzlewicz (2015), and Lam (2016), among others; [1]; second, the present paper opens up the code of the QuEST function and its Jacobian to the general public for inspection and potential improvements; and third, Section 10 provides an extensive Monte Carlo study with nearly a third of a million simulations across a variety of challenging scenarios.

Apart from El Karoui (2008), other proposals have been put forward, making this field one of the most active ones in multivariate analysis in recent years.

- Rao et al. (2008) provide a solution when the population spectrum has a staircase structure, typically with half of the eigenvalues equal to one and the rest equal to two. The ability of this approach to handle the general case where there can be up to p distinct population eigenvalues, with p going to infinity, is not established.

- Mestre (2008) provides a solution when the concentration ratio $c_n := p/n$ is sufficiently small and/or the distinct population eigenvalues sufficiently far from one another, that is, when the sample eigenvalues display what is known as "spectral separation". This is a favorable situation where the sample eigenvalues are grouped into easily identifiable clusters, each cluster corresponding to one single population eigenvalue (which can have multiplicity higher than one). His Monte Carlo simulations assume no more than four distinct population eigenvalues.

  Although his algorithm is computationally elegant, the method is severely limited. It cannot be applied when (i) the limiting distribution of population eigenvalues is discrete but spectral separation does not occur or (ii) the limiting distribution of population eigenvalues has a continuous part. Many applications of interest fall under the latter category.

[1] The code can be downloaded at http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html under the link "Programming Code".

- Bai et al. (2010) propose a solution based on the method of moments when the parametric dimension of the population spectrum is finite. They demonstrate good behavior up to order four.

- Chen et al. (2011) elaborate on the previous paper by providing more rigorous justification of the method when the model order is unknown. But Monte Carlo simulations only go to order three.

- Yao et al. (2012) can be seen as a cross between the papers of Mestre (2008) and Bai et al. (2010), but also requiring a finite number of distinct population eigenvalues. In practice, Monte Carlo simulations provided by the authors do not go above three distinct population eigenvalues.

The common point between all these other methods is that they do not purport to address the general case. They work with a finite number of degrees of freedom (in practice no more than four) in the choice of the population spectral distribution, whereas the real number is p, which goes to infinity. This is why it is important to avoid the criticisms that have been levelled at the only other ostensibly general approach, that of El Karoui (2008), by fully explaining how to numerically implement the QuEST function, and by providing extensive Monte Carlo simulations showing that it works in practice under a wide variety of circumstances.

Finally, we should note that Dobriban (2015) also provides a numerical method for solving the so-called 'Fundamental Equation' of Random Matrix Theory; see Section 3. He does not compute the QuEST function explicitly, and does not provide the Jacobian analytically. As a result, numerical inversion is very difficult; but his paper is not focused on the problem of recovering the population eigenvalues.

2.2 Potential Applications

The numerical implementation of the QuEST function given in this paper is essential for the estimation of the population eigenvalues, which in turn is essential for nonlinear shrinkage estimation of the covariance matrix under large-dimensional asymptotics; see Ledoit and Wolf (2012, 2015). Many fields are interested in shrinking the covariance matrix when the number of variables is high:

Acoustics Optimally removing noise from signals captured from an array of hydrophones (Zhang et al., 2009).

Cancer Research Mapping out the influence of the Human Papillomavirus (HPV) on gene expression (Pyeon et al., 2007).

Chemistry Estimating the temporal autocorrelation function (TACF) for fluorescence correlation spectroscopy (Guo et al., 2012).

Civil Engineering Detecting and identifying vibration–based bridge damage through Random Coefficient Pooled (RCP) models (Michaelides et al., 2011).

Climatology Detecting trends in average global temperature through the optimal fingerprinting method (Ribes et al., 2013).

Econometrics Specifying the target covariance matrix in the Dynamic Conditional Correlation (DCC) model to capture time-series effects in the second moments (Hafner and Reznikova, 2012).

Electromagnetics Studying correlation between reverberation chamber measurements collected at different stirrer positions (Pirkl et al., 2012)

Entertainment Technology Designing a video game controlled by performing tricks on a skateboard (Anlauff et al., 2010).

Finance Reducing the risk in large portfolios of stocks (Jagannathan and Ma, 2003).

Genetics Inferring large-scale covariance matrices from functional genomic data (Schäfer and Strimmer, 2005).

Geology Modeling multiphase flow in subsurface petroleum reservoirs with the iterative stochastic ensemble method (ISEM) on inverse problems (Elsheikh et al., 2013).

Image Recognition Detecting anomalous pixels in hyperspectral imagery (Bachega et al., 2011).

Neuroscience Calibrating brain-computer interfaces (Lotte and Guan, 2009).

Psychology Modeling co-morbidity patterns among mental disorders (Markon, 2010).

Road Safety Research Developing an emergency braking assistance system (Haufe et al., 2011).

Signal Processing Combining data recorded by an array of sensors to minimize the noise (Chen et al., 2010).

Speech Recognition Automatically transcribing records of phone conversations (Bell and King, 2009).

Up until now, these fields have had to satisfy themselves with *linear* shrinkage estimation of the covariance matrix (Ledoit and Wolf, 2003, 2004). However this approach is asymptotically suboptimal in the class of rotation-equivariant estimators relative to nonlinear shrinkage, which requires numerical implementation of the QuEST function. The present paper makes this new and improved method universally available in practice.

3 Definition of the QuEST Function

The mathematical definition of the QuEST function is given by Ledoit and Wolf (2015); it is reproduced here for convenience. In particular, the large-dimensional asymptotic framework assumes that the number of population eigenvalues, p, goes to infinity together with the sample size, n, with their ratio $c_n \coloneqq p/n$ converging to a finite, positive constant $c \neq 1$, which is called the *limiting concentration (ratio)*. The reader is referred to Ledoit and Wolf (2015, Section 2.1) for a detailed description of the asymptotic framework.

For any positive integers n and p, the QuEST function, denoted by $Q_{n,p}$, is the nonrandom multivariate function given by

$$Q_{n,p} : [0,\infty)^p \longrightarrow [0,\infty)^p \tag{3.1}$$

$$\mathbf{t} \coloneqq (t_1,\ldots,t_p)' \longmapsto Q_{n,p}(\mathbf{t}) \coloneqq \left(q^1_{n,p}(\mathbf{t}),\ldots,q^p_{n,p}(\mathbf{t})\right)', \tag{3.2}$$

where

$$\forall i = 1,\ldots,p \qquad q^i_{n,p}(\mathbf{t}) \coloneqq p\int_{(i-1)/p}^{i/p} \left(F^{\mathbf{t}}_{n,p}\right)^{-1}(v)\,dv\ , \tag{3.3}$$

$$\forall v \in [0,1] \qquad \left(F^{\mathbf{t}}_{n,p}\right)^{-1}(v) \coloneqq \sup\{x \in \mathbb{R} : F^{\mathbf{t}}_{n,p}(x) \leq v\}\ , \tag{3.4}$$

$$\forall x \in \mathbb{R} \qquad F^{\mathbf{t}}_{n,p}(x) \coloneqq \begin{cases} \max\left(1-\dfrac{n}{p}, \dfrac{1}{p}\displaystyle\sum_{i=1}^{p} \mathbf{1}_{\{t_i=0\}}\right) & \text{if } x = 0\ , \\ \displaystyle\lim_{\eta\to 0^+} \frac{1}{\pi}\int_{-\infty}^{x} \mathsf{Im}\left[m^{\mathbf{t}}_{n,p}(\xi+i\eta)\right]\,d\xi & \text{otherwise}\ , \end{cases}\ , \tag{3.5}$$

and $\forall z \in \mathbb{C}^+ \quad m \coloneqq m^{\mathbf{t}}_{n,p}(z)$ is the unique solution in the set

$$\left\{m \in \mathbb{C} : -\frac{n-p}{nz} + \frac{p}{n}\,m \in \mathbb{C}^+\right\} \tag{3.6}$$

to the equation

$$m = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{t_i \left(1 - \frac{p}{n} - \frac{p}{n} z\, m\right) - z} \,. \tag{3.7}$$

The QuEST function is a natural discretization of equation (1.4) of Silverstein (1995), which is itself a reformulation of equation (1.14) of Marčenko and Pastur (1967). We refer to this as the 'Fundamental Equation' (FE) of Random Matrix Theory. For detailed description, see Ledoit and Wolf (2015, Section 2.1).

The basic idea is that p represents the matrix dimension, n the sample size, $\mathbf{t} := (t_1, \ldots, t_p)'$ the population eigenvalues, $Q_{n,p}(\mathbf{t}) := \left(q^1_{n,p}(\mathbf{t}), \ldots, q^p_{n,p}(\mathbf{t})\right)'$ the sample eigenvalues, $F^{\mathbf{t}}_{n,p}$ the limiting empirical c.d.f of sample eigenvalues, and $m^{\mathbf{t}}_{n,p}$ its Stieltjes (1894) transform. A fundamental result in large-dimensional asymptotics is that the relationship between the population spectral distribution and the sample spectral distribution is nonrandom in the limit. Figure 1, publicized by Jianfeng Yao (2015) in a conference presentation, gives a heuristic view of the area where Marčenko-Pastur asymptotic theory is more useful (labelled "MP area") vs. the area where standard fixed-dimension asymptotic theory applies (labelled "Low-dim area").

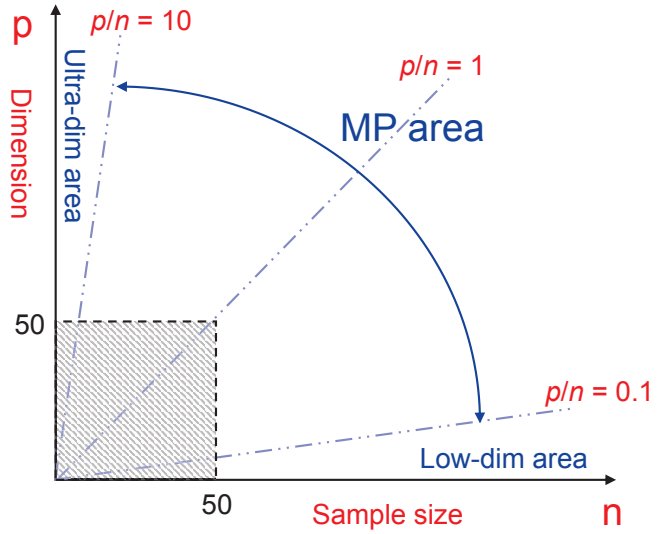

Figure 1: Heuristic comparison of the area of relevance of Marčenko-Pastur asymptotics vs. traditional fixed-dimension asymptotics.

This insight is further developed in the recent book by Yao et al. (2015). Readers interested in the background from probability theory may also consult the authoritative monograph by Bai and Silverstein (2010).

The importance of the QuEST function is twofold. First, inverting it numerically yields an estimator of the population eigenvalues that is consistent under large-dimensional

asymptotics. Second, once this has been achieved, it is possible to use Theorem 2 of Ledoit and Péché (2011) to construct shrinkage estimators of the covariance matrix that are asymptotically optimal with respect to a given loss function in the p-dimensional space of rotation-equivariant estimators introduced by Stein (1975, 1986). Ledoit and Wolf (2014b) derive the optimal shrinkage formula for five different loss functions, and Ledoit and Wolf (2014a) for a sixth.

The numerical implementation of the QuEST function consists of a series of six successive operations: 1) finding the support of $F_{n,p}^{\mathbf{t}}$; 2) choosing a grid that covers the support; 3) solving equation (3.7) on the grid; 4) computing the sample spectral density; 5) integrating it to obtain the empirical c.d.f. of sample eigenvalues; and 6) interpolating the c.d.f to compute sample eigenvalues as per equation (3.3). Each of these steps is detailed below.

4 Support

In what follows we omit the subscripts and superscript of $F_{n,p}^{\mathbf{t}}$ and c_n in order to simplify the notation. We do not work directly with F but with u, which is defined by

$$\begin{aligned} u \coloneqq u(z) &\coloneqq -\frac{1}{m_{\underline{F}}(z)} \\ m_{\underline{F}}(z) &\coloneqq \frac{c-1}{z} + c\, m_F(z) \\ m_F(z) &\coloneqq \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dF(\lambda) \ . \end{aligned}$$

There is a direct mapping between F-space and u-space, as explained in Ledoit and Wolf (2012, Section 2). Numerically it is more judicious to work in u-space.

To determine the image of the support of F in u-space, we first need to group together the population eigenvalues $\tau_1, \ldots, \tau_p$ that are equal to one another and, if necessary, discard those that are equal to zero. Let us say that there are K distinct nonzero population eigenvalues $0 < t_1 < \ldots < t_K$. We can associate them with their respective weights: if j elements of the vector $(\tau_1, \ldots, \tau_p)$ are equal to t_k then the corresponding weight is $w_k \coloneqq j/p$.

4.1 Spectral Separation

Now we look for spectral separation between t_k and t_{k+1} $(k = 1, \ldots, K-1)$. This is done in two stages. First we run a quick test to see whether we can rule out spectral separation

a priori. Second, if the test is inconclusive, we do the full analysis to ascertain whether spectral separation does indeed occur.

4.1.1 Necessary Condition for Spectral Separation

Spectral separation occurs at some $u \in (t_k, t_{k+1})$ if and only if

$$\forall v \in (0, +\infty) \qquad \mathsf{Im}\left[(u+iv) - c(u+iv)\sum_{j=1}^{K} \frac{w_j t_j}{t_j - (u+iv)}\right] \neq 0 \ ,$$

which is equivalent to

$$\sum_{j=1}^{K} \frac{w_j t_j^2}{(t_j - u)^2} < \frac{1}{c} \ . \tag{4.1}$$

Equation (4.1) is equivalent to the function $x_F(m)$ defined in equation (1.6) of Silverstein and Choi (1995) being strictly increasing at $m = -1/u$. Silverstein and Choi (1995, Section 4) explain how this enables us to determine the support.

Call $\varphi(u)$ the function on the left-hand side of equation (4.1). We can decompose it into

$$\begin{aligned} \varphi(u) &= \theta_k(u) + \psi_k^L(u) + \psi_k^R(u) \ , \\ \text{where} \qquad \theta_k(u) &\coloneqq \frac{w_k t_k^2}{(t_k - u)^2} + \frac{w_{k+1} t_{k+1}^2}{(t_{k+1} - u)^2} \ , \\ \psi_k^L(u) &\coloneqq \sum_{j=1}^{k-1} \frac{w_j t_j^2}{(t_j - u)^2} \ , \\ \text{and} \qquad \psi_k^R(u) &\coloneqq \sum_{j=k+2}^{K} \frac{w_j t_j^2}{(t_j - u)^2} \ . \end{aligned}$$

It is easy to see that the function $\theta_k(\cdot)$ is convex over the interval (t_k, t_{k+1}), diverges to $+\infty$ near t_k and t_{k+1}, and attains its minimum at

$$\widehat{x}_k \coloneqq (t_k t_{k+1})^{2/3} \, \frac{w_k^{1/3} t_{k+1}^{1/3} + w_{k+1}^{1/3} t_k^{1/3}}{w_k^{1/3} t_k^{2/3} + w_{k+1}^{1/3} t_{k+1}^{2/3}} \ , \tag{4.2}$$

therefore a lower bound for $\theta_k(\cdot)$ on (t_k, t_{k+1}) is $\theta_k(\widehat{x}_k)$.

It is also easy to see that the function $\psi_k^L(\cdot)$ is decreasing over the interval (t_k, t_{k+1}); therefore, it attains its minimum at t_{k+1} and is bounded from below by $\psi_k^L(t_{k+1})$. Conversely, the function $\psi_k^R(\cdot)$ is increasing over the interval (t_k, t_{k+1}), attains its minimum at t_k and

is bounded from below by $\psi_k^R(t_k)$. Putting these three results together yields the following lower bound for $\varphi(\cdot)$:

$$\forall u \in (t_k, t_{k+1}) \qquad \varphi(u) \geq \frac{w_k t_k^2}{(t_k - \widehat{x}_k)^2} + \frac{w_{k+1} t_{k+1}^2}{(t_{k+1} - \widehat{x}_k)^2} + \sum_{j=1}^{k-1} \frac{w_j t_j^2}{(t_j - t_{k+1})^2} + \sum_{j=k+2}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2} \ ,$$

where $\widehat{x}_k$ is given by equation (4.2).

Combining this bound with equation (4.1) means that

$$\frac{w_k t_k^2}{(t_k - \widehat{x}_k)^2} + \frac{w_{k+1} t_{k+1}^2}{(t_{k+1} - \widehat{x}_k)^2} + \sum_{j=1}^{k-1} \frac{w_j t_j^2}{(t_j - t_{k+1})^2} + \sum_{j=k+2}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2} < \frac{1}{c} \tag{4.3}$$

is a necessary (but not sufficient) condition for spectral separation to occur between t_k and t_{k+1}, that is, at *some* $u \in (t_k, t_{k+1})$. Thus, the numerical procedure can be made more efficient by first computing the quantity on the left-hand side of equation (4.3), comparing it to $1/c$, and asserting the absence of spectral separation in the interval (t_k, t_{k+1}) in the case where it is higher than $1/c$. If, on the other hand, it is strictly lower than $1/c$, then further work is needed to ascertain whether spectral separation does indeed occur. In practice, checking this condition seems to save a lot of time by eliminating many intervals (t_k, t_{k+1}), except perhaps when c is very small and the population eigenvalues are very spread out.

4.1.2 Necessary and Sufficient Condition for Spectral Separation

Consider now some $k \in \{1, 2, \ldots, K-1\}$ for which the condition in equation (4.3) holds. Given equation (4.1), we need to find the minimum of $\varphi(\cdot)$ over (t_k, t_{k+1}) and compare it to $1/c$. It is easy to check that $\varphi(\cdot)$ is strictly convex over (t_k, t_{k+1}), therefore this minimum exists, is unique, and is the only zero in (t_k, t_{k+1}) of the derivative function

$$\varphi'(u) = 2 \sum_{j=1}^{K} \frac{w_j t_j^2}{(t_j - u)^3} \ .$$

Most numerical algorithms that find the zero of a function require as inputs two points $\underline{x}$ and $\overline{x}$ such that the sign of the function is not the same at $\underline{x}$ as at $\overline{x}$. Finding two such points is the goal of the next step. There are three cases, depending on the sign of $\varphi'(\widehat{x}_k)$.

- $\varphi'(\widehat{x}_k) = 0$: Then the search is immediately over because $\varphi(\cdot)$ attains its minimum at $x_k^* := \widehat{x}_k$. This would not happen generically unless $K = 2$.
- $\varphi'(\widehat{x}_k) < 0$: In this case, given that $\varphi'(\cdot)$ is strictly increasing, the minimizer of $\varphi(\cdot)$

lies in the interval $(\widehat{x}_k, t_{k+1})$. We can feed the lower bound $\underline{x} = \widehat{x}_k$ into the numerical procedure that will find the zero of $\varphi'(\cdot)$. It would be also tempting to set $\overline{x} := t_{k+1}$, but unfortunately doing so would not be practicable because $\lim_{u \nearrow t_{k+1}} \varphi'(u) = +\infty$, and most numerical procedures perform poorly near singularity points. Therefore we need to find some $\overline{x} \in (\widehat{x}_k, t_{k+1})$ such that $\varphi'(\overline{x}) > 0$. Let x_k^* denote the unique value in $(\widehat{x}_k, t_{k+1})$ such that $\varphi'(x_k^*) = 0$. Then the fact that $w_j t_j^2/(t_j - u)^3$ is increasing in u for any $j \in \{1, \ldots, K\}$ implies that the following inequalities hold:

$$
\begin{aligned}
\forall u \in (\widehat{x}_k, t_{k+1}) \qquad \varphi'(u) &> 2\frac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} + 2\frac{w_{k+1} t_{k+1}^2}{(t_{k+1} - u)^3} + \psi_k^{L\prime}(\widehat{x}_k) + \psi_k^{R\prime}(\widehat{x}_k) \\
0 &> 2\frac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} + 2\frac{w_{k+1} t_{k+1}^2}{(t_{k+1} - x_k^*)^3} + \psi_k^{L\prime}(\widehat{x}_k) + \psi_k^{R\prime}(\widehat{x}_k) \\
-2\frac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} - \psi_k^{L\prime}(\widehat{x}_k) - \psi_k^{R\prime}(\widehat{x}_k) &> 2\frac{w_{k+1} t_{k+1}^2}{(t_{k+1} - x_k^*)^3} \\
t_{k+1} - x_k^* &> \left(\frac{2w_{k+1} t_{k+1}^2}{-2\dfrac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} - \psi_k^{L\prime}(\widehat{x}_k) - \psi_k^{R\prime}(\widehat{x}_k)} \right)^{1/3} \\
x_k^* &< t_{k+1} - \left(\frac{2w_{k+1} t_{k+1}^2}{-2\dfrac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} - \psi_k^{L\prime}(\widehat{x}_k) - \psi_k^{R\prime}(\widehat{x}_k)} \right)^{1/3} \\
x_k^* &< t_{k+1} - \left(\frac{2w_{k+1} t_{k+1}^2}{-2\dfrac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} - \varphi'(\widehat{x}_k)} \right)^{1/3},
\end{aligned}
$$

where the last inequality follows from $\theta_k'(\widehat{x}_k) = 0$. Thus if we set

$$
\overline{x} := t_{k+1} - \left(\frac{2w_{k+1} t_{k+1}^2}{-2\dfrac{w_k t_k^2}{(t_k - \widehat{x}_k)^3} - \varphi'(\widehat{x}_k)} \right)^{1/3},
$$

we know that $\varphi'(\overline{x}) > 0$. Launching a zero-finding algorithm for $\varphi'(\cdot)$ on the interval $[\underline{x}, \overline{x}]$ as defined above yields a unique solution x_k^*.

- $\varphi'(\widehat{x}_k) < 0$: A similar line of reasoning points us to

$$\underline{x} \coloneqq t_k + \left(\frac{2w_k t_k^2}{2\dfrac{w_{k+1} t_{k+1}^2}{(t_{k+1} - \widehat{x}_k)^3} + \varphi'(\widehat{x}_k)} \right)^{1/3},$$

  $\overline{x} = \widehat{x}_k$, and yields a unique zero x_k^* for $\varphi'(\cdot)$ over the interval $[\underline{x}, \overline{x}]$.

Across all three cases, the outcome of this procedure is $x_k^* = \operatorname{argmin}_{u \in (t_k, t_{k+1})} \varphi(u)$. Spectral separation occurs between t_k and t_{k+1} if and only if $\varphi(x_k^*) < 1/c$.

If there is no spectral separation, then we can dismiss the interval (t_k, t_{k+1}); otherwise, we need some additional work to compute spectrum boundaries.

4.1.3 Interval Boundaries

Consider now some $k \in \{1, 2, \ldots, K-1\}$ for which $x_k^* = \operatorname{argmin}_{u \in (t_k, t_{k+1})} \varphi(u)$ is known and $\varphi(x_k^*) < 1/c$. Spectral separation means that the support ends at some point in (t_k, x_k^*) and starts again at some point in (x_k^*, t_{k+1}). The equation that characterizes support endpoints is $\varphi(x) = 1/c$. Thus we need to find two zeros of the function $\varphi(\cdot) - 1/c$, one in the interval (t_k, x_k^*) and the other in the interval (x_k^*, t_{k+1}).

Let us start with the first zero of the function $\varphi(\cdot) - 1/c$, the one that lies in the interval (t_k, x_k^*). Once again, we employ an off-the-shelf univariate zero-finding routine that takes as inputs two points $\underline{x}$ and $\overline{x}$ such that $\varphi(\underline{x}) > 1/c$ and $\varphi(\overline{x}) < 1/c$. The obvious candidate for $\overline{x}$ is $\overline{x} \coloneqq x_k^*$. For $\underline{x}$, however, we cannot use t_k because $\lim_{x \searrow t_k} \varphi(x) = +\infty$. Therefore we need to find some $\underline{x} \in (t_k, x_k^*)$ that verifies $\varphi(\underline{x}) > 1/c$. Such an $\underline{x}$ can be found by considering the following series of inequalities, which hold for all $x \in (t_k, x_k^*)$:

$$\begin{aligned}
\varphi(x) &> \frac{w_k t_k^2}{(t_k - x)^2} + \sum_{j=1}^{k-1} \frac{w_j t_j^2}{(t_j - x_k^*)^2} + \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2} \\
\varphi(x) - \varphi(x_k^*) &> \frac{w_k t_k^2}{(t_k - x)^2} - \frac{w_k t_k^2}{(t_k - x_k^*)^2} + \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2} - \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - x_k^*)^2} \\
\varphi(x) - \frac{1}{c} &> \frac{w_k t_k^2}{(t_k - x)^2} - \frac{w_k t_k^2}{(t_k - x_k^*)^2} + \left[\varphi(x_k^*) - \frac{1}{c} \right] + \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2} - \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - x_k^*)^2}.
\end{aligned}$$

Call the function on the right-hand side of the last inequality $g_k(x)$. This function is defined

on $(t_k, +\infty)$. It is of the form

$$g_k(x) = \frac{w_k t_k^2}{(t_k - x)^2} + C_k \ ,$$

where C_k is a negative constant. $g_k(\cdot)$ is strictly decreasing, with $\lim_{x \searrow t_k} g_k(x) = +\infty$, and $\lim_{x \nearrow +\infty} g_k(x) = C_k < 0$. Therefore, it admits a unique zero in $(t_k, +\infty)$, given by

$$\underline{x} \coloneqq t_k + \sqrt{\frac{w_k t_k^2}{\dfrac{w_k t_k^2}{(t_k - x_k^*)^2} + \left[\dfrac{1}{c} - \varphi(x_k^*)\right] + \displaystyle\sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - x_k^*)^2} - \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2}}} \ .$$

Notice that

$$g_k(x_k^*) = \left[\varphi(x_k^*) - \frac{1}{c}\right] + \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - t_k)^2} - \sum_{j=k+1}^{K} \frac{w_j t_j^2}{(t_j - x_k^*)^2} < 0.$$

Combining this fact with the fact that $g_k(\cdot)$ is strictly decreasing implies that $\underline{x} < x_k^*$.

Since $g_k(\underline{x}) = 0$ by construction, $\varphi(\underline{x}) > 1/c$. Feeding $(\underline{x}, \overline{x})$ thus defined into the zero-finding numerical routine with the function $\varphi(\cdot) - 1/c$ yields an endpoint of the support.

A similar line of reasoning leads to setting $\underline{x} \coloneqq x_k^*$,

$$\overline{x} \coloneqq t_{k+1} - \sqrt{\frac{w_{k+1} t_{k+1}^2}{\dfrac{w_{k+1} t_{k+1}^2}{(t_{k+1} - x_k^*)^2} + \left[\dfrac{1}{c} - \varphi(x_k^*)\right] + \displaystyle\sum_{j=1}^{k-1} \frac{w_j t_j^2}{(t_j - x_k^*)^2} - \sum_{j=1}^{k-1} \frac{w_j t_j^2}{(t_j - t_{k+1})^2}}} \ ,$$

and running a numerical routine to find a zero of the function $\varphi(\cdot) - 1/c$ on the interval $(\underline{x}, \overline{x}) \subset (x_k^*, t_{k+1})$. This zero will also be a support endpoint.

4.2 Extremities of the Support

The procedure described so far identifies all support endpoints lying in the interval $[t_1, t_K]$. In order to complete the determination of the support, we must find the support endpoint that lies in the interval $(-\infty, t_1)$ and the support endpoint that lies in the interval $(t_K, +\infty)$.

4.2.1 Minimum of the Support

Let us start with the first support endpoint, the one lying in the interval $(-\infty, t_1)$. The equation that characterizes this point is the same as before: $\varphi(x) = 1/c$. In order to

employ the zero-finding numerical routine, we must find two bounds $\underline{x}$ and $\overline{x}$, both strictly less than t_1, such that $\varphi(\underline{x}) < 1/c$ and $\varphi(\overline{x}) > 1/c$. The left-hand side bound $\underline{x}$ can be obtained by considering the following inequalities:

$$\forall x \in (-\infty, t_1) \quad \forall j = 1, \ldots, K \qquad \frac{w_j t_j^2}{(x - t_j)^2} \leq \frac{w_j t_j^2}{(x - t_1)^2}$$

$$\forall x \in (-\infty, t_1) \qquad \varphi(x) \leq \frac{\sum_{j=1}^{K} w_j t_j^2}{(x - t_1)^2} \ . \tag{4.4}$$

Note that if we set

$$\underline{x} \coloneqq t_1 - \sqrt{c \sum_{j=1}^{K} w_j t_j^2} - 1 \ ,$$

then

$$\frac{\sum_{j=1}^{K} w_j t_j^2}{(\underline{x} - t_1)^2} < \frac{1}{c} \ ,$$

which in turn implies by equation (4.4) that $\varphi(\underline{x}) < 1/c$, as desired.

The right-hand side bound $\overline{x}$ can be found by considering a different inequality:

$$\forall x \in (-\infty, t_1) \qquad \varphi(x) \geq \frac{w_1 t_1^2}{(x - t_1)^2} \ . \tag{4.5}$$

Note that if we set

$$\overline{x} \coloneqq t_1 - \frac{\sqrt{c w_1 t_1^2}}{2} \ ,$$

then

$$\frac{w_1 t_1^2}{(\overline{x} - t_1)^2} > \frac{1}{c} \ ,$$

which in turn implies by equation (4.5) that $\varphi(\overline{x}) > 1/c$, as desired. Launching the numerical routine to find a zero of the function $\varphi(\cdot) - 1/c$ over the interval $(\underline{x}, \overline{x})$ thus defined yields the first endpoint of the support.

4.2.2 Maximum of the Support

For the last endpoint of the support, the one that lies in the interval $(t_K, +\infty)$, a similar line of reasoning leads us to define

$$\underline{x} \coloneqq t_K + \frac{\sqrt{c w_K t_K^2}}{2} \qquad \text{and} \qquad \overline{x} \coloneqq t_K + \sqrt{c \sum_{j=1}^{K} w_j t_j^2} + 1 \ .$$

Launching the numerical routine to find a zero of the function $\varphi(\cdot) - 1/c$ over the interval $(\underline{x}, \overline{x})$ thus defined yields the last endpoint of the support.

4.3 Output

The main outputs of this procedure are $\nu \geq 1$, the number of distinct intervals that constitute the support, and $u_1, \ldots, u_{2\nu}$, the support endpoints. The support in u-space is $S_U = [u_1, u_2] \cup \cdots \cup [u_{2\nu-1}, u_{2\nu}]$.

Another output of this procedure is a set of positive integers $\omega_1, \ldots, \omega_\nu$ summing up to p that tell us how many population eigenvalues correspond to each support interval. If $\nu = 1$ then there is no spectral separation and $\omega_1 = p$. If $\nu \geq 2$ and the first spectral separation occurs between t_k and t_{k+1} for some $k = 1, \ldots, K-1$, then $\omega_1 = p \sum_{j=1}^{k} w_j$. If some population eigenvalues are equal to zero, then ω_1 needs to be augmented accordingly; see the discussion below Lemma 4 of Bai and Silverstein (1999) for details.

If $\nu \geq 2$ and the last spectral separation occurs between $t_{k'}$ and $t_{k'+1}$ for some $k' = 1, \ldots, K-1$, then $\omega_\nu = p \sum_{j=k'+1}^{K} w_j$. If $\nu \geq 3$ and the i^{th} support interval (for $i = 2, \ldots, \nu-1$) is delimited on the left-hand side by spectral separation occurring between t_k and t_{k+1}, and on the right-hand side by spectral separation occurring between $t_{k'}$ and $t_{k'+1}$ (where $1 \leq k < k' \leq K$), then $\omega_i = p \sum_{j=k+1}^{k'} w_j$. This information will turn out to be useful in subsequent operations.

4.4 Derivative of the Support Endpoints

If the QuEST function defined by equations (3.1)–(3.7) is to be used efficiently in an optimization algorithm, it is desirable to be able to compute its derivative analytically. Since this function is constructed as a chain of six successive operations, the first of which is the determination of support endpoints, its derivative can be computed in the same way, provided that we start by computing analytically the derivative of support endpoints with respect to τ_k for all $k = 1, \ldots, K$.

Every u_i for $i = 1, \ldots, 2\nu$ is a zero of the function

$$\widetilde{\varphi}(u; \tau_1, \ldots, \tau_p) := \frac{1}{p} \sum_{j=1}^{p} \frac{\tau_j^2}{(\tau_j - u)^2} - \frac{1}{c} \ .$$

By differentiating the equation $\widetilde{\varphi}(u; \tau_1, \ldots, \tau_p) = 0$ we get:

$$\frac{\partial \widetilde{\varphi}}{\partial u} \cdot du + \frac{\partial \widetilde{\varphi}}{\partial \tau_k} \cdot d\tau_k = 0 \ ,$$

$$\frac{\partial u}{\partial \tau_k} = -\frac{\dfrac{\partial \widetilde{\varphi}}{\partial \tau_k}}{\dfrac{\partial \widetilde{\varphi}}{\partial u}} \ .$$

The partial derivatives of the function $\widetilde{\varphi}$ are as follows:

$$\frac{\partial \widetilde{\varphi}}{\partial u}(u; \tau_1, \ldots, \tau_p) = \frac{2}{p} \sum_{j=1}^{p} \frac{\tau_j^2}{(\tau_j - u)^3}$$
$$\frac{\partial \widetilde{\varphi}}{\partial \tau_k}(u; \tau_1, \ldots, \tau_p) = -\frac{2}{p} \frac{\tau_k u}{(\tau_k - u)^3} \ ;$$

therefore,

$$\forall i = 1, \ldots, 2\nu \quad \forall k = 1, \ldots, p \qquad \frac{\partial u_i}{\partial \tau_k} = \frac{\dfrac{\tau_k u_i}{(\tau_k - u_i)^3}}{\displaystyle\sum_{j=1}^{p} \frac{\tau_j^2}{(\tau_j - u_i)^3}} \ . \tag{4.6}$$

5 Grid

The first operation generated the support in u-space $S_U = [u_1, u_2] \cup \cdots \cup [u_{2\nu-1}, u_{2\nu}]$ and the number of population eigenvalues corresponding to each interval: $\omega_1, \ldots, \omega_\nu$. The goal of the second operation is to produce a grid that covers this support. This problem can be broken down by considering each interval $i = 1, \ldots, \nu$ separately.

5.1 Formula for the Grid Points

Take some $i \in \{1, \ldots, \nu\}$. How shall we determine a grid that covers the interval $[u_{2i-1}, u_{2i}]$? The number of points on the grid will be a function of ω_i. Specifically, we shall take ω_i points in the open interval (u_{2i-1}, u_{2i}), plus the two endpoints u_{2i-1} and u_{2i}. Thus, the total number of points covering the closed interval $[u_{2i-1}, u_{2i}]$ will be $\omega_i + 2$. Let us call these points $\xi_0^i, \ldots, \xi_{\omega_i+1}^i$, with the convention that $\xi_0^i := u_{2i-1}$ and $\xi_{\omega_i+1}^i := u_{2i}$. Thus, what is left is to define $\xi_1^i, \ldots, \xi_{\omega_i}^i$.

There are many ways to choose such a grid, depending on how densely we want to cover the various parts of the interval. The simplest idea would be to have uniform coverage through a linearly spaced grid. But it is more judicious to increase coverage density near the edges of the interval because this is where a lot of the action is taking place. Silverstein and Choi (1995) demonstrate that the limiting density of sample eigenvalues has "square root"-type behavior near boundary points. This fact points us towards the

inverse c.d.f. function of the beta distribution with parameters $(0.5, 0.5)$, also known as the *arcsine* distribution:

$$\forall j \in \{0, \ldots, \omega_i + 1\} \qquad \xi_j^i := u_{2i-1} + (u_{2i} - u_{2i-1}) \sin^2\left[\frac{\pi j}{2(\omega_i + 1)}\right] . \tag{5.1}$$

Compared to the beta distribution with parameters $(1, 1)$, which is the uniform distribution, reducing both parameters from 1 to 0.5 increases coverage density near the edges of the interval. Note that the density of the arcsine distribution goes to infinity at the edges of the interval (as does the derivative of the square root function), but the c.d.f., its inverse and the grid all remain well-behaved. The goal here is to enhance numerical accuracy.

5.2 Derivative of the Grid Points

In keeping with our earlier stated objective (see Section 4.4) of building towards an analytical formula for the partial derivative of λ_i with respect to τ_k for all $i, k \in \{1, \ldots, p\}$, at this stage we need to compute $\partial \xi_j^i / \partial \tau_k$ for all $j \in \{1, \ldots, \omega_i\}$. From equation (5.1) we can see immediately that it is

$$\frac{\partial \xi_j^i}{\partial \tau_k} = \left\{1 - \sin^2\left[\frac{\pi j}{2(\omega_i + 1)}\right]\right\} \frac{\partial u_{2i-1}}{\partial \tau_k} + \sin^2\left[\frac{\pi j}{2(\omega_i + 1)}\right] \frac{\partial u_{2i}}{\partial \tau_k} , \tag{5.2}$$

where $\partial u_{2i-1} / \partial \tau_k$ and $\partial u_{2i} / \partial \tau_k$ are given by equation (4.6).

6 Solving the Fundamental Equation in u-Space

In this section we will assume that the interval index $i \in \{1, \ldots, \nu\}$ is fixed.

6.1 Statement of the Problem

Given a grid coverage $(\xi_j^i)_{j=0,\ldots,\omega_i}$ of the i^{th} support interval, the third operation solves the Fundamental Equation at ξ_j^i. For every $j = 0, \ldots, \omega_i + 1$, define the function

$$\forall y \in [0, +\infty) \qquad \Gamma_j^i(y) := \frac{1}{p} \sum_{k=1}^{p} \frac{\tau_k^2}{(\tau_k - \xi_j^i)^2 + y^2} - \frac{1}{c}.$$

Since by definition of the grid ξ_j^i lies in the support S_U, equation (4.1) implies that $\Gamma_j^i(0) \geq 0$. In addition, it is easy to verify that Γ_j^i is strictly decreasing on $[0, +\infty)$ and that $\lim_{y \to +\infty} \Gamma_j^i(y) = -1/c$.

The solution to the Fundamental Equation at ξ_j^i is the thus unique $y \in [0, +\infty)$ such that

$$\Gamma_j^i(y) = 0 \ . \tag{6.1}$$

Call it y_j^i; this line of attack is directly inspired by Ledoit and Wolf (2012, Section 2.3).

From the definition of the ξ_j^i's in Section 5.1, it is obvious that $y_1^i = y_{\omega_i+1}^i = 0$. What remains to be determined is $(y_j^i)_{j=1,\ldots,\omega_i}$. In the remainder of this section we will assume that j is fixed in the set $\{1, \ldots, \omega_i\}$.

The solution y to the equation $\Gamma_j^i(y) = 0$ is computed by some standard numerical routine that finds the zero of a real univariate function. As usual, we need to input into this routine a lower bound $\underline{y}_j^i \in [0, +\infty)$ such that $\Gamma_j^i(\underline{y}_j^i) \in (0, +\infty)$ and an upper bound $\overline{y}_j^i \in (0, +\infty)$ such that $\Gamma_j^i(\overline{y}_j^i) < 0$.

6.2 Lower Bound

From Section 4, $(t_1, \ldots, t_K)$ is the vector of unique nonzero population eigenvalues, with corresponding weights $(w_1, \ldots, w_K)$. Let $\delta_j^i \coloneqq \min_{k \in \{1,\ldots,K\}}(t_k - \xi_i^j)^2$ and $\Omega_i^j \coloneqq \left\{k \in \{1, \ldots, K\} : (t_k - \xi_i^j)^2 = \delta_j^i\right\}$. Then we have

$$\Gamma_j^i(y) \geq \frac{\sum_{k \in \Omega_i^j} w_k t_k^2}{\delta_j^i + y^2} - \frac{1}{c} \ . \tag{6.2}$$

Looking at the right-hand side of equation (6.2), we see that

$$\frac{\sum_{k \in \Omega_i^j} w_k t_k^2}{\delta_j^i + y^2} - \frac{1}{c} \geq 0 \iff y^2 \leq c \sum_{k \in \Omega_i^j} w_k t_k^2 - \delta_j^i \ .$$

Therefore, if we set

$$\underline{y}_j^i \coloneqq \frac{\sqrt{\max\left(0, c \sum_{k \in \Omega_i^j} w_k t_k^2 - \delta_j^i\right)}}{2} \ ,$$

then $\Gamma(\underline{y}_j^i) \in (0, +\infty)$, as desired.

6.3 Upper Bound

We use the inequalities

$$\forall k \in \{1, \ldots, K\} \qquad \frac{1}{(t_k - \xi_j^i)^2 + y^2} \leq \frac{1}{\delta_j^i + y^2}$$

$$\Gamma_j^i(y) \leq \frac{\sum_{k=1}^K w_k t_k^2}{\delta_j^i + y^2} - \frac{1}{c} \ . \tag{6.3}$$

Note that if we set

$$\overline{y}_j^i := \sqrt{c \sum_{k=1}^K w_k t_k^2 - \delta_j^i + 1} \ ,$$

then

$$\frac{\sum_{k=1}^K w_k t_k^2}{\delta_j^i + (\overline{y}_j^i)^2} - \frac{1}{c} < 0 \ ;$$

therefore, by equation (6.3), $\Gamma_j^i(\overline{y}_j^i) < 0$, as desired.

6.4 Output

Launching a standard numerical routine to find the zero of the function $\Gamma_j^i(\cdot)$ over the interval $(\underline{y}_j^i, \overline{y}_j^i)$ yields y_j^i, the solution to the Fundamental Equation at ξ_j^i. The output of this operation is more conveniently expressed as the complex number $z_j^i := \xi_j^i + \sqrt{-1}\, y_j^i$.

6.5 Derivative

The derivative of the real part of z_j^i with respect to τ_k has been computed in Section 5.2. As for the derivative of the imaginary part, y_j^i, consider the function

$$\widetilde{\Gamma}_j^i(y; \tau_1, \ldots, \tau_p) := \frac{1}{p} \sum_{k=1}^p \frac{\tau_k^2}{\left(\tau_k - \xi_j^i\right)^2 + y^2} - \frac{1}{c} \ .$$

We can view y_j^i as a function of $(\tau_1, \ldots, \tau_p)$: $y_j^i = \widetilde{y}_j^i(\tau_1, \ldots, \tau_p)$. Then the manner in which y_j^i is obtained in Section 6.1 can be expressed through the equation

$$\widetilde{\Gamma}_j^i \left(\widetilde{y}_j^i(\tau_1, \ldots, \tau_p); \tau_1, \ldots, \tau_p \right) = 0.$$

Taking the partial derivative with respect to τ_k while holding the other population eigenvalues constant yields

$$\frac{\partial \widetilde{\Gamma}_j^i}{\partial y} \cdot \frac{\partial \widetilde{y}_j^i}{\partial \tau_k} + \frac{\partial \widetilde{\Gamma}_j^i}{\partial \tau_k} = 0 \ ,$$

$$\frac{\partial \widetilde{y}_j^i}{\partial \tau_k} = -\frac{\dfrac{\partial \widetilde{\Gamma}_j^i}{\partial \tau_k}}{\dfrac{\partial \widetilde{\Gamma}_j^i}{\partial y}} \ .$$

The partial derivatives of the function $\widetilde{\Gamma}_j^i$ are

$$\frac{\partial \widetilde{\Gamma}_j^i}{\partial \tau_k}(y; \tau_1, \ldots, \tau_p) = \frac{2\tau_k}{\left(\tau_k - \xi_j^i\right)^2 + y^2} - \frac{2\tau_k^2\left(\tau_k - \xi_j^i\right)}{\left[\left(\tau_k - \xi_j^i\right)^2 + y^2\right]^2}$$
$$\frac{\partial \widetilde{\Gamma}_j^i}{\partial y}(y; \tau_1, \ldots, \tau_p) = -2\sum_{l=1}^{p} \frac{\tau_l^2 y}{\left[\left(\tau_l - \xi_j^i\right)^2 + y^2\right]^2} \ .$$

Therefore,

$$\frac{\partial \widetilde{y}_j^i}{\partial \tau_k}(\tau_1, \ldots, \tau_p) = \frac{\dfrac{\tau_k}{(\tau_k - \xi_j^i)^2 + (y_j^i)^2} - \dfrac{\tau_k^2(\tau_k - \xi_j^i)}{[(\tau_k - \xi_j^i)^2 + (y_j^i)^2]^2}}{\displaystyle\sum_{l=1}^{p} \frac{\tau_l^2\, y_j^i}{[(\tau_l - \xi_j^i)^2 + (y_j^i)^2]^2}} \ . \tag{6.4}$$

Now this is only part of the answer because in this analysis we held ξ_j^i constant, whereas in reality it is also a function of the population eigenvalues. Thus, the partial derivative of y_j^i with respect to τ_k is given by the formula

$$\frac{\partial y_j^i}{\partial \tau_k} = \frac{\partial \widetilde{y}_j^i}{\partial \tau_k} + \frac{\partial y_j^i}{\partial \xi_j^i} \cdot \frac{\partial \xi_j^i}{\partial \tau_k} \ , \tag{6.5}$$

where $\partial \widetilde{y}_j^i / \partial \tau_k$ is given by equation (6.4) and $\partial \xi_j^i / \partial \tau_k$ is given by equation (5.2). All that remains to be computed is $\partial y_j^i / \partial \xi_j^i$. This is done by temporarily ignoring direct dependency on population eigenvalues and setting up the function

$$\widehat{\Gamma}(y; \xi) := \frac{1}{p}\sum_{k=1}^{p} \frac{\tau_k^2}{\left(\tau_j - \xi\right)^2 + y^2} - \frac{1}{c} \ .$$

Differentiating the equation $\widehat{\Gamma}(y; \xi) = 0$ yields:

$$\frac{\partial \widehat{\Gamma}}{\partial y}\, dy + \frac{\partial \widehat{\Gamma}}{\partial \xi}\, d\xi = 0 \quad \Longrightarrow \quad \frac{\partial y}{\partial \xi} = -\frac{\dfrac{\partial \widehat{\Gamma}}{\partial \xi}}{\dfrac{\partial \widehat{\Gamma}}{\partial y}} \ .$$

The partial derivatives of the function $\widehat{\Gamma}$ are

$$\frac{\partial\widehat{\Gamma}}{\partial\xi}(y;\xi) = 2\sum_{l=1}^{p}\frac{\tau_l^2(\tau_l-\xi)}{\left[(\tau_l-\xi)^2+y^2\right]^2}$$
$$\text{and}\qquad \frac{\partial\widehat{\Gamma}}{\partial y}(y;\xi) = -2\sum_{l=1}^{p}\frac{\tau_l^2\, y}{\left[(\tau_l-\xi)^2+y^2\right]^2}\ ;$$

therefore,

$$\frac{\partial y_j^i}{\partial\xi_j^i} = \frac{\displaystyle\sum_{l=1}^{p}\frac{\tau_l^2(\tau_l-\xi_j^i)}{\left[\left(\tau_l-\xi_j^i\right)^2+\left(y_j^i\right)^2\right]^2}}{\displaystyle\sum_{l=1}^{p}\frac{\tau_l^2\, y_j^i}{\left[\left(\tau_l-\xi_j^i\right)^2+\left(y_j^i\right)^2\right]^2}}\ .$$

Plugging this formula into equation (6.5) yields the partial derivative of y_j^i with respect to τ_k.

7 Density of the Limiting Distribution of the Sample Eigenvalues

7.1 Mapping

This is the operation where we leave u-space and map back to $(x, F(x))$ where F is the limiting distribution of sample eigenvalues. The underlying mathematics for this mapping can be found in equations (2.7)–(2.8) of Ledoit and Wolf (2012). The mapping can be expressed with the notation of the present paper as

$$x := u - c\,u\,\frac{1}{p}\sum_{k=1}^{p}\frac{\tau_k}{\tau_k-u}\ .$$

In the remainder of this section, we will assume that the interval index $i \in \{1,\ldots,\nu\}$ is fixed. For every $j \in \{0,1,\ldots,\omega_i+1\}$, map z_j^i into

$$x_j^i = z_j^i - c\,z_j^i\,\frac{1}{p}\sum_{k=1}^{p}\frac{\tau_k}{\tau_k-z_j^i}\ . \tag{7.1}$$

Even though z_j^i is generally a complex number, equation (6.1) guarantees that x_j^i is real.

Using Section (2.3) of Ledoit and Wolf (2012), we can also obtain the value of the

limiting sample spectral density F' evaluated at x_j^i as $F'(x_j^i) = f_j^i$ where

$$f_j^i = \frac{1}{c\pi} \mathsf{Im}\left[-\frac{1}{z_j^i}\right] = \frac{1}{c\pi} \frac{y_j^i}{(x_j^i)^2 + (y_j^i)^2} \ . \tag{7.2}$$

Note that $f_1^i = f_{\omega_i+1}^i = 0$.

The output of this operation is $(x_j^i, f_j^i)_{j=0,1,\ldots,\omega_i+1}$, for every $i \in \{1, \ldots, \nu\}$.

7.2 Derivative

From equation (7.2), it is easy to compute the partial derivative of f_j^i with respect to τ_k as

$$\frac{\partial f_j^i}{\partial \tau_k} = \frac{1}{c\pi} \, \mathsf{Im}\left[\frac{\partial z_j^i}{\partial \tau_k} \cdot \frac{1}{\left(z_j^i\right)^2}\right] \ , \tag{7.3}$$

$$\text{where} \qquad \frac{\partial z_j^i}{\partial \tau_k} = \frac{\partial x_j^i}{\partial \tau_k} + \sqrt{-1} \, \frac{\partial y_j^i}{\partial \tau_k} \ , \tag{7.4}$$

$\partial x_j^i / \partial \tau_k$ is given by equation (5.2), and $\partial y_j^i / \partial \tau_k$ is given by equation (6.5).

In order to differentiate equation (7.1) more easily, introduce the function m_{LH} defined as per Section 2.2 of Ledoit and Wolf (2012):

$$\forall z \in \mathbb{C}^+ \qquad m_{LH}(z; \tau_1, \ldots, \tau_p) \coloneqq \frac{1}{p} \sum_{l=1}^{p} \frac{\tau_l}{\tau_l - z} = 1 + z \, \frac{1}{p} \sum_{l=1}^{p} \frac{1}{\tau_l - z} \ .$$

This enables us to rewrite equation (7.1) as

$$x_j^i = z_j^i - c \, z_j^i \, m_{LH}(z_j^i; \tau_1, \ldots, \tau_p) \ . \tag{7.5}$$

The full derivative of $m_{LH}(z_j^i; \tau_1, \ldots, \tau_p)$ with respect to τ_k is

$$\frac{d m_{LH}}{d \tau_k} = \frac{\partial m_{LH}}{\partial \tau_k} + \frac{\partial m_{LH}}{\partial z_j^i} \cdot \frac{\partial z_j^i}{\partial \tau_k} \ ,$$

where the last term is given by equation (7.4). The partial derivatives of m_{LH} are

$$\frac{\partial m_{LH}}{\partial \tau_k} = -z_j^i \times \frac{1}{p} \frac{1}{(\tau_k - z_j^i)^2}$$

$$\text{and} \qquad \frac{\partial m_{LH}}{\partial z_j^i} = \frac{1}{p} \sum_{l=1}^{p} \frac{\tau_l}{(\tau_l - z_j^i)^2} \ ;$$

therefore,

$$\frac{dm_{LH}}{d\tau_k} = -z_j^i \times \frac{1}{p}\frac{1}{(\tau_k - z_j^i)^2} + \frac{\partial z_j^i}{\partial \tau_k} \times \frac{1}{p}\sum_{l=1}^{p}\frac{\tau_l}{(\tau_l - z_j^i)^2} \ .$$

Finally, differentiating equation (7.5) enables us to compute the partial derivative of x_j^i with respect to τ_k as follows:

$$\frac{\partial x_j^i}{\partial \tau_k} = \frac{\partial z_j^i}{\partial \tau_k} \times \left[1 - c \cdot m_{LH}(z_j^i; \tau_1, \ldots, \tau_p)\right] - c \cdot z_j^i \, \frac{dm_{LH}}{d\tau_k}(z_j^i; \tau_1, \ldots, \tau_p) \ . \tag{7.6}$$

8 Cumulative Distribution Function

8.1 Numerical Integration of the Density

The objective is to compute $F_j^i \coloneqq F(x_j^i)$. We know that

$$F(0) = F_0^1 = \max\left(0, 1 - \frac{1}{c}\right) \ . \tag{8.1}$$

Since the support of F is $\cup_{i=1}^{\nu}[x_0^i, x_{\omega_i+1}^i]$ (with the possible addition of $\{0\}$ if $p > n$), as soon as ν is greater than or equal to two, $F_0^{i+1} = F_{\omega_i+1}^i$, for $i = 1, \ldots, \nu - 1$. Bai and Silverstein (1999) show that

$$\forall i = 1, \ldots, \nu \qquad F_{\omega_i+1}^i = \frac{1}{p}\sum_{j=1}^{i}\omega_j \ . \tag{8.2}$$

All that remains is to compute F_j^i for $j \in \{1, \ldots, \omega_i\}$. First, we will get an approximation of F_j^i by using the trapezoidal integration formula over $[x_0^i, x_j^i]$. Then we will refine this approximation using the fact stated in equation (8.2). The trapezoidal method yields the approximation:

$$\forall j = 1, \ldots, \omega_i + 1 \qquad \widetilde{F}_j^i \coloneqq F_0^i + \frac{1}{2}\sum_{l=1}^{j}(x_l^i - x_{l-1}^i)(f_l^i + f_{l-1}^i) \ . \tag{8.3}$$

Now the problem is that $\widetilde{F}_{\omega_i+1}^i$ thus defined would generally differ from $\sum_{j=1}^{i}\omega_j/p$ due to numerical error in the integration formula. This is why, in a second step, we refine the approximation by computing

$$F_j^i \coloneqq F_0^i + \left(\widetilde{F}_j^i - F_0^i\right)\frac{F_{\omega_i+1}^i - F_0^i}{\widetilde{F}_{\omega_i+1}^i - F_0^i} \qquad \text{for} \quad j = 1, \ldots, \omega_i \ . \tag{8.4}$$

8.2 Derivatives with Respect to Population Eigenvalues

The computation of these derivatives is subdivided into two steps that mirror the ones performed in Section 8.1. First, by differentiating equation (8.3) with respect to x_l^i and f_l^i we obtain:

$$\forall j = 1, \ldots, \omega_i + 1 \qquad \frac{\partial \widetilde{F}_j^i}{\partial \tau_k} = \frac{1}{2} \sum_{l=1}^{j} \left(\frac{\partial x_l^i}{\partial \tau_k} - \frac{\partial x_{l-1}^i}{\partial \tau_k} \right) \left(f_l^i + f_{l-1}^i \right) + \frac{1}{2} \sum_{l=1}^{j} \left(x_l^i - x_{l-1}^i \right) \left(\frac{\partial f_l^i}{\partial \tau_k} + \frac{\partial f_{l-1}^i}{\partial \tau_k} \right), \tag{8.5}$$

where the partial derivatives of x_l^i and f_l^i with respect to τ_k are given by equations (7.6) and (7.3), respectively. Second, differentiating equation (8.4) with respect to $\widetilde{F}_j^i$ and $\widetilde{F}_{\omega_i+1}^i$ yields:

$$\frac{\partial F_j^i}{\partial \tau_k} = \left(F_{\omega_i+1}^i - F_0^i \right) \frac{\dfrac{\partial \widetilde{F}_j^i}{\partial \tau_k} - F_0^i}{\widetilde{F}_{\omega_i+1}^i - F_0^i} - \left(F_{\omega_i+1}^i - F_0^i \right) \frac{\partial \widetilde{F}_{\omega_i+1}^i}{\partial \tau_k} \cdot \frac{\widetilde{F}_j^i - F_0^i}{\left(\widetilde{F}_{\omega_i+1}^i - F_0^i \right)^2}, \tag{8.6}$$

where the partial derivatives of $\widetilde{F}_j^i$ and $\widetilde{F}_{\omega_i+1}^i$ with respect to τ_k are given by equation (8.5).

9 Discretization of the Sample Spectral C.D.F.

9.1 Sample Eigenvalues

The final operation involves extracting from F a set of p sample eigenvalues $(\lambda_1, \ldots, \lambda_p)$. First, we take care of zero eigenvalues when $c > 1$. By equation (8.1) we know that

$$\text{if} \quad p < n \quad \text{then} \quad \lambda_1, \ldots, \lambda_{p-n} = 0 \; .$$

In what follows we will assume that we have fixed an interval index i in the set $\{1, \ldots, \nu\}$.

Let the function $X^i(\alpha)$ denote the approximation to $\int_{F_0^i}^{\alpha} F^{-1}(x)dx$ that is obtained by fitting a piecewise linear function to $F^{-1}(\cdot)$ over the interval $[F_0^i, F_{\omega_i+1}^i]$. This piecewise linear function passes through every point $(F_j^i, x_j^i)_{j=0,\ldots,\omega_i+1}$. Using once again the

trapezoidal integration formula, we get:

$$\forall j = 0, \ldots, \omega_i \qquad \int_{F_j^i}^{F_{j+1}^i} F^{-1}(x)\, dx \approx X^i(F_{j+1}^i) - X^i(F_j^i) = \left(F_{j+1}^i - F_j^i\right) \frac{x_j^i + x_{j+1}^i}{2} \ . \tag{9.1}$$

For every integer κ such that $pF_0^i \leq \kappa < pF_{\omega_i+1}^i$, define $j(\kappa)$ as the unique integer in $\{0, \ldots, \omega_i\}$ such that $F_{j(\kappa)}^i \leq \kappa < F_{j(\kappa)+1}^i$. Then we have:

$$\int_{F_{j(\kappa)}^i}^{\kappa/p} F^{-1}(x)\, dx \approx X^i(\kappa/p) - X^i\left(F_{j(\kappa)}^i\right), \quad \text{where}$$

$$\begin{aligned} X^i(\kappa/p) - X^i\left(F_{j(\kappa)}^i\right) &= \left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right) \left[x_{j(\kappa)}^i + \frac{\frac{\kappa}{p} - F_{j(\kappa)}^i}{2\left(F_{j(\kappa)+1}^i - F_{j(\kappa)}^i\right)} \left(x_{j(\kappa)+1}^i - x_{j(\kappa)}^i\right)\right] \\ &= \left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right) x_{j(\kappa)}^i + \frac{\left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right)^2}{2\left(F_{j(\kappa)+1}^i - F_{j(\kappa)}^i\right)} \left(x_{j(\kappa)+1}^i - x_{j(\kappa)}^i\right) \ . \end{aligned} \tag{9.2}$$

Putting together equations (9.1)–(9.2) yields:

$$\begin{aligned} X^i(\kappa/p) = &\sum_{l=0}^{j(\kappa)-1} \left(F_{l+1}^i - F_l^i\right) \frac{x_l^i + x_{l+1}^i}{2} \\ &+ \left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right) x_{j(\kappa)}^i + \frac{\left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right)^2}{2\left(F_{j(\kappa)+1}^i - F_{j(\kappa)}^i\right)} \left(x_{j(\kappa)+1}^i - x_{j(\kappa)}^i\right) \ . \end{aligned}$$

Finally, we can define the sample eigenvalues that belong to the i^{th} support interval as:

$$\forall \kappa \in \{pF_0^i + 1, pF_0^i + 2, \ldots, pF_{\omega_i+1}^i\} \qquad \lambda_\kappa := X^i\left(\frac{\kappa}{p}\right) - X^i\left(\frac{\kappa - 1}{p}\right) \ . \tag{9.3}$$

9.2 Partial Derivatives of Sample Eigenvalues w.r.t. Population Eigenvalues

As in Section 9.1, we handle separately the zero eigenvalues when the sample covariance matrix is singular:

$$\text{if} \quad p < n \quad \text{then} \quad \forall \kappa = 1, \ldots, p - n \qquad \frac{\partial \lambda_\kappa}{\partial \tau_k} = 0.$$

In the remainder of this section we will assume that we have fixed an interval index i in the set $\{1, \ldots, \nu\}$. Differentiating equation (9.1) with respect to F_j^i and x_j^i yields:

$$\forall j = 0, \ldots, \omega_i \qquad \frac{\partial X^i}{\partial \tau_k}(F_{j+1}^i) - \frac{\partial X^i}{\partial \tau_k}(F_j^i) = \frac{1}{2}\left(\frac{\partial F_{j+1}^i}{\partial \tau_k} - \frac{\partial F_j^i}{\partial \tau_k}\right)\left(x_j^i + x_{j+1}^i\right) + \frac{1}{2}\left(F_{j+1}^i - F_j^i\right)\left(\frac{\partial x_j^i}{\partial \tau_k} + \frac{\partial x_{j+1}^i}{\partial \tau_k}\right), \tag{9.4}$$

where the partial derivatives of F_j^i and x_j^i with respect to τ_k are given by equations (8.6) and (7.6), respectively. Similarly, differentiating equation (9.2) yields:

$$\begin{aligned}\frac{\partial X^i}{\partial \tau_k}\left(\frac{\kappa}{p}\right) - \frac{\partial X^i}{\partial \tau_k}\left(F_{j(\kappa)}^i\right) = {} & \left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right)\frac{\partial x_{j(\kappa)}^i}{\partial \tau_k} - \frac{\partial F_{j(\kappa)}^i}{\partial \tau_k}\, x_{j(\kappa)}^i \\ & - \frac{\partial F_{j(\kappa)}^i}{\partial \tau_k} \times \frac{\left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right)^2}{F_{j(\kappa)+1}^i - F_{j(\kappa)}^i}\left(x_{j(\kappa)+1}^i - x_{j(\kappa)}^i\right) \\ & + \frac{\left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right)^2}{2\left(F_{j(\kappa)+1}^i - F_{j(\kappa)}^i\right)}\left(\frac{\partial x_{j(\kappa)+1}^i}{\partial \tau_k} - \frac{\partial x_{j(\kappa)}^i}{\partial \tau_k}\right) \\ & - \left(\frac{\partial F_{j(\kappa)+1}^i}{\partial \tau_k} - \frac{\partial F_{j(\kappa)}^i}{\partial \tau_k}\right)\frac{\left(\frac{\kappa}{p} - F_{j(\kappa)}^i\right)^2}{2\left(F_{j(\kappa)+1}^i - F_{j(\kappa)}^i\right)^2}\left(x_{j(\kappa)+1}^i - x_{j(\kappa)}^i\right).\end{aligned} \tag{9.5}$$

We obtain the partial derivative of X^i with respect to τ_k evaluated at κ/p from equations (9.4)–(9.5) in the following way:

$$\frac{\partial X^i}{\partial \tau_k}\left(\frac{\kappa}{p}\right) = \sum_{l=0}^{j(\kappa)-1}\left[\frac{\partial X^i}{\partial \tau_k}(F_{l+1}^i) - \frac{\partial X^i}{\partial \tau_k}(F_l^i)\right] + \left[\frac{\partial X^i}{\partial \tau_k}\left(\frac{\kappa}{p}\right) - \frac{\partial X^i}{\partial \tau_k}\left(F_{j(\kappa)}^i\right)\right],$$

which enables us to compute the partial derivatives of the sample eigenvalues that belong to the i^{th} support interval with respect to the population eigenvalues as:

$$\forall \kappa \in \{pF_0^i + 1, pF_0^i + 2, \ldots, pF_{\omega_i+1}^i\} \quad \forall k = 1, \ldots, p \quad \frac{\partial \lambda_\kappa}{\partial \tau_k} = \frac{\partial X^i}{\partial \tau_k}\left(\frac{\kappa}{p}\right) - \frac{\partial X^i}{\partial \tau_k}\left(\frac{\kappa - 1}{p}\right). \tag{9.6}$$

This derivation concludes the description of the numerical implementation of the QuEST function and its analytical Jacobian.

10 Monte Carlo Simulations

Section 5.1.1 of Ledoit and Wolf (2015) already provides some preliminary evidence documenting the accuracy of the estimator of the population eigenvalues obtained by numerically inverting the QuEST function. The simulations presented below are more extensive. They highlight the convergence rate in log-log scale for various shapes of the population spectrum.

10.1 Population Spectrum

The population eigenvalues are taken from the distribution of $1 + (\kappa - 1)X$, where κ is the condition number and X is a random variable whose support is the compact interval $[0, 1]$. Throughout the whole simulation study, we carry four different shapes for the distribution of X.

1. The original shape is left-skewed: it is the Kumaraswamy (1980) distribution with parameters $(3, 1/3)$. The Kumaraswamy family is similar in spirit to the Beta family, but more tractable: the density, the c.d.f. and the quantile function are all available in closed form. For reference, the c.d.f. of Kumaraswamy$(3, 1/3)$ is

$$\forall x \in [0,1] \qquad H_1(x) = 1 - \left(1 - x^3\right)^{1/3} \ . \tag{10.1}$$

   All the other shapes are derived from this one.

2. The next shape is right-skewed, obtained by taking the mirror image of the density about the midpoint of the support. Jones (2009, p. 73) observes that there is "a pleasing symmetry" in this case: it is equivalent to taking the mirror image of the c.d.f. about the 45 degrees line, that is, replacing it with its inverse, the quantile function:

$$\forall x \in [0,1] \qquad H_2(x) = \left[1 - (1 - x)^3\right]^{1/3} \ . \tag{10.2}$$

3. A symmetric bimodal distribution is generated by combining right-skewness on $[0, 1/2]$ with left-skewness on $[1/2, 1]$:

$$\forall x \in [0,1] \qquad H_3(x) = \begin{cases} \dfrac{1}{2}\left[1 - (1 - 2x)^3\right]^{1/3} & \text{if } x \in [0, 1/2] \ , \\ 1 - \dfrac{\left[1 - (2x - 1)^3\right]^{1/3}}{2} & \text{if } x \in [1/2, 1] \ . \end{cases} \tag{10.3}$$

4. Finally a symmetric unimodal distribution is generated by combining left-skewness on $[0, 1/2]$ with right-skewness on $[1/2, 1]$:

$$\forall x \in [0,1] \qquad H_4(x) = \begin{cases} \dfrac{1 - [1 - (2x)^3]^{1/3}}{2} & \text{if } x \in [0, 1/2] \ , \\ \dfrac{1 + [1 - (2 - 2x)^3]^{1/3}}{2} & \text{if } x \in [1/2, 1] \ . \end{cases} \tag{10.4}$$

All four densities diverge to infinity, so the set of shapes chosen is a challenging one.

10.2 Intuition

Given the sample eigenvalues $\lambda_{n,1} \leq \lambda_{n,2} \leq \cdots \leq \lambda_{n,p}$, we estimate the population eigenvalues $\tau_{n,1} \leq \tau_{n,2} \leq \cdots \leq \tau_{n,p}$ by numerically inverting the QuEST function:

$$\widehat{\boldsymbol{\tau}}_n := \underset{\mathbf{t} \in [0,\infty)^p}{\operatorname{argmin}} \ \frac{1}{p} \sum_{i=1}^{p} \left[q_{n,p}^i(\mathbf{t}) - \lambda_{n,i} \right]^2 \ . \tag{10.5}$$

The simulation study presented below centers on the base-case scenario where the condition number is $\kappa = 10$, variates are normally distributed, and the concentration ratio is $c = 1/3$. For dimension $p = 1,000$, Figure 2 provides a side-by-side comparison of the population spectra specified in Section 10.1 with their sample counterparts.

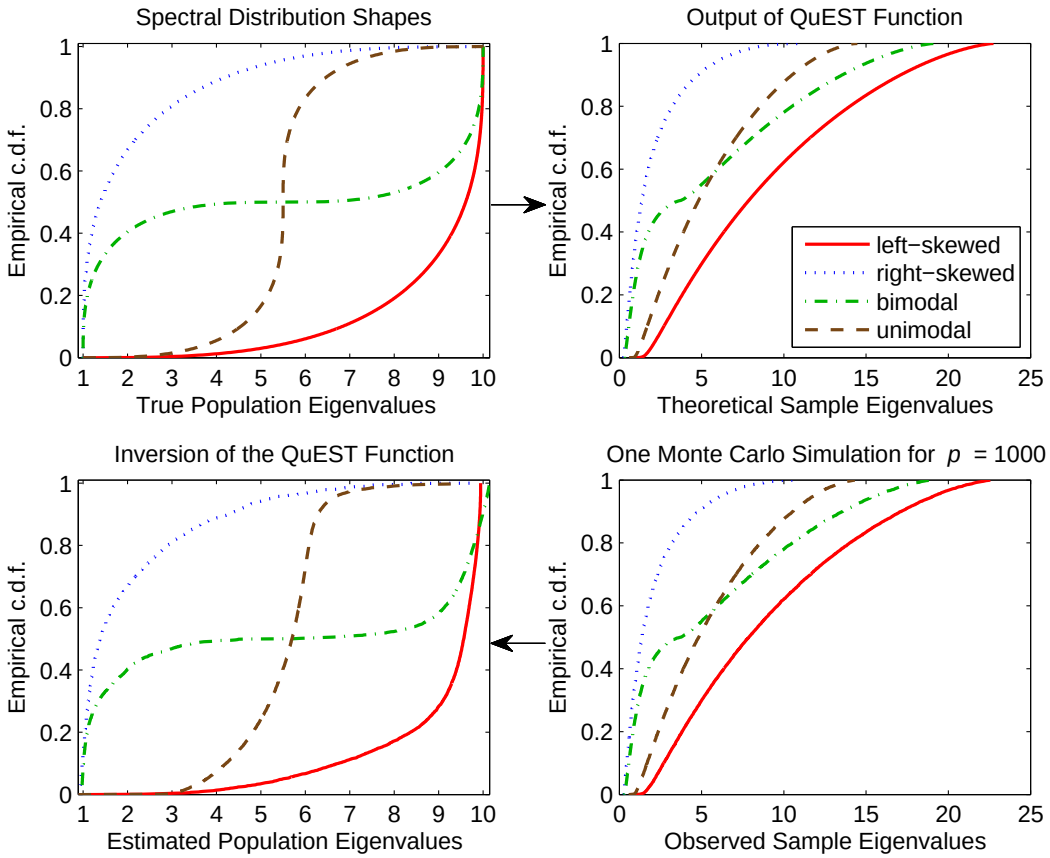

Figure 2: Population vs. sample spectrum. Top panel is the direct QuEST function, bottom panel the inverse of the QuEST function for estimation purposes.

Let us start with the top panel. It shows the effect of the QuEST function. For each of the four distribution shapes, the population eigenvalues in the top left graph get mapped into the limiting sample spectra shown in the top right graph. This illustrates the transformation coming out of the Fundamental Equation. There is a lot of distortion, but the relative positions of the four color-coded c.d.f.'s have been preserved. Therefore, the information has not been destroyed: it is just waiting to be deciphered by a suitable method. We are essentially facing a severe nonlinear bias-correction problem.

The bottom panel goes in the opposite direction: the QuEST function gets inverted. At the bottom right are sample eigenvalues generated in one Monte Carlo simulation. Observe how closely they match the nonrandom distributions in the top right. This is because, as mentioned above, in the large-dimensional asymptotic limit randomness vanishes. Then numerically inverting the QuEST function yields the estimator of population eigenvalues shown in the bottom left graph. It closely matches the truth (shown top left). The distortion has been undone, and the original shapes of the spectral

distributions have been restored. The bottom panel is our estimation procedure in a nutshell.

10.3 Base-Case Scenario

Ledoit and Wolf (2015, Theorem 2.2) prove that the mean squared deviation between estimated and true population eigenvalues $p^{-1}\sum_{i=1}^{p}\left[\widehat{\tau}_{n,i}-\tau_{n,i}\right]^2$ converges almost surely to zero under large-dimensional asymptotics. This quantity is scale-sensitive, whereas the problem is scale-invariant. This is why we study in Monte Carlo simulations the scale-adjusted quantity

$$\frac{\frac{1}{p}\sum_{i=1}^{p}\left[\widehat{\tau}_{n,i}-\tau_{n,i}\right]^2}{\left(\frac{1}{p}\sum_{i=1}^{p}\tau_{n,i}\right)^2} \tag{10.6}$$

instead, called the (empirical) *normalized mean squared error*. This change in performance measure does not make any difference to strong the consistency result, given that Ledoit and Wolf (2015) assume that the population eigenvalues are bounded away from zero and infinity. But we do not want to give the visual impression that covariance matrices with a larger trace are estimated less accurately, since on a relative basis it is not true.

The matrix dimension ranges from $p = 30$ to $p = 1,000$. Convergence of the scale-adjusted mean squared deviation defined by equation (10.6) is displayed in Figure 3 on a log-log scale for the four distribution shapes.

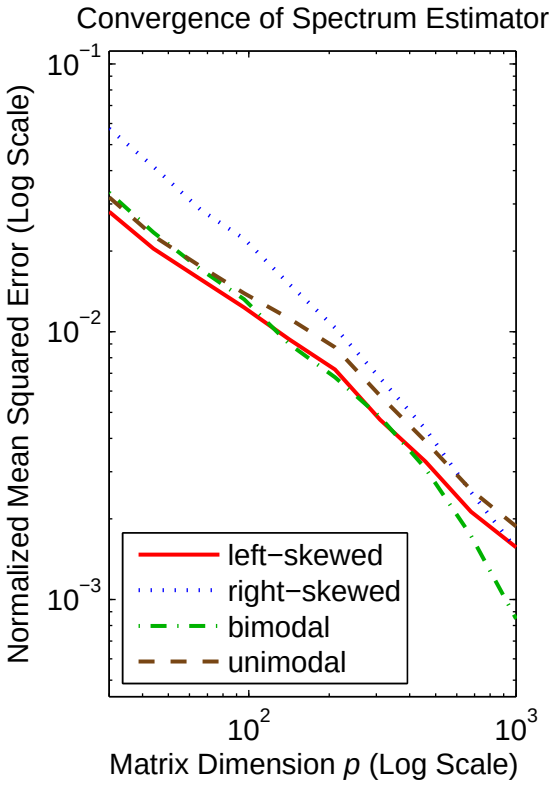

Figure 3: Consistency of the estimator of population eigenvalues in the base case scenario.

In all log-log graphs presented in this paper, including this one, the scales of the x- and y-axes have been equalized, so that the $-45°$ line corresponds to a convergence rate of p. Each point in the curves corresponds to the average across $1,000$ Monte Carlo simulations.

In terms of speed, Figure 4 shows that the numerical recipe presented in this paper for the implementation of the QuEST function is sufficiently fast for practical purposes.[2]

[2]These numbers were run using Matlab R2014b on an Apple Mac Pro with a 3.5 GHz Intel Xeon E5 processor.

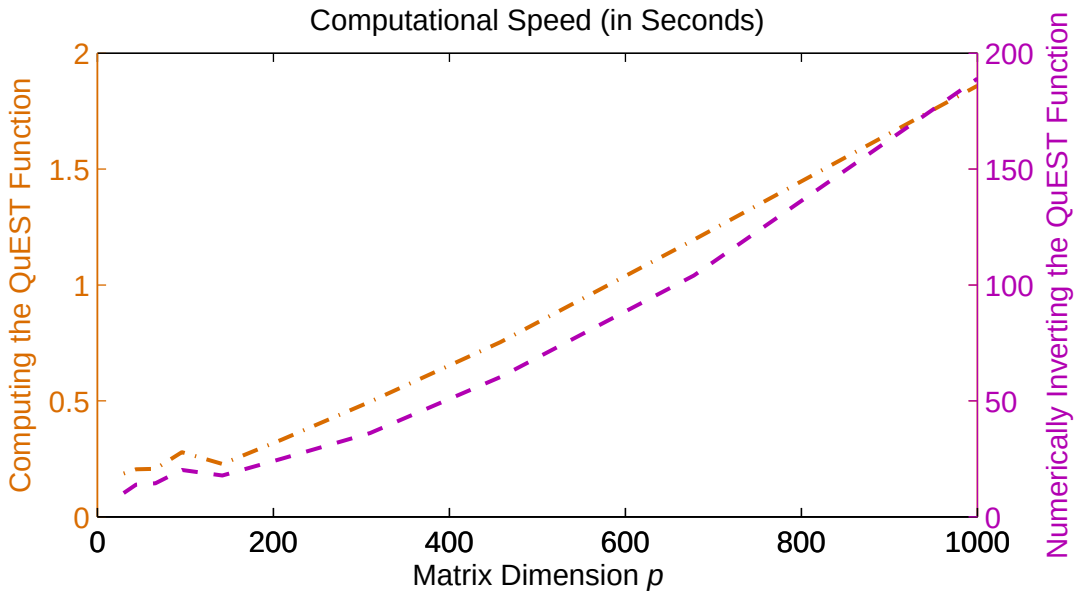

Figure 4: Speed benchmark for computing the QuEST function and estimating population eigenvalues.

The remainder of Section 10 is dedicated to demonstrating the robustness of the base–case convergence pattern in three directions: different concentration ratios $c = p/n$, condition numbers κ, and variate distributions D.

10.4 Concentration Ratio

First, we increase the concentration ratio $c = p/n$. We pick two values: $c = 1$ and $c = 2$. The first case is not covered by the mathematical theory of Ledoit and Wolf (2015), but the numerical results displayed on the left panel of Figure 5 seem to indicate that satisfactory convergence is achieved nonetheless. In the second case, we manage to consistently estimate p eigenvalues, in spite of the fact that the sample covariance matrix has only $n = p/2$ nontrivial eigenvalues. Note that this is the only graph where we let n (instead of p) range from 30 to 1,000, because of $n < p$.

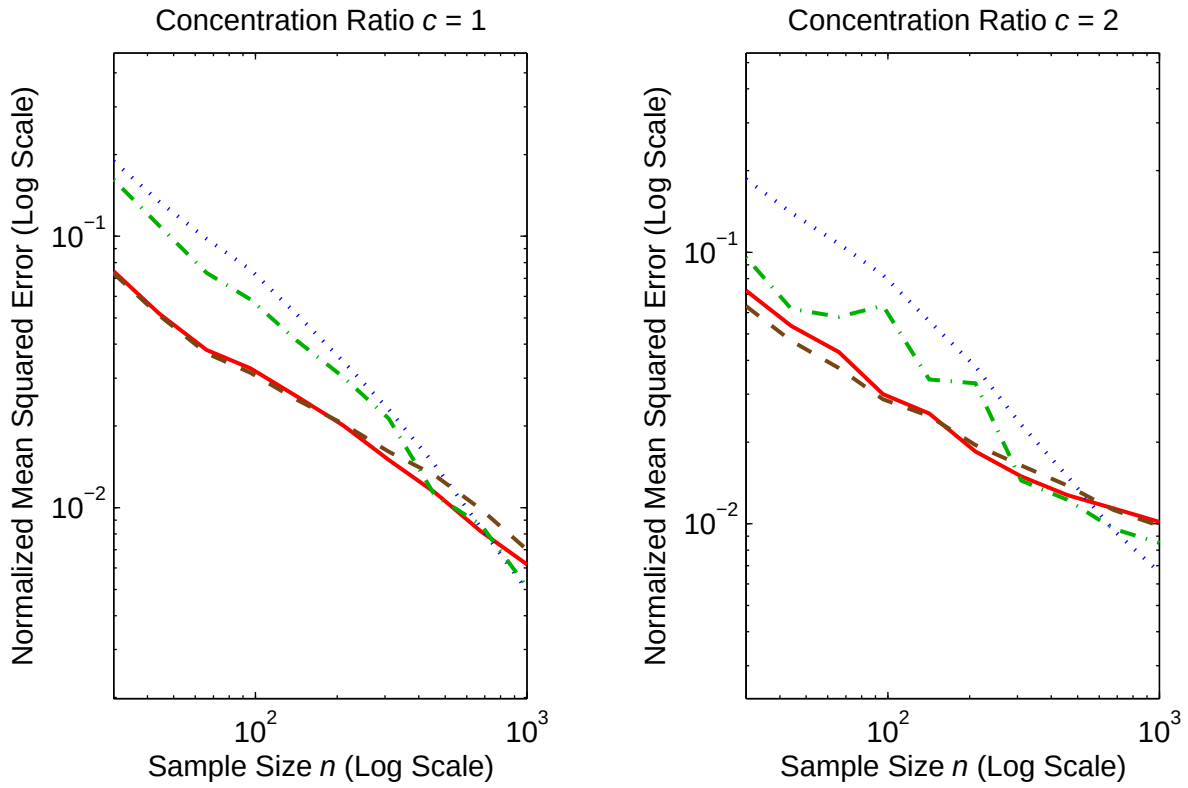

Figure 5: Consistency of the estimator of population eigenvalues for higher concentration ratios. Color and line-style code are as in Figures 2 and 3.

10.5 Condition Number

The second axis of deviation from the baseline case is to look at condition numbers other than $\kappa = 10$. We consider a smaller condition number, $\kappa = 2$, and a larger one, $\kappa = 100$. The results are displayed in Figure 6.

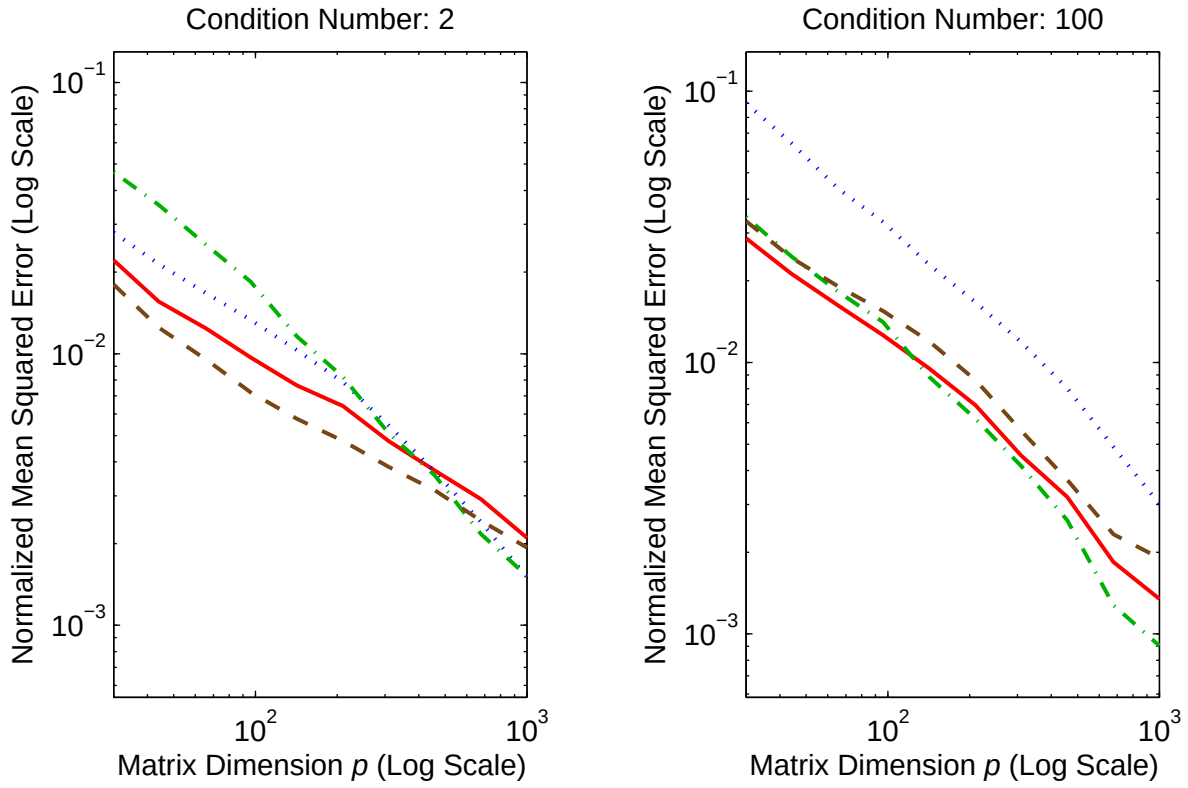

Figure 6: Consistency of the estimator of population eigenvalues for various condition numbers. Color and line-style code are as in Figures 2 and 3.

These results show that we can still obtain convergence in spite of changes in the condition number.

10.6 Distribution of the Variates

Finally, we deviate from the base-case scenario in the direction of having other distributions than Gaussian for the random variates. First, we take a fat-tailed distribution: the "Student" t-distribution with 5 degrees of freedom; and second, the most thin-tailed of all distributions: the Bernoulli coin toss distribution with probability 1/2. The results are displayed in Figure 7.

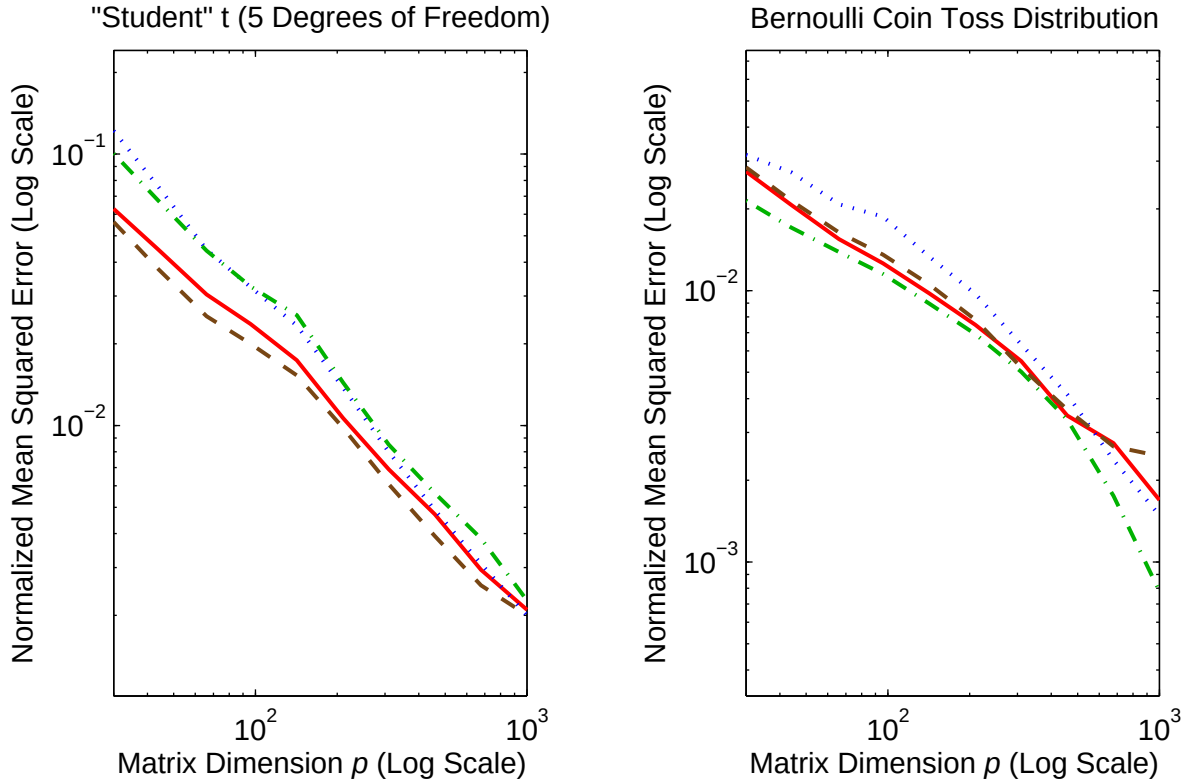

Figure 7: Consistency of the estimator of population eigenvalues when the variates have thick or thin tails. Color and line-style code are as in Figures 2 and and 3.

We also consider a skewed distribution: the exponential. The results are displayed in Figure 8.

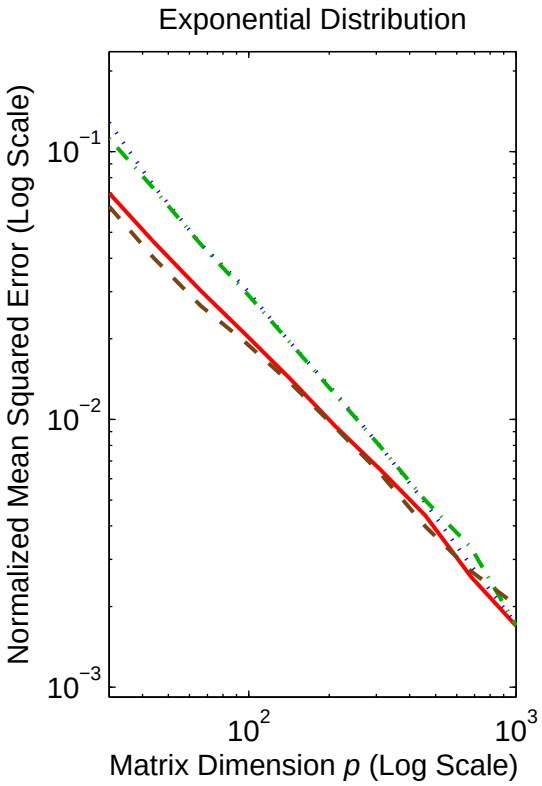

Figure 8: Consistency of the estimator of population eigenvalues when the variates are skewed. Color and line-style code are as in Figures 2 and and 3.

These results show that we can obtain convergence across a variety of variate distributions.

10.7 Overview of the Simulation Results

The Monte Carlo simulations presented above illustrate the ability of the estimator of population eigenvalues constructed by numerically inverting the QuEST function to get closer to the truth as the matrix dimension and the sample size go to infinity together. This exercise has been extensive, involving a grand total of $320,000$ Monte Carlo simulations. The point was to build practical comfort around the theoretical result. Best-fit lines in log-log space have slopes that vary in the range from -0.70 to -1.10, giving some empirical indication about the exponent of the convergence rate of the mean squared deviation between true and estimated population eigenvalues.

11 Conclusion

When the matrix dimension is not negligible with respect to the sample size, finite-dimensional asymptotic approximations are no longer close to the truth: We enter the Marčenko and Pastur (1967) zone instead. In this zone, the sample eigenvalues are a very distorted version of their population counterparts. Only after the publication of El Karoui (2008) and Mestre (2008) did researchers in the field of large-dimensional multivariate statistics start to harbor any hope of unwinding this distortion.

Ledoit and Wolf (2015) put forward a natural discretization of the Fundamental Equation of Random Matrix Theory that can be inverted numerically. Even though the sample eigenvalues are far from their population counterparts, the distortion can be inverted through this particular procedure. The present paper describes in great detail how to discretize the Marčenko-Pastur equation. We also provide extensive Monte Carlo simulations demonstrating the practical effectiveness of the method in terms of recovering the population eigenvalues. There are many applications in the field of multivariate statistics, starting with nonlinear shrinkage estimation of covariance matrices as proposed by Ledoit and Wolf (2012, 2015).

References

Anlauff, J., Weitnauer, E., Lehnhardt, A., Schirmer, S., Zehe, S., and Tonekaboni, K. (2010). A method for outdoor skateboarding video games. In *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology*, pages 40–44. ACM.

Bachega, L. R., Theiler, J., Bouman, C., et al. (2011). Evaluating and improving local hyperspectral anomaly detectors. In *Applied Imagery Pattern Recognition Workshop (AIPR), 2011 IEEE*, pages 1–8. IEEE.

Bai, Z., Chen, J., and Yao, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Australian & New Zealand Journal of Statistics*, 52(4):423–437.

Bai, Z. D. and Silverstein, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Annals of Probability*, 27(3):1536–1555.

Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large-Dimensional Random Matrices*. Springer, New York, second edition.

Bell, P. and King, S. (2009). Diagonal priors for full covariance speech recognition. In *Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on*, pages 113–117. IEEE.

Chen, J., Delyon, B., and Yao, J.-F. (2011). On a model selection problem from high-dimensional sample covariance matrices. *Journal of Multivariate Analysis*, 102(10):1388–1398.

Chen, Y., Wiesel, A., Eldar, Y. C., and Hero, A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *Signal Processing, IEEE Transactions on*, 58(10):5016–5029.

Dobriban, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices: Theory and Applications*, 04(04):1550019.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, 36(6):2757–2790.

Elsheikh, A. H., Wheeler, M. F., and Hoteit, I. (2013). An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *Journal of Computational Physics*, 242:696–714.

Guo, S.-M., He, J., Monnier, N., Sun, G., Wohland, T., and Bathe, M. (2012). Bayesian approach to the analysis of fluorescence correlation spectroscopy data II: application to simulated and in vitro data. *Analytical Chemistry*, 84(9):3880–3888.

Hafner, C. M. and Reznikova, O. (2012). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis*, 56(11):3533–3545.

Haufe, S., Treder, M. S., Gugler, M. F., Sagebaum, M., Curio, G., and Blankertz, B. (2011). EEG potentials predict upcoming emergency brakings during simulated driving. *Journal of Neural Engineering*, 8(5):056001.

Huang, N. and Fryzlewicz, P. (2015). NOVELIST estimator of large correlation and covariance matrices and their inverses. Technical report, Department of Statistics, London School of Economics and Political Science.

Ito, T. and Kubokawa, T. (2015). Linear ridge estimator of high-dimensional precision matrix using random matrix theory. Technical Report F-995, CIRJE, Faculty of Economics, University of Tokyo.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684.

Jones, M. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6(1):70–81.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1):79–88.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Annals of Statistics*, 44(3):928–953.

Ledoit, O. and Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 150(1–2):233–264.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2014a). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Working Paper ECON 137, Department of Economics, University of Zurich.

Ledoit, O. and Wolf, M. (2014b). Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Working Paper ECON 122, Department of Economics, University of Zurich.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139(2):360–384.

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

Lotte, F. and Guan, C. (2009). An efficient p300-based brain-computer interface with minimal calibration time. In *Assistive Machine Learning for People with Disabilities Symposium (NIPS'09 Symposium)*.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483.

Markon, K. (2010). Modeling psychopathology structure: a symptom-level analysis of axis I and II disorders. *Psychological Medicine*, 40(02):273–288.

Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11):5113–5129.

Michaelides, P., Apostolellis, P., and Fassois, S. (2011). Vibration-based damage diagnosis in a laboratory cable–stayed bridge model via an rcp–arx model based method. In *Journal of Physics: Conference Series*, volume 305, page 012104. IOP Publishing.

Pirkl, R. J., Remley, K., Patané, C. S. L., et al. (2012). Reverberation chamber measurement correlation. *Electromagnetic Compatibility, IEEE Transactions on*, 54(3):533–545.

Pyeon, D., Newton, M. A., Lambert, P. F., Den Boon, J. A., Sengupta, S., Marsit, C. J., Woodworth, C. D., Connor, J. P., Haugen, T. H., et al. (2007). Fundamental differences in cell cycle deregulation in human papillomavirus–positive and human papillomavirus–negative head/neck and cervical cancers. *Cancer Research*, 67(10):4605–4619.

Rao, N. R., Mingo, J. A., Speicher, R., and Edelman, A. (2008). Statistical eigen-inference from large Wishart matrices. *Annals of Statistics*, 36(6):2850–2885.

Ribes, A., Planton, S., and Terray, L. (2013). Application of regularised optimal fingerprinting to attribution. part I: method, properties and idealised analysis. *Climate Dynamics*, 41(11-12):2817–2836.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1). Article 32.

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

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

Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences*, 34(1):1373–1403.

Stieltjes, T. J. (1894). Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse* 1^{re} *Série*, 8(4):J1–J122.

Welsing, S. (2015). Nonlinear shrinkage estimation of covariance matrices for portfolio selection. Master's thesis, Department of Mathematics, Technical University of Munich.

Yao, J. (2015). Identifying the number of factors from singular values of a large sample auto-covariance matrix. Presentation given at the conference on *Complex Systems in Time Series* organized by the London School of Economics on December 5, 2015.

Yao, J., Kammoun, A., and Najim, J. (2012). Eigenvalue estimation of parameterized covariance matrices of large dimensional data. *Signal Processing, IEEE Transactions on*, 60(11):5893–5905.

Yao, J., Zheng, S., and Bai, Z. (2015). *Large Sample Covariance Matrices and High-dimensional Data Analysis*. Cambridge University Press.

Zhang, Y., Sun, D., and Zhang, D. (2009). Robust adaptive acoustic vector sensor beamforming using automated diagonal loading. *Applied Acoustics*, 70(8):1029–1033.