Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices.

Alexei Onatski

March 13, 2007

Abstract

This paper establishes the fact that the joint distribution of the centered and scaled several largest eigenvalues of a p-dimensional complex Wishart matrix $W_{\mathbb{C}}(\Omega, n)$ converge to the joint Tracy-Widom distribution when n and p tend to infinity so that n/p remains in a compact subset of $(0, \infty)$. Our result extends Baik et al. (2005) and El Karoui (2007) who study the asymptotic distribution of the single largest eigenvalue of $W_{\mathbb{C}}(\Omega, n)$ as n and p tend to infinity so that n/p remains in a compact subset of $[1, \infty)$. We show how our result can be used to find a 95% confidence set for the number of common factors in excess stock returns.

1 Introduction

The goal of this paper is to establish the joint asymptotic distribution of a finite number of properly scaled and centered largest eigenvalues of a p-dimensional complex Wishart matrix $W_{\mathbb{C}}(\Omega, n)$ as both n and p go to infinity so that n/p remains in a compact subset of $(0, \infty)$. The paper extends Baik et al. (2005) and El Karoui (2007) who find the asymptotic distribution of the scaled and centered single largest eigenvalue of $W_{\mathbb{C}}(\Omega, n)$ under an assumption that n/p remains in a compact subset of $[1, \infty)$. When n/p is less than 1, the Wishart matrix is singular. The main contribution of this paper is extending Baik et al. (2005) and El Karoui (2007) to the singular Wishart case.

A need for such an extension arise in a separate paper of mine (2007) which develops statistical tests of various hypotheses about the number of factors in Chamberlain and Rothschild's (1983) approximate factor model, which is widely used in empirical macroeconomics and finance. The model considers a double-infinite sequence of random variables $\{\xi_{it},\ i, t \in \mathbb{N}\}$ such that, for any $i, t \in \mathbb{N}$:

$$\xi_{it} = \Lambda_i' F_t + \eta_{it} \tag{1}$$

where F_t and Λ_i are k-dimensional ($k < \infty$) vectors of unobserved common factors and factor loadings, respectively, and η_{it} is an unobserved idiosyncratic component of ξ_{it}. In contrast to the classical factor model (see Anderson (1984), chapter 14) the idiosyncratic components are allowed to be correlated over the i-dimension. The identification of the idiosyncratic components is achieved by assuming that the largest eigenvalue of the covariance matrix of vector $\{\eta_{it}\}_{1 \le i \le p}$ stays bounded as p tends to infinity whereas the smallest non-zero eigenvalue of the covariance matrix of vector $\{\Lambda_i' F_t\}_{1 \le i \le p}$ diverges to infinity as fast as p. These assumptions are often interpreted as formalizing the requirements that the common factors non-trivially influence all data points whereas the idiosyncratic components have only local effect.

Researchers in macroeocnomics and finance use Chamberlain and Rothschild's (1983) model to handle high-dimensional data sets. They interpret F_t as a vector of factors non-trivially influencing hundreds of available macroeconomic indicators (see, for example, Stock and Watson (2002)) or, in the case of finance, as a vector of the risk factors common to hundreds of stock returns (see, for example, Connor and Korajczyk (1993)). An important practical question is how many such factors there are. Under assumptions of the Chamberlain-Rothschild model, one can equivalently ask how many eigenvalues of XX'/n, where X is the data matrix $\{\xi_{it}\}_{1\leq i\leq p,1\leq t\leq n}$, diverge to infinity as both p and n tend to infinity so that n/p remains in a compact subset of $(0,\infty)$.

Using the eigenvalue perturbation theory, Onatski (2007) shows that if the true number of factors is k_0, then the asymptotic distribution of the scaled and centered k_0+1th, k_0+2th, etc. eigenvalues of XX'/n is the same as that of the 1st, 2nd, etc. eigenvalues of the sample covariance matrix of the idiosyncratic components[1]. Therefore, assuming that $\{\eta_{it}\}_{1\leq i\leq p}$ are i.i.d. (over $t\in\mathbb{N}$) complex Gaussian $N_{\mathbb{C}}(0,\Sigma_p)$ vectors[2], a test (described in more detail in Section 4 below) of the null hypothesis that the true number of factors equals k_0 against an alternative of more than k_0 factors can be based on checking whether the k_0+1th, k_0+2th, etc. eigenvalues of the sample covariance matrix of the data are drawn from the joint distribution of the largest eigenvalues of $W_{\mathbb{C}}\left(\Sigma_p/n,n\right)$. Since in macroeconomics and finance the cross-sectional dimension of data p is often larger than their time series dimension n, to obtain the asymptotic critical values of the test we have to analyze the joint asymptotic distribution of the largest eigenvalues of a singular complex Wishart matrix.

El Karoui (2007) proves that the asymptotic distribution of the properly scaled and centered largest eigenvalue of a non-singular complex Wishart matrix $W_{\mathbb{C}}\left(\Sigma_p/n,n\right)$ is the Tracy-Widom distribution of type two (TW_2). TW_2 refers to a distribution with the cumulative distribution function

$$F(x)\equiv\exp\left(-\int_x^\infty(x-s)q^2(s)ds\right),$$

where $q(s)$ is the solution of an ordinary differential equation

$$q''(s)=sq(s)+2q^3(s),$$

[1] Slightly abusing a standard definition, we will call matrix $\sum_{i=1}^n v_iv_i'/n$ the sample covariance matrix of vectors $v_1,...,v_n$.

[2] To make such an assumption realistic, we perform a preliminary transformation of real-valued data into a complex-valued form. The reason to work with complex-valued data is that much more is known about the largest eigenvalues of the complex Wishart matrices than about those of the real Wishart matrices.

which is asymptotically equivalent to the Airy function $Ai(s)$ as $s \to \infty$. It plays an important role in the large random matrix theory (see Mehta, 2004) because it is the asymptotic distribution of the scaled and centered largest eigenvalue of a matrix from the so called Gaussian Unitary Ensemble (GUE) as the size of the matrix goes to infinity.

GUE is a collection of all $N \times N$ Hermitian matrices with i.i.d. complex Gaussian $N_{\mathbb{C}}(0, 1/N)$ lower triangular entries and (independent from them) i.i.d. real Gaussian $N(0, 1/N)$ diagonal entries. Let $d_1 \geq ... \geq d_N$ be eigenvalues of a matrix from GUE. Define $\tilde{d}_i = N^{2/3}(d_i - 2)$. Tracy and Widom (1994) studied the asymptotic distribution of a few largest eigenvalues of matrices from GUE when $N \to \infty$. They described the asymptotic marginal distributions of $\tilde{d}_i$, $i = 1, ..., m$, where m is any fixed positive integer, in terms of a solution of a completely integrable system of partial differential equations.[3] The system simplifies to a single ordinary differential equation given above when we are interested in the asymptotic distribution of the largest eigenvalue only.

In this paper, we extend El Karoui's (2007) results to show that the asymptotic distribution of the scaled and centered m largest eigenvalues ($m < \infty$) of a possibly singular complex Wishart matrix $W_{\mathbb{C}}(\Sigma_p/n, n)$ is the same as the joint asymptotic distribution of $\tilde{d}_1, ..., \tilde{d}_m$. We follow Soshnikov (2002) in calling such a joint distribution the joint Tracy-Widom distribution.

The large random matrix theory has developed the following powerful method of analysis of the joint asymptotic distribution of a few largest eigenvalues of various random matrices as the dimensionality of the matrices goes to infinity. First, the joint distribution of a few largest eigenvalues is expressed through the probabilities $P(i_1, ..., i_m; J_1, ..., J_m)$ that disjoint subsets $J_1, ..., J_m$ of real line contain exactly $i_1, ..., i_m$ eigenvalues. Then, the latter probabilities are represented in the form of Fredholm determinants of operators indexed by the dimensionality of the analyzed random matrix. Finally, it is proven that the operators converge in the trace-class norm as the dimensionality goes to infinity and the corresponding

[3] For a discussion of the *joint* asymptotic distribution of $\tilde{d}_i$, $i = 1, ..., m$, see Theorem 1.4 in Johansson (2001).

limits are found. The outcome of such an analysis is an expression of the joint distribution of a few largest eigenvalues through Fredholm determinants of the limiting integral operators. Often, kernels of these operators have a relatively simple form which insures further detailed analysis of the joint distribution of the largest eigenvalues.

Let λ_j be the j-th largest eigenvalue of a complex Wishart matrix $W_{\mathbb{C}}\left(\Sigma_p/n, n\right)$. Then, the first step of the above method is performed as follows. For any real $s_1 > ... > s_m > 0$,

$$\Pr\left(\lambda_1 \leq s_1, ..., \lambda_m \leq s_m\right) = \sum_I P\left(i_1, ..., i_m; \left(s_1, \infty\right), \left(s_2, s_1\right], ..., \left(s_m, s_{m-1}\right]\right), \tag{2}$$

where I consists of all sets of m non-negative integers $i_1, ..., i_m$ such that $i_1 = 0$ and $i_{j+1} \leq j - i_j - ... - i_1$ for $j = 1, ..., m-1$. In the special case when only the largest eigenvalue is analyzed, we have: $\Pr\left(\lambda_1 \leq s_1\right) = P\left(0, \left(s_1, \infty\right)\right) = E\prod_{j=1}^{p}\left[1 - \chi_{\left(s_1, \infty\right)}\left(\lambda_j\right)\right]$, where $\chi_J\left(\lambda\right)$ denotes the indicator function of the set J, and the expectation is taken with respect to the joint distribution of $\lambda_1, ..., \lambda_p$. For this special case, Baik et al. (2005) perform the second step of the above method.[4] Assuming that $n \geq p$, they show that $E\prod_{j=1}^{p}\left[1 - \chi_{\left(s_1, \infty\right)}\left(\lambda_j\right)\right]$ equals the Fredholm determinant $\det\left(1 - K_{n,p}\right)$, where $K_{n,p}$ is an operator acting on $L^2\left(\left(s_1, \infty\right)\right)$ with a kernel that has a convenient integral representation.

To establish the asymptotic distribution of the largest eigenvalue of $W_{\mathbb{C}}\left(\Sigma_p/n, n\right)$, El Karoui (2007) starts from Baik et al.'s (2005) result. He then finds centering and scaling sequences $\mu_{n,p}$ and $\sigma_{n,p}$ such that as both n and p go to infinity, the re-centered and re-scaled version of the operator $K_{n,p}$, $S_{n,p}$, converges in the trace-class norm to an operator $E \cdot Ai \cdot E^{-1}$ acting on $L^2\left(\left(s_1, \infty\right)\right)$, where E is an operator of the multiplication by a certain function and Ai is an integral operator with the Airy kernel $Ai(x, y) = \int Ai(x+u)Ai(y+u)du$, where $Ai(x)$ is the Airy function (see Olver, 1974). Since Fredholm determinant is continuous with respect to the trace-class norm and since it is invariant with respect to conjugation, El Karoui (2007) concludes that the distribution of the centered and scaled largest eigenvalue of $W_{\mathbb{C}}\left(\Sigma_p/n, n\right)$ converges to the distribution defined by $\det\left(I - Ai\right)$, which is TW_2 (see

[4] They also perform the final step of the method, but for a special form of matrix Σ_p.

Tracy and Widom (1994)).

A careful inspection of El Karoui's (2007) proofs reveals that he uses the assumption that n/p remains in a compact subset of $[1, \infty)$ only to be able to use the determinantal representation of the cumulative distribution function of the largest eigenvalue of $W_{\mathbb{C}}(\Sigma_p/n, n)$ established by Baik et al. (2005). We therefore, first, extend Baik et al. (2005) to the case of a singular complex Wishart matrix. Somewhat unexpectedly, we find that not only the determinantal representation of the cumulative distribution function of the largest eigenvalue of $W_{\mathbb{C}}(\Sigma_p/n, n)$ but also all the rest of their results hold for the singular Wishart case without any extra qualifications. Our extension of El Karoui (2007) easily follows from the extension of Baik et al. (2005).

The rest of this paper is organized as follows. In Section 2, we prove our generalization of Baik et al. (2005). Section 3 generalizes El Karoui (2007) to the case of several eigenvalues of a possibly singular complex Wishart matrix. Section 4 contains an application to the determination of the number of the common risk factors in stock return data. Section 5 concludes. The Appendix contains proofs of some of the less important statements of this paper.

2 Extension of Baik et al. (2005)

In this section, we extend Baik et al.'s (2005) analysis to the singular situation when $n < p$ and to the case of several largest eigenvalues of $W_{\mathbb{C}}(\Sigma_p/n, n)$. Note that for a general positive integer m, we have:

$$P(i_1, ..., i_m; J_1, ..., J_m) = \frac{1}{i_1!...i_m!} \frac{\partial^{i_1+...+i_m}}{\partial z_1^{i_1}...\partial z_m^{i_m}} E \prod_{j=1}^{p} \left[1 + \sum_{k=1}^{m} (z_k - 1) \chi_{J_k}(\lambda_j) \right] \Bigg|_{z_1=...=z_m=0} \tag{3}$$

(see, for example, Formula 4.1 in Tracy and Widom, 1998). Below, we establish a determinantal representation of $E \prod_{j=1}^{p} \left[1 + \sum_{k=1}^{m} (z_k - 1) \chi_{J_k}(\lambda_j) \right]$ which does not depend on whether $n < p$ or $n \geq p$.

First, for the case $n < p$, we will find a convenient expression for the joint density of the non-zero eigenvalues $\lambda_1, ..., \lambda_n$ of a singular complex Wishart matrix $W_{\mathbb{C}}(\Sigma_p/n, n)$. Let π_j be the *inverse* of the j-th largest eigenvalue of Σ_p. Define $\vec{\lambda} = (\lambda_1, ..., \lambda_n)'$, $\vec{\pi} = (\pi_1, ..., \pi_p)'$, $\Lambda = \operatorname{diag}(\lambda_1, ..., \lambda_n)$, and $\Pi = \operatorname{diag}(\pi_1, ..., \pi_p)$. As shown in Ratnarajah and Vaillanourt (2005), Formula (25), the joint density equals

$$f\left(\vec{\lambda}\right) = \text{const} \cdot V\left(\vec{\lambda}\right)^2 \prod_{j=1}^{n} \lambda_j^{p-n} \int_{Q_1 \in CV(p,n)} e^{-n \operatorname{tr}\left(\Pi Q_1 \Lambda Q_1^*\right)} \left(Q^* dQ_1\right), \tag{4}$$

where $V(\vec{\lambda}) = \prod_{1 \leq i < j \leq n} (\lambda_j - \lambda_i)$, $CV(p, n)$ denotes the complex Stiefel manifold of $p \times n$ matrices with orthonormal columns, and $(Q^* dQ_1)$ is the exterior differential form representing the uniform measure on the complex Stiefel manifold. Throughout the paper, "const" denotes possibly different constants that may depend on p, n and $\vec{\pi}$, but not on $\vec{\lambda}$.

Note that

$$\int_{Q_1 \in CV(p,n)} e^{-n \operatorname{tr}\left(\Pi Q_1 \Lambda Q_1^*\right)} \left(Q^* dQ_1\right) = \frac{\int_{R \in U(p)} e^{-n \operatorname{tr}\left(\Pi R_1 \Lambda R_1^*\right)} \left(R^* dR\right)}{\operatorname{Vol}\left\{U(p-n)\right\}}, \tag{5}$$

where R_1 is a $p \times n$ matrix of the first n columns of matrix $R \equiv [R_1, R_2]$ and $U(p)$ is the set of all $p \times p$ unitary matrices. A proof of formula (5) can be found in the Appendix. Now, since the unitary group is compact, we have:

$$\int_{R \in U(p)} e^{-n \operatorname{tr}\left(\Pi R_1 \Lambda R_1^*\right)} \left(R^* dR\right) = \lim_{\varepsilon \to 0} \int_{R \in U(p)} e^{-n \operatorname{tr}\left(\Pi R \Lambda_\varepsilon R^*\right)} \left(R^* dR\right) \tag{6}$$

where $\Lambda_\varepsilon := \operatorname{diag}(\lambda_1, ..., \lambda_n, \varepsilon, 2\varepsilon, ..., \varepsilon(p-n))$.

The integral on the right hand side of (6) is called the Harish-Chandra-Itzykson-Zuber integral (see Mehta, 2004, Appendix A5 and p. 648). It can be simplified as follows. Define $\vec{\lambda}_\varepsilon = (\lambda_1, ..., \lambda_n, \varepsilon, 2\varepsilon, ..., \varepsilon(p-n))'$ and let $\lambda_{\varepsilon k}$ be the k-th component of vector $\vec{\lambda}_\varepsilon$. Then we have:

$$\int_{R\in U(p)} e^{-n\,\mathrm{tr}(\Pi R\Lambda_\varepsilon R^*)}\,(R^*dR) = \mathrm{const}\cdot\left(V(\vec{\pi})V(\vec{\lambda}_\varepsilon)\right)^{-1}\det\left(e^{-n\pi_j\lambda_{\varepsilon k}}\right)_{1\le j,k\le p}. \tag{7}$$

Here, we assume that all π_i are different. If some of π_i are equal, then the formula should be changed according to l'Hospital's theorem.

Let $\{i_1,...,i_p\}$ be a set of indices equal to the set $\{1,2,...,p\}$ and such that $i_1<...<i_n$ and $i_{n+1}<...<i_p$. Denote the multi-index $(i_1,...,i_n)$ as α and the multi-index $(i_{n+1},...,i_p)$ as $\bar{\alpha}$, and let x_α denote $(x_{i_1},...,x_{i_n})'$, $x_{\alpha(k)}$ denote x_{i_k}, $x_{\bar{\alpha}}$ denote $\left(x_{i_{n+1}},...,x_{i_p}\right)'$, and $x_{\bar{\alpha}(k)}$ denote $x_{i_{n+k}}$. Finally, let $|\alpha|$ denote $i_1+...+i_n$. Then, by the Laplace expansion theorem, $\det\left(e^{-n\pi_j\lambda_{\varepsilon k}}\right)_{1\le j,k\le p}$ is equal by absolute value to

$$\sum_\alpha(-1)^{|\alpha|}\det\left(e^{-n\pi_{\alpha(j)}\lambda_k}\right)_{1\le j,k\le n}\det\left(\left(e^{-n\pi_{\bar{\alpha}(j)}\varepsilon}\right)^k\right)_{1\le j,k\le p-n}. \tag{8}$$

The second determinant in the above sum is a Vandermonde determinant. Hence,

$$\det\left(\left(e^{-n\pi_{\bar{\alpha}(j)}\varepsilon}\right)^k\right)_{1\le j,k\le p-n} = e^{-n\varepsilon\sum_{j=1}^{p-n}\pi_{\bar{\alpha}(j)}}\prod_{1\le j<k\le p-n}\left(e^{-n\pi_{\bar{\alpha}(k)}\varepsilon}-e^{-n\pi_{\bar{\alpha}(j)}\varepsilon}\right) \tag{9}$$

Further, note that

$$V(\vec{\lambda}_\varepsilon) = \mathrm{const}\cdot\varepsilon^{\binom{p-n}{2}}V(\vec{\lambda})\prod_{i=1}^{n}\prod_{k=n+1}^{p}\left(\varepsilon(k-n)-\lambda_i\right). \tag{10}$$

Combining Formulas (5) through (10) and taking limit as $\varepsilon\to 0$, we obtain:

$$\begin{aligned}&\int_{Q_1\in CV(p,n)} e^{-n\,\mathrm{tr}\left(\Pi Q_1\Lambda Q_1^*\right)}\,(Q^*dQ_1) \\ =\ &\mathrm{const}\cdot\left(V(\vec{\pi})V(\vec{\lambda})\right)^{-1}\prod_{i=1}^{n}\lambda_i^{n-p}\sum_\alpha(-1)^{|\alpha|}V\left(\pi_{\bar{\alpha}}\right)\det\left(e^{-n\pi_{\alpha(j)}\lambda_k}\right)_{1\le j,k\le n}\end{aligned} \tag{11}$$

Using (11) in (4), we find that the joint density of the non-zero eigenvalues $\vec{\lambda}=(\lambda_1,...,\lambda_n)$

of a singular complex Wishart matrix $W_C(\Sigma_p/n, n)$ equals

$$f\left(\vec{\lambda}\right) = \mathrm{const} \cdot V\left(\vec{\lambda}\right) \sum_{\alpha} (-1)^{|\alpha|} V(\pi_{\bar{\alpha}}) \det\left(e^{-n\pi_{\alpha(j)}\lambda_k}\right)_{1\leq j,k\leq n}. \tag{12}$$

Now we are ready to generalize Proposition 2.1 of Baik et al. (2005) which establish a determinantal representation of $E\prod_{j=1}^{p}\left[1-\chi_{(s_1,\infty)}(\lambda_j)\right]$ to the case of $E\prod_{j=1}^{p}\left[1+\sum_{k=1}^{m}(z_k-1)\chi_{J_k}(\lambda_j)\right]$ and general n and p.

Proposition 1 *For any fixed q satisfying $0 < q < \min\{\pi_j\}_{j=1}^{p}$, let $K_{n,p}$ be the operator acting on $L^2((0,\infty))$ with kernel*

$$K_{n,p}(\eta,\zeta) = \frac{n}{(2\pi i)^2}\int_{\Gamma} dz \int_{\Sigma} dw e^{-\eta n(z-q)+\zeta n(w-q)} \frac{1}{w-z}\left(\frac{z}{w}\right)^n \prod_{k=1}^{p}\frac{\pi_k - w}{\pi_k - z}$$

where Σ is a simple closed contour enclosing 0 and lying in $\{w : \mathrm{Re}(w) < q\}$, and Γ is a simple closed contour enclosing $\pi_1, ..., \pi_p$ and lying in $\{z : \mathrm{Re}(z) > q\}$, both oriented counter-clockwise. Then for any real-valued measurable bounded function $f(x)$ which equals 0 for any $x \leq 0$

$$E\prod_{j=1}^{p}(1+f(\lambda_j)) = \det(1+K_{n,p}f),$$

where 1 is the identity operator and f is the operator of multiplication by function $f(\cdot)$.

Proof. Let us, first, focus on the case when $n < p$. Then the eigenvalues $\lambda_{n+1}, ..., \lambda_p$ equal zero and we have: $E\prod_{j=1}^{p}(1+f(\lambda_j)) = E\prod_{j=1}^{n}(1+f(\lambda_j))$. Using the equality $V\left(\vec{\lambda}\right) = \det\left(\lambda_k^{j-1}\right)_{1\leq j,k\leq n}$ and Formula (12) we get:

$$E\prod_{j=1}^{n}(1+f(\lambda_j)) =$$
$$\mathrm{const}\cdot\sum_{\alpha}(-1)^{|\alpha|}V(\pi_{\bar{\alpha}})\int_0^{\infty}...\int_0^{\infty}\det\left(\lambda_k^{j-1}\right)\det\left(e^{-n\pi_{\alpha(j)}\lambda_k}\right)\prod_{k=1}^{n}(1+f(\lambda_k))\,d\lambda_k$$

Using Andreief's (1883) identity $\int...\int\det(f_j(x_k))\det(g_j(x_k))\prod_k d\mu(x_k) =$

$\det\left(\int f_j(x)g_k(x)d\mu(x)\right)$, we find:

$$E\prod_{j=1}^{n}\left(1+f\left(\lambda_j\right)\right)=\text{const}\cdot\sum_{\alpha}(-1)^{|\alpha|}V\left(\pi_{\bar{\alpha}}\right)\det\left(\int_0^\infty\left(1+f\left(\lambda\right)\right)\lambda^{j-1}e^{-n\pi_{\alpha(k)}\lambda}d\lambda\right)_{1\le j,k\le n} \tag{13}$$

Now define $\nu=n-p$ (note that it is less than zero) and set

$$\phi_j\left(\lambda\right)=\begin{cases}0 & \text{if } j\le-\nu \\ \frac{n^{j+\nu}}{\Gamma(j+\nu)}\lambda^{j-1+\nu}e^{-nq\lambda} & \text{if } j>-\nu\end{cases},\ \text{for } j=1,...,p,$$
$$\Phi_k\left(\lambda\right)=e^{-n(\pi_k-q)\lambda},\ \text{for } k=1,...,p$$

for any $0<q<\min\left\{\pi_j\right\}_{j=1}^{p}$. Also let

$$A=\left(A_{jk}\right)_{1\le j,k\le p},\ A_{jk}=\pi_k^{-j-\nu}.$$

Note that $A_{jk}=\int_0^\infty\phi_j\left(\lambda\right)\Phi_k\left(\lambda\right)d\lambda$ for $j>-\nu$. Since A is a simple modification of a Vandermond matrix, we have:

$$\det A=\prod_{j=1}^{p}\frac{1}{\pi_j^{\nu+1}}\prod_{1\le j<k\le p}\left(\pi_k^{-1}-\pi_j^{-1}\right) \tag{14}$$

Thus A is invertible when all π's are distinct.

Next, define the operators $B:L^2\left((0,\infty)\right)\to l^2\left(\{1,...,p\}\right)$, and $C:l^2\left(\{1,...,p\}\right)\to L^2\left((0,\infty)\right)$ by

$$B\left(j,\lambda\right)=\phi_j\left(\lambda\right),C\left(\lambda,k\right)=\Phi_k\left(\lambda\right).$$

and let f be the operator $L^2\left((0,\infty)\right)\to L^2\left((0,\infty)\right)$ of multiplication by a real-valued measurable bounded function $f(x)$. Then since

$$\frac{n^j}{\Gamma(j)}\int_0^\infty f\left(\lambda\right)\lambda^{j-1}e^{-n\pi_{\alpha(k)}\lambda}d\lambda=\left(BfC\right)\left(j-\nu,\alpha(k)\right) \tag{15}$$

for $j = 1, ..., n$, we find, using (13), that

$$E \prod_{j=1}^{n} \left(1 + f\left(\lambda_j\right)\right) = \text{const} \cdot \sum_{\alpha} (-1)^{|\alpha|} V\left(\pi_{\bar{\alpha}}\right) \det\left(A^{(\alpha)} + B^{(\alpha)} f C^{(\alpha)}\right), \tag{16}$$

where $A^{(\alpha)}$ is a submatrix of A that consists of the intersection of its columns with numbers $\alpha(1), ..., \alpha(n)$ and its last n rows, that is $A^{(\alpha)} = \left(A_{j-\nu,\alpha(k)}\right)_{1 \le j,k \le n}$. Similarly, $B^{(\alpha)}$ is an operator with kernel that consists of the last n elements of the kernel of B, and $C^{(\alpha)}$ is an operator with kernel that consists of elements of the kernel of C with numbers $\alpha(1), ..., \alpha(n)$.

Note that the right hand side of (16) is proportional to the Laplace expansion of $\det\left(A - BfC\right)$. To see this, use (15) and observe that the kernel of $A - BfC$ has the following form:

$$\begin{pmatrix} \pi_1^{-1-\nu} & \cdots & \pi_p^{-1-\nu} \\ \vdots & & \vdots \\ 1 & \cdots & 1 \\ \pi_1^{-1} + \frac{n}{\Gamma(1)} \int_0^\infty f(\lambda) e^{-n\pi_1 \lambda} d\lambda & \cdots & \pi_p^{-1} + \frac{n}{\Gamma(1)} \int_0^\infty f(\lambda) e^{-n\pi_p \lambda} d\lambda \\ \vdots & & \vdots \\ \pi_1^{-n} + \frac{n^n}{\Gamma(n)} \int_0^\infty f(\lambda) \lambda^{n-1} e^{-n\pi_1 \lambda} d\lambda & \cdots & \pi_p^{-n} + \frac{n^n}{\Gamma(n)} \int_0^\infty f(\lambda) \lambda^{n-1} e^{-n\pi_p \lambda} d\lambda \end{pmatrix}.$$

Hence

$$E \prod_{j=1}^{n} \left(1 + f\left(\lambda_j\right)\right) = \text{const} \cdot \det\left(A + BfC\right) = \text{const} \cdot \det\left(A^{-1}\right) \det\left(1 + A^{-1} B f C\right).$$

So, interchanging the order of the composition of operators under the determinant,

$$E \prod_{j=1}^{n} \left(1 + f\left(\lambda_j\right)\right) = \text{const} \cdot \det\left(1 + C A^{-1} B f\right).$$

By setting $f(\cdot)$ equal to the minus indicator function of (s, ∞) and letting $s \to \infty$ in both sides of the above equality, we find that "const" in the above formula equals 1. Thus

$$E\prod_{j=1}^{n}\left(1+f\left(\lambda_j\right)\right)=\det\left(1+CA^{-1}Bf\right).$$

After this point, the proof of Baik et al. (2005) goes practically without any changes. We will provide it here to make this paper self-contained. The kernel of the operator $CA^{-1}B$ in the above determinant is

$$CA^{-1}B\left(\eta,\zeta\right)=\sum_{j=1}^{p}C\left(\eta,k\right)\left(A^{-1}B\right)\left(j,\zeta\right),\qquad \eta,\zeta>0.$$

Further, from Cramer's rule,

$$\left(A^{-1}B\right)\left(j,\zeta\right)=\frac{\det A^{(j)}\left(\zeta\right)}{\det A} \tag{17}$$

where $A^{(j)}\left(\zeta\right)$ is the matrix given by A with jth column replaced by the vector $\left(\phi_1\left(\zeta\right),...,\phi_p\left(\zeta\right)\right)'$. To compute $\det A^{(j)}$, note that

$$\frac{1}{2\pi i}\int_{\Sigma}\frac{e^w}{w^a}dw=\begin{cases}\frac{1}{\Gamma(a)} \text{ if } a \text{ is positive integer}\\ 0 \text{ is } a \text{ is zero or negative}\end{cases},$$

where Σ is any simple closed contour enclosing the origin 0 with counter-clockwise orientation. By replacing $w\to\zeta nw$ and setting $a=k+\nu$, this implies that

$$\frac{\zeta^{-(k-1+\nu)}}{2\pi i}\int_{\Sigma}e^{\zeta nw}\frac{n}{\left(nw\right)^{k+\nu}}dw=\begin{cases}\frac{1}{\Gamma(k+\nu)} \text{ if } k>-\nu\\ 0 \text{ if } k\leq-\nu\end{cases}$$

and therefore

$$\phi_k\left(\zeta\right)=\frac{1}{2\pi i}\int_{\Sigma}e^{\zeta n(w-q)}\frac{n}{w^{k+\nu}}dw.$$

Substituting this formula for $\phi_k\left(\zeta\right)$ in the j-th column of $A^{(j)}$, and pulling out the integrals over w,

$$\det A^{(j)}(\zeta) = \frac{1}{2\pi i}\int_{\Sigma} e^{\zeta n(w-q)} \det\left(A'(w)\right) n dw,$$

where the entries of $A'(w)$ are $A'_{ab}(w) = 1/p_b^{a+\nu}$, where $p_b = \pi_b$ when $b \neq j$ and $p_b = w$ when $b = j$. Hence, by the formula for the Vandermonde determinant,

$$\det A^{(j)}(\zeta) = \prod_{k \neq j} \frac{1}{\pi_k^{1+\nu}} \frac{1}{2\pi i}\int_{\Sigma} e^{\zeta n(w-q)} \prod_{1\leq a<b\leq n}\left(p_b^{-1} - p_a^{-1}\right)\frac{ndw}{w^{1+\nu}}$$

and so, using (14) and (17), we obtain:

$$\left(A^{-1}B\right)(j,\zeta) = \frac{n\pi_j^{p+\nu}}{2\pi i}\int_{\Sigma} e^{\zeta n(w-q)} \prod_{k\neq j}\frac{w-\pi_k}{\pi_j - \pi_k}\frac{dw}{w^{p+\nu}}$$

But for any simple closed contour Γ that encloses $\pi_1, ..., \pi_p$ but excludes w, and is oriented counter-clockwise:

$$\frac{1}{2\pi i}\int_{\Gamma} z^n e^{-\eta n z}\frac{1}{w-z}\prod_{k=1}^{p}\frac{w-\pi_k}{z-\pi_k}dz = \sum_{j=1}^{p}\pi_j^n e^{-n\pi_j\eta}\prod_{k\neq j}\frac{w-\pi_k}{\pi_j-\pi_k}.$$

Therefore, we find

$$CA^{-1}B(\eta,\zeta) = \frac{n}{(2\pi i)^2}\int_{\Gamma} dz \int_{\Sigma} dw e^{-\eta n(z-q)+\zeta n(w-q)}\frac{1}{w-z}\prod_{k=1}^{p}\frac{w-\pi_k}{z-\pi_k}\left(\frac{z}{w}\right)^n,$$

which completes the proof when all π_j are distinct. When some π_j are equal, the formula for the kernel $CA^{-1}B(\eta,\zeta)$ follows by taking proper limits and using l'Hospital's theorem.

For the case when $n \geq p$ and $f(x)$ equals minus the indicator function for the interval (s,∞), where $s \in R$, the proposition is equivalent to Proposition 2.1 of Baik et al. (2005). Extending Baik et al.'s proof to the case of general $f(x)$ while keeping their assumption that $n \geq p$ is straightforward. To save space, we omit such an extension from the proof.□

In the next section, we will use Proposition 1 to extend the results of El Karoui (2007) to the case of several largest eigenvalues of a complex singular Wishart matrix. In conclusion of this section we would like to note that our extension of Proposition 2.1 of Baik et al. (2005) implies that the main results of that paper, namely Theorem 1.1 and Theorem 1.2, hold under an assumption that n/p (M/N in the notations of Baik et al. (2005)) remains in a compact subset of $(0,+\infty)$ as both n and p go to infinity. This assumption relaxes Baik et al.'s requirement that n/p remains in a compact subset of $[1,+\infty)$. The Appendix contains a brief list of changes that should be made to the proofs of Baik et al. (2005) (beyond the extension of Proposition 2.1) to justify such a relaxation.

3 Extension of El Karoui (2007)

In this section, I will prove that the joint distribution of the first m scaled and centered eigenvalues of a complex Wishart matrix $W_{\mathbb{C}}(\Sigma_p/n, n)$ weakly converges to the m-dimensional joint Tracy-Widom distribution. Such a converges takes place in both cases: $n \geq p$ and $n < p$. The scaling and centering sequences are the same for all of the m eigenvalues and have the form proposed by El Karoui (2007).

Proposition 2. *Let H_p be the spectral distribution of Σ_p. Let c be the unique solution in $[0, \pi_1)$ of the equation*

$$\int \left(\frac{\lambda c}{1-\lambda c}\right)^2 dH_p(\lambda) = \frac{n}{p}.$$

Assume that n/p remains in a compact subset of $(0,\infty)$, $\limsup \pi_1^{-1} < \infty$, $\liminf \pi_p^{-1} > 0$, and $\limsup c/\pi_1 < 1$. Define

$$\begin{aligned} \mu_{n,p} &= \frac{1}{c}\left(1 + \frac{p}{n}\int \frac{\lambda c}{1-\lambda c} dH_p(\lambda)\right), \text{ and} \\ \sigma_{n,p} &= \frac{1}{n^{2/3}c}\left(1 + \frac{p}{n}\int \left(\frac{\lambda c}{1-\lambda c}\right)^3 dH_p(\lambda)\right)^{1/3}. \end{aligned}$$

Then, as n and p go to infinity, the joint distribution of the first m centered and scaled eigenvalues $\sigma_{n,p}^{-1}(\lambda_1 - \mu_{n,p}), \ldots, \sigma_{n,p}^{-1}(\lambda_m - \mu_{n,p})$ of matrix $W_{\mathbb{C}}(\Sigma_p/n, n)$ weakly converges to *the m-dimensional joint Tracy-Widom distribution.*

Proof: For a short proof of the uniqueness of c, see El Karoui's (2007) Formula (11) and a discussion that follows the formula. Let us, first, prove that $\liminf c > 0$ and $\limsup c < \infty$. Since by assumption $\liminf n/p > 0$, there exist $\gamma > 0$ such that $n/p > \gamma^2$. We have:

$$\gamma^2 < \frac{n}{p} = \int \left(\frac{\lambda c}{1 - \lambda c} \right)^2 dH_p(\lambda) \leq \left(\frac{c/\pi_1}{1 - c/\pi_1} \right)^2$$

and therefore $c > \pi_1 \frac{\gamma}{1+\gamma}$. This implies that $\liminf c > 0$ because, by assumption, $\limsup \pi_1^{-1} < \infty$. Further, since $c < \pi_1 \leq \pi_p$, the assumption that $\liminf \pi_p^{-1} > 0$ implies that $\limsup c < \infty$. Note that the just established facts that $\liminf c > 0$ and $\limsup c < \infty$, and the assumptions that $\limsup c/\pi_1 < 1$ and that n/p remains in a compact subset of $(0, \infty)$ imply that $\mu_{n,p}$ remains in a compact subset of $(0, \infty)$ and $\sigma_{n,p}$ decays to zero as fast as $n^{-2/3}$ when n goes to infinity.

Now, let $x_1 > ... > x_m$ be any real numbers. Since $\mu_{n,p}$ remains in a compact subset of $(0, \infty)$ whereas $\sigma_{n,p}$ goes to zero as $n \to \infty$, there exist $N > 0$ such that for any $n > N$, $s_i = \mu_{n,p} + \sigma_{n,p} x_i$, $i = 1, ..., m$, are positive numbers. In what follows, we will always take $n > N$. Consider function $f_{z_1,...,z_m}(x) = \sum_{k=1}^m (z_k - 1) \chi_{J_k}(x)$, where $J_1, ..., J_k$ equal $(s_1, \infty), (s_2, s_1], ..., (s_m, s_{m-1}]$ and $z_1, ..., z_m$ are any complex numbers. According to Formulas (2) and (3),

$$\Pr(\lambda_1 \leq s_1, ..., \lambda_m \leq s_m) = \sum_{\{i_1,...,i_m\} \in I} \frac{1}{i_1! ... i_m!} \frac{\partial^{i_1 + ... + i_m}}{\partial z_1^{i_1} ... \partial z_m^{i_m}} E \prod_{j=1}^p \left[1 + f_{z_1,...,z_m}(\lambda_j)\right] \Bigg|_{z_1 = ... = z_m = 0} \tag{18}$$

By Proposition 1, we equivalently can say that $\Pr(\lambda_1 \leq s_1, ..., \lambda_m \leq s_m)$ can be expressed as a sum of a few coefficients in the power expansion of $\det(1 + K_{n,p} f_{z_1,...,z_m})$.

Consider a re-scaled and re-centered version of the kernel of the operator $K_{n,p}$:

$$S_{n,p}(u, v) = \sigma_{n,p} K_{n,p} \left(\mu_{n,p} + \sigma_{n,p} u, \mu_{n,p} + \sigma_{n,p} v \right).$$

Let $S_{n,p}$ be an operator with kernel $S_{n,p}(u,v)$, which acts on $L^2((x_m,\infty))$. Note that

$$\det(1+K_{n,p}f_{z_1,...,z_m}) = \det(1+S_{n,p}g_{z_1,...,z_m})$$

where $g_{z_1,...,z_m}(x) = \sum_{k=1}^{m}(z_k-1)\chi_{R_k}(x)$ and $R_1,...,R_m$ equal $(x_1,\infty),(x_2,x_1],...,(x_m,x_{m-1}]$, respectively.

Under the conditions of Proposition 2 and an additional condition that $n \geq p$, El Karoui (2007) proves that there exists $\varepsilon > 0$ such that $S_{n,p}$ converges in trace class norm to an operator $E \cdot Ai \cdot E^{-1}$ from the trace class, where E is the operator of multiplication by $e^{-\varepsilon x}$ and Ai is an integral operator acting on $L^2((x_m,\infty))$, which has the Airy kernel $Ai(x,y) = \int Ai(x+u)Ai(y+u)du$. A careful reading of his proofs reveals that he only needs the additional condition $n \geq p$ to be able to use Proposition 1.2 of Baik et al. (2005). For all other purposes, the inequality $n/p \geq 1$ in his proofs can be substituted by $n/p > \gamma^2 > 0$ without changing validity of the proofs. Therefore, our Proposition 1 and El Karoui's (2007) result imply the convergence of $S_{n,p}$ to $E \cdot Ai \cdot E^{-1}$ without the extra condition that $n \geq p$.

Since the trace class operators form an ideal in the algebra of linear bounded operators, the operators $S_{n,p} \cdot g_{z_1,...,z_m}$ and $E \cdot Ai \cdot E^{-1} \cdot g_{z_1,...,z_m}$ must be from the trace class. Further, since $g_{z_1,...,z_m}$ is a bounded function for all $z_1,...,z_m$, $\|S_{n,p} \cdot g_{z_1,...,z_m} - E \cdot Ai \cdot E^{-1} \cdot g_{z_1,...,z_m}\|_1 \leq \|S_{n,p} \cdot g_{z_1,...,z_m} - E \cdot Ai \cdot E^{-1}\|_1 \|g_{z_1,...,z_m}\|$, which converges to zero as n and p go to infinity. Here $\|K\|_1$ denotes the trace class norm of operator K and the above norm inequality follows from the inequalities of Theorem 1.6 in Simon (2005). Hence, $S_{n,p} \cdot g_{z_1,...,z_m}$ converges to $E \cdot Ai \cdot E^{-1} \cdot g_{z_1,...,z_m}$ in the trace class norm.

Now, since Fredholm determinant is continuous with respect to the trace class norm, $\det(1+S_{n,p} \cdot g_{z_1,...,z_m})$ converges to $\det(1+E \cdot Ai \cdot E^{-1} \cdot g_{z_1,...,z_m})$ for any $z_1,...,z_m$. Further, since E^{-1} and $g_{z_1,...,z_m}$ commute, $\det(1+E \cdot Ai \cdot E^{-1} \cdot g_{z_1,...,z_m}) = \det(1+E \cdot Ai \cdot g_{z_1,...,z_m} \cdot E^{-1})$. But determinants are invariant with respect to conjugation which leaves an operator in the trace class (see Remark 2.1 in Baik et al. (2005)). Therefore, we have

$$\det\left(1+S_{n,p}\cdot g_{z_1,\dots,z_m}\right)\to\det\left(1+Ai\cdot g_{z_1,\dots,z_m}\right)$$

for any $z_1,\dots,z_m$.

Since $\det\left(1+S_{n,p}\cdot g_{z_1,\dots,z_m}\right)$ exactly equals $E\left(\prod_{j=1}^{p}\left(1+\sum_{k=1}^{m}\left(z_k-1\right)\chi_{J_k}\left(\lambda_j\right)\right)\right)$, it is a finite order polynomial in $z_1,\dots,z_m$, and, hence, an analytic function of $z_1,\dots,z_m$. Further, as follows, for example, from Formulas 1.30 and 1.32 in Soshnikov (2000),

$$\left|\det\left(1+S_{n,p}\cdot g_{z_1,\dots,z_m}\right)\right|\leq\text{const}\cdot\exp\left(\max_{j=1,\dots,m}\left|z_j-1\right|\left\|S_{n,p}\right\|_1\right)\tag{19}$$

(see also Lemma 3.3 of Simon, 2005, for the case when $m=1$). Since $S_{n,p}$ converges in the trace class norm to $E\cdot Ai\cdot E^{-1}$, there exists a constant M such that $\left\|S_{n,p}\right\|_1<M$ for all n and p. This fact together with inequality (19) imply that $\det\left(1+S_{n,p}\cdot g_{z_1,\dots,z_m}\right)$ form a normal family of analytic functions (see Rudin, 1980, p. 5). Hence, convergence of $\det\left(1+S_{n,p}\cdot g_{z_1,\dots,z_m}\right)$ to $\det\left(1+Ai\cdot g_{z_1,\dots,z_m}\right)$ is uniform on any compact set in $\mathbb{C}^m$ and therefore all derivatives of $\det\left(1+S_{n,p}\cdot g_{z_1,\dots,z_m}\right)$, and thus, all derivatives of $E\prod_{j=1}^{p}\left[1+f_{z_1,\dots,z_m}(\lambda_j)\right]$, converge to the corresponding derivatives of $\det\left(1+Ai\cdot g_{z_1,\dots,z_m}\right)$ at $z_1=\dots=z_m=0$.

As shown by Johansson (2001) (see his Formulas (1.19), (3.46) and (3.48)), $F(x_1,\dots,x_m)$ defined as

$$F(x_1,\dots,x_m)=\sum_{\{i_1,\dots,i_m\}\in I}\frac{1}{i_1!\dots i_m!}\frac{\partial^{i_1+\dots+i_m}}{\partial z_1^{i_1}\dots\partial z_m^{i_m}}\det\left(1+Ai\cdot g_{z_1,\dots,z_m}\right)\Bigg|_{z_1=\dots=z_m=0}$$

is the distribution function for the m-dimensional joint Tracy-Widom distribution. Using formula (18) and the just established convergence of the derivatives of $E\prod_{j=1}^{p}\left[1+f_{z_1,\dots,z_m}(\lambda_j)\right]$ to those of $\det\left(1+Ai\cdot g_{z_1,\dots,z_m}\right)$, we conclude that

$$\Pr\left(\frac{\lambda_1-\mu_{n,p}}{\sigma_{n,p}}\leq x_1,\dots,\frac{\lambda_m-\mu_{n,p}}{\sigma_{n,p}}\leq x_m\right)\to F(x_1,\dots,x_m).$$

Since $F(x_1, ..., x_m)$ is a continuous function, such a convergence implies that the joint distribution of $\sigma_{n,p}^{-1}(\lambda_1 - \mu_{n,p}), ..., \sigma_{n,p}^{-1}(\lambda_m - \mu_{n,p})$ weakly converges to the m-dimensional joint Tracy-Widom distribution.$\square$

In conclusion of this section we note that since El Karoui (2007) used the assumption that $n/p \geq 1$ only to be able to use Proposition 2.1 of Baik et al. (2005), our finding that that proposition holds for $n < p$ proves that all El Karoui's (2007) results remain true for n/p remaining in a compact subset of $(0, \infty)$.

4 Application

In this section I will show how Proposition 2 can be used in the analysis of excess stock return data generated by the approximate factor model (1). An approximate factor model for asset returns form the core of Chamberlain and Rothschild's (1983) extension of the arbitrage pricing theory (APT) of Ross (1976). The APT is one of the most important finance theories that shows that asset prices must be well explained by covariances of asset returns with a few common risk factors. An important practical question is how many such common factors exist.

This question has attracted considerable research attention. Roll and Ross (1980, p.1092) find that "at least three factors are important for pricing, but that it is unlikely that more than four are present". Brown and Weinstein (1983, p.713) "find evidence that there may be as few as 3 economywide factors, and certainly no more than 5 if the arbitrage pricing model is correct". Trzcinka (1986) finds that there may be one to five common risk factors. Connor and Korajczyk (1993) report one or two factors in non-January months but three to six factor for January returns. Huang and Jo (1995, p.988) find that "the evidence supports only a small number of factors, generally one and at most two". Bai and Ng (2002) estimate the number of common factors in stock returns to be two. Two is also preferred number of factors in Onatski (2005). Makarov and Papanikolaou (2007) find evidence that there are four factors in stock returns.

In general, researchers find a small number of factors in the approximate factor model for excess stock returns.[5] Often, they are uncertain about their point estimates. The uncertainty about the point estimates is also reflected in the fact that different researchers often provide conflicting estimates. Even though the uncertainty is well recognized, its amount has never been formally quantified. Below, I will try to quantify this uncertainty. Precisely, I will find an asymptotic 95% confidence set for the number of factors by inverting a statistical test for the number of factors partially developed in a companion paper Onatski (2007).

In the companion paper, I am interested in testing the null of k_0 factors vs. the alternative that the number of factors k is larger than k_0 but smaller than $k_{\max}+1$, where $k_{\max}$ is an *a priori* maximum number of factors. I assume that the real-valued data $\{\xi_{it}\}_{1\leq i\leq p,1\leq t\leq N}$, where $N=2n$, is generated by model (1), where the vectors of idiosyncratic components $\{\eta_{it}\}_{1\leq i\leq p}$ are i.i.d. Gaussian $N(0,\Sigma_p/2)$ and independent from factors $\{F_t\}_{1\leq t\leq N}$. To test the hypothesis, I propose, first, to construct a new complex-valued data set $\left\{\tilde{\xi}_{it}\right\}_{1\leq i\leq p,1\leq t\leq n}$, where $\tilde{\xi}_{it}=\xi_{it}+\sqrt{-1}\xi_{i,t+n}$, and then to compute a test statistic $\max_{k_0<i\leq k_{\max}}\left(\gamma_i-\gamma_{i+1}\right)/\left(\gamma_{i+1}-\gamma_{i+2}\right)$, where γ_i is the i-th largest eigenvalue of the sample covariance matrix of the new dataset (this matrix is defined as $\frac{\tilde{X}\tilde{X}^*}{n/2}$ where $\tilde{X}=\left\{\tilde{\xi}_{it}\right\}_{1\leq i\leq p,1\leq t\leq n}$). I show that under the null, the asymptotic distribution of the proposed test statistic as n and p rise so that n/p remains in a compact subset of $(0,\infty)$ is the same as the asymptotic distribution of $\max_{0<i\leq k_{\max}-k_0}\left(\lambda_i-\lambda_{i+1}\right)/\left(\lambda_{i+1}-\lambda_{i+2}\right)$, where λ_i is the i-th largest eigenvalue of a $W_{\mathbb{C}}\left(\Sigma_p/n,n\right)$ matrix. Under the alternative, the test statistics explodes in probability as n and p rise.

Proposition 2 implies that the asymptotic distribution of $\max_{0<i\leq k_{\max}-k_0}\left(\lambda_i-\lambda_{i+1}\right)/\left(\lambda_{i+1}-\lambda_{i+2}\right)$ equals the distribution of $\max_{0<i\leq k_{\max}-k_0}\left(\mu_i-\mu_{i+1}\right)/\left(\mu_{i+1}-\mu_{i+2}\right)$, where $\mu_1,...,\mu_{k_{\max}-k_0}$ have the joint $(k_{\max}-k_0)$-dimensional Tracy-Widom distribution. This result allows us to

[5] See, however, Dhrymes et al. (1984) who find that the estimated number of factors rises with the dimensionality of data. It is important to realize that Dhrymes et al. (1984) study the classical factor model whereas we are interested in the approximate factor model. Many "classical factors" would be considered by an approximate factor model as part of the idiosyncratic component of data.

%	$k_{\max} - k_0$							
	1	2	3	4	5	6	7	8
50	1.27	1.95	2.30	2.54	2.74	2.92	3.09	3.24
60	1.53	2.24	2.59	2.88	3.10	3.31	3.49	3.65
70	1.86	2.61	3.01	3.32	3.59	3.82	4.01	4.20
80	2.37	3.19	3.65	4.02	4.32	4.59	4.83	5.05
85	2.75	3.62	4.15	4.54	4.89	5.20	5.45	5.70
90	3.33	4.31	4.91	5.40	5.77	6.13	6.42	6.66
91	3.50	4.49	5.13	5.62	6.03	6.39	6.67	6.92
92	3.69	4.72	5.37	5.91	6.31	6.68	6.95	7.25
93	3.92	4.99	5.66	6.24	6.62	7.00	7.32	7.59
94	4.20	5.31	6.03	6.57	7.00	7.41	7.74	8.04
95	4.52	5.73	6.46	7.01	7.50	7.95	8.29	8.59
96	5.02	6.26	6.97	7.63	8.16	8.61	9.06	9.36
97	5.62	6.91	7.79	8.48	9.06	9.64	10.11	10.44
98	6.55	8.15	9.06	9.93	10.47	11.27	11.75	12.13
99	8.74	10.52	11.67	12.56	13.42	14.26	14.88	15.25

Table 1: Approximate percentiles of the test statistics for the tests of k_0 factors vs. an alternative of more than k_0 but less than $k_{\max} + 1$ factors

tabulate the critical values of the asymptotic distribution by using Monte Carlo simulations of large dimensional matrices from GUE to approximate the joint Tracy-Widom distribution. Table 1 contains such critical values[6] for $k_{\max} - k_0 = 1, 2, ..., 8$. For example, the approximate 5% critical value of the test of 3 factors versus the alternative $3 < k \leq 10$ is in the 5th row (counting from the bottom up) and 2nd column (counting from the right) of the table. It equals 8.29.

Our test procedure can be interpreted as formalizing the widely used empirical method of the number of (classical) factors determination based on the visual inspection of the scree plot introduced by Cattell (1966). The scree plot is a line that connects the decreasing eigenvalues of the sample covariance matrix of the data plotted against their respective order numbers. In practice, it often happens that the scree plot shows a sharp break where the true number of factors ends and "debris" corresponding to the idiosyncratic influences

[6] We approximate the joint 10-dimensional Tracy-Widom distribution of type two by the distribution of 10 largest eigenvalues of a 1000×1000 matrix from the Gaussian Unitary Ensemble. We obtain an approximation for the latter distribution by simulating 30,000 independent matrices from the ensemble and numerically computing their 10 first eigenvalues.

appears. Our test statistic effectively measures the curvature of the scree plot at a would-be break point under the alternative hypothesis. When the alternative hypothesis is true, the curvature asymptotically goes to infinity. In contrast, under the null, the curvature has a non-degenerate asymptotic distribution that does not depend on the model's parameter Σ_p.

To construct a 95% confidence interval for the number of common factors in excess stock returns, I use the data on monthly returns on $p = 972$ stocks traded on the NYSE, AMEX, and NASDAQ during the period from January 1983 to December 2006 provided by the Center for Research in Security Prices (CRSP). My data set includes those and only those companies for which CRSP provides monthly holding period return data for all months in the studied time interval. To obtain the excess returns on the stocks I subtract 1-month risk-free rate provided by CRSP from the stock returns.

Since previous empirical research suggests that the number of common risk factors may be different in January and non-January months, I drop January months from the data which leaves me with $N = 264$ time observations of real-valued data. To get a complex-valued data set, I divide the real-valued data into two parts: the first containing all observations from February 1983 to December 1994, and the second containing all observations from February 1995 to December 2006. Then I add the data from the first sub-period and the imaginary unit times the data from the second sub-period. Hence, the dimensionality of my complex-valued data set is $p = 972$, $n = 132$. Note that p is larger than n.

I maintain an assumption that the true number of factors is strictly less than seven. I choose seven as an upper bound for the number of factors because it is consistent with previous studies. I will include $j < 7$ into the asymptotic 95% confidence set for the number of factors if a 5%-asymptotic-size test of the null of j factors versus the alternative that the number of factors is larger than j but no larger than seven does not reject the null.

Table 2 contains the first nine eigenvalues $\gamma_1, ..., \gamma_9$ of the sample covariance matrix of my complex-valued data set, the quantities $\left(\gamma_i - \gamma_{i+1}\right) / \left(\gamma_{i+1} - \gamma_{i+2}\right)$, $i = 1, ..., 7$, the test statistics $\max_{k_0 < i \leq 7} \left(\gamma_i - \gamma_{i+1}\right) / \left(\gamma_{i+1} - \gamma_{i+2}\right)$ for the tests of the nulls of $k_0 = 0$, $k_0 = 1$,

i	1	2	3	4	5	6	7	8	9
γ_i	3.99	1.16	0.76	0.66	0.50	0.48	0.40	0.37	0.34
$\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$	7.14	3.77	0.65	12.73	0.15	2.98	1.18		
k_0	0	1	2	3	4	5	6		
$\max_{k_0<i\leq 7}\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$	12.73	12.73	12.73	12.73	2.98	2.98	1.18		
Critical values	8.29	7.95	7.50	7.01	6.46	5.73	4.52		

Table 2: The largest eigenvalues of the sample covariance matrix of the complex-valued data and the test statistics for the tests of hypotheses that the number of factors equals k_0 versus the alternative of more than k_0 but no more than 7 factors.

..., $k_0 = 6$ factors versus the alternatives that the number of factors is larger than k_0 but no larger than 7, and the corresponding 5% critical values taken from Table 1.

I reject the nulls of $0, 1, 2,$ and 3 factors by the tests of asymptotic size 5%, but cannot reject the nulls of 4, 5, and 6 factors. Hence, my 95% confidence set for the true number of common factors in the excess stock returns is $\{4, 5, 6\}$. This set intersects with the range of estimates proposed by Roll and Ross (1980), Brown and Weinstein (1983), and Trzcinka (1986). It includes the point estimate 4 found in Makarov and Papanikolaou (2007). It is disjoint with the set of estimates reported by Connor and Korajczyk (1993), Huang and Jo (1995), and Bai and Ng (2002). My own previous estimate 2 reported in Onatski (2005) is not included in the set.

The 95% confidence set $\{4, 5, 6\}$ should, perhaps, be appealing to the proponents of the multifactor financial models. The good news is at least that 0 and 1 do not enter the set. However, this result should be taken with a grain of salt. The reason is that the rejection of the nulls of 0, 1, 2 and 3 factors may be due to a failure of some of the primitive assumptions of the tests. For example, no one would truly believe that the idiosyncratic components of the excess stock returns are Gaussian and independent over time. Soshnikov's (2002) results on the universality of the Tracy-Widom limit for the largest eigenvalue of sample covariance matrices require n/p approaching to 1 as n goes to infinity and assume that the tails of the distribution of the data points are relatively thin. For the financial data, the tails of the distribution may be heavy. Then, the above test for the number of factors will be invalid.

In fact, the approximate factor model would no longer be a plausible description of the data because it assumes the existence of the second moments of the data.

Another discouraging possibility is that the asymptotics that the tests rely on may poorly approximate the finite sample situation. Perhaps most importantly, the result of Onatski (2007) that the asymptotic distribution of $\max_{k_0<i\leq k_{\max}}\left(\gamma_i-\gamma_{i+1}\right)/\left(\gamma_{i+1}-\gamma_{i+2}\right)$ is the same as that of $\max_{0<i\leq k_{\max}-k_0}\left(\lambda_i-\lambda_{i+1}\right)/\left(\lambda_{i+1}-\lambda_{i+2}\right)$, where λ_i is the i-th largest eigenvalue of a $W_{\mathbb{C}}\left(\Sigma_p/n,n\right)$ matrix, substantially uses the fact that if the true number of factors is k_0, then under the null, $\gamma_{k_0}/\gamma_{k_0+1}\to\infty$. A casual inspection of the second raw of Table 2 reveals that none of the ratios γ_i/γ_{i+1} are large. Hence, to comfortably use the obtained 95% confidence set one should check whether the asymptotic requirements in Onatski (2007) can be relaxed. I leave such a check for future research.

5 Conclusion

In this paper, I have shown that the joint distribution of the centered and normalized several largest eigenvalues of a p-dimensional complex Wishart matrix $W_{\mathbb{C}}\left(\Omega,n\right)$ converges to the joint Tracy-Widom distribution as n and p tend to infinity so that n/p remains in a compact subset of $(0,\infty)$. This result extends Baik et al. (2005) and El Karoui (2007) in two directions. First, a several largest eigenvalues as opposed to the single largest eigenvalue has been analyzed. Second, and most importantly, n is allowed to be smaller than p, a situation corresponding to $W_{\mathbb{C}}\left(\Omega,n\right)$ being a singular matrix.

I also have shown that all results of Baik et al. (2005) and El Karoui (2007) remain true if the assumption that n/p remains in a compact subset of $[1,\infty)$, which these two papers make, is substituted by a less restricting assumption that n/p remains in a compact subset of $(0,\infty)$.

Finally, I have demonstrated how the theoretical result of this paper can be used to find a 95% confidence set for the number of common factors in excess stock returns. The established confidence set is $\{4,5,6\}$. Such a set formally quantifies the uncertainty about the true number of factors in excess stock returns present in the previous studies of the number of factors. The set supports some of the previous studies, but not others. I have pointed out possible drawbacks in the proposed methodology of obtaining the 95% confidence set, which suggest future directions of research.

6 Appendix

A proof of formula (5).

Consider the following analytic homeomorphism g (see James, 1954, for a useful summary of concepts from the differential geometry) of almost all the $U(p)$ on almost all the product $CV(p,n) \times U(p-n)$. Let $g(R) = \{Q_1, S\}$, where $Q_1 = R_1$, $S = H_{R_1}^* R_2$, H_{R_1} is such that $[R_1, H_{R_1}] \in U(p)$ and the elements of H_{R_1} are analytic functions of R_1. Columns of H_{R_1} can, for example, be obtained by applying the Gram-Schmidt orthogonalization procedure to the projections of the first $p-n$ vectors of a fixed basis in $\mathbb{C}^p$ on the $p-n$-dimensional subspace orthogonal to the columns of R_1. Such a construction will work for all R such that the first $p-n$ vectors of the fixed basis and the columns of R_1 are linearly independent. Hence, it will work for almost all the $U(p)$. The inverse of g is given by $g^{-1}(\{Q_1, S\}) = [Q_1, H_{Q_1} S]$.

The homeomorphism g maps the differential form $(R^* dR) \equiv (R_1^* dR_1)(R_2^* dR_1)(R_2^* dR_2)$ to the product of forms $(Q_1^* dQ_1)$, $\left(S^* H_{Q_1}^* dQ_1\right)$, and $\left(S^* H_{Q_1}^* d\left(H_{Q_1} S\right)\right)$. We have: $(Q_1^* dQ_1)\left(S^* H_{Q_1}^* dQ_1\right) = (Q_1^* dQ_1)\,|S|\left(H_{Q_1}^* dQ_1\right) = (Q^* dQ_1)$, where $Q = [Q_1, H_{Q_1}]$. Further, the form $\left(S^* H_{Q_1}^* d\left(H_{Q_1} S\right)\right)$ can be represented as a sum of $\left(S^* H_{Q_1}^* dH_{Q_1} S\right)$ and $(S^* dS)$. But the product of $(Q^* dQ_1)$ with $\left(S^* H_{Q_1}^* dH_{Q_1} S\right)$ is zero because $(Q^* dQ_1)$ is a form of maximum degree on the Stiefel manifold. Therefore,

$$\begin{aligned}
\int_{R \in U(p)} e^{-\operatorname{tr}\left(\Pi R_1 \Lambda R_1^*\right)} (R^* dR) &= \int_{S \in U(p-n)} \int_{Q_1 \in CV(p,n)} e^{-\operatorname{tr}\left(\Pi Q_1 \Lambda Q_1^*\right)} (Q^* dQ_1)(S^* dS) \\
&= \operatorname{Vol}\{U(p-n)\} \int_{Q_1 \in CV(p,n)} e^{-\operatorname{tr}\left(\Pi Q_1 \Lambda Q_1^*\right)} (Q^* dQ_1). \square
\end{aligned}$$

A list of changes to the proofs in Baik et al. (2005)

Our extension of Proposition 2.1 of Baik et al. (2005) to the case when $n/p < 1$ ($M/N < 1$ in Baik et al.'s notations) implies that the main results of that paper contained in Theorems 1.1 and 1.2 hold if the requirement that M/N remains in a compact subset of $[1, +\infty)$ as both N and M tend to infinity is relaxed to the requirement that M/N remains in a compact subset of $(0, +\infty)$. Here is a list of few extra necessary changes to the proofs of Theorems 1.1 and 1.2 We use the notations of Baik et al. (2005).

1) Formula (134) should be complemented by the statement: clearly, $T_1(t) > 0$ for $0 < \gamma < 1$.

2) The statement after Formula (139) should be changed into: $1 \le \mu = (\gamma + 1)^2 / \gamma^2 < \infty$

3) Formula (141) should be changed into: $\bar{\gamma} \le \gamma \le \gamma_0$ for fixed $\gamma_0 \ge 1$ and $0 < \bar{\gamma} \le 1$.

4) Formula (143) should be changed into: $0 < \delta < \min \left\{ \frac{\nu^3}{6C_0}, \frac{\bar{\gamma}}{(1+\gamma_0)(1+\bar{\gamma})} \right\}$, $C_0 := \frac{(1+\gamma_0)^4 (1+\bar{\gamma})^4}{4\gamma_0^4 \bar{\gamma}^4} \left(1 + \bar{\gamma}^2 \gamma_0^4\right)$.

5) Formula (144) should use definition of C_0 from the change 4)

6) Formula (177) should be complemented by the following reasoning: Define $T_1(t) = (\gamma+1)^2 t^2 + (\gamma^2 - 1) t + 2\gamma$. Note that $\min_{t \in [0, 2p_c]} T_1(t) = \begin{cases} T_1\left(\frac{1-\gamma^2}{2(\gamma+1)^2}\right) & \text{for } 1/5 \le \gamma \le 1 \\ T_1(2p_c) & \text{for } 0 < \gamma < 1/5 \end{cases}$.
But $T_1\left(\frac{1-\gamma^2}{2(\gamma+1)^2}\right) \ge 6/25$ for $1/5 \le \gamma \le 1$ and $T_1(2p_c) = 6\gamma^2$ for $0 < \gamma < 1/5$. Hence $T_1(t) > 0$ for $0 < \gamma < 1$.

7) Formula (237) should be changed in the same way as Formula (141). See change 3) above.

8) Formula (239) should be changed into: $0 < \delta < \min \left\{ \frac{\Pi}{2}, \frac{1}{2(1+\gamma_0)}, \frac{\nu^2}{4C_1} \right\}$, $C_1 := \frac{8}{3}\left(\frac{1}{\Pi^3} + \frac{(1+\gamma_0)^3}{\bar{\gamma}^2}\right)$.

9) Formula (241) should be changed into: $\left|f^{(3)}(s)\right| = \left|\frac{2}{s^3} - \frac{2}{\gamma^2 (s-1)^3}\right| \le \frac{2}{(\pi_1 - \delta)^3} + \frac{2}{\bar{\gamma}^2 (1 - \pi_1 - \delta)^3} \le \frac{16}{\Pi^3} + \frac{16(1+\gamma_0)^3}{\bar{\gamma}^2} = 6C_1$.

References

[1] Anderson, T.W. (1984) An introduction to multivariate statistical analysis. John Wiley & Sons. New York, Chichester, Brisbane, Toronto, Singapore.

[2] Andreief, C. (1883) "Note sur une relation les integrales definies des produits des fonctions". Mem. de la Soc. Sci. Bordeaux, 2.

[3] Bai, J. and Ng, S (2002). "Determining the number of factors in approximate factor models", Econometrica, 70, pp 191-221

[4] Baik, J., Ben Arous, G. and S. Peche (2005) "Phase transition of the largest eigenvalue for non-null complex sample covariance matrices", Annals of Probability , 33(5):1643-1697.

[5] Brown, S. and Weinstein, M., (1983) "A new approach to testing asset pricing models: the bilinear paradigm", Journal of Finance, 38, pp 711-743.

[6] Connor, G. and Korajczyk, R. (1993) "A test for the number of factors in an approximate factor model", The Journal of Finance, 58, pp. 1263-1291

[7] Chamberlain, G. and Rothschild, M. (1983) "Arbitrage, factor structure, and mean-variance analysis on large asset markets", Econometrica, 51, pp.1281-1304.

[8] Dhrymes, P., Friend. I., and Gultekin, N. (1984) "A critical reexamination of the empirical evidence on the arbitrage pricing theory", The Journal of Finance, 39, pp 323-346.

[9] El Karoui, N. (2007) "Tracy-Widom Limit for the Largest Eigenvalue of a Large Class of Complex Wishart Matrices", Annals of Probability, Vol. 35, No 2

[10] Huang, R. and Jo, H. (1995). "Data frequency and the number of factors in stock returns", Journal of Banking and Finance, 19, pp 987-1003.

[11] James, A. (1954) "Normal multivariate analysis and the orthogonal group", Annals of Mathematical Statistics, 25, No.1, 40-75.

[12] James, A. (1964) "Distributions of matrix variates and latent roots derived from normal samples", Annals of Mathematical Statistics, 35, 475-501

[13] Johansson, K. (2001) "Discrete orthogonal polynomial ensembles and the Plancherel measure", Annals of Mathematics, 153, 259-296.

[14] Johnstone, I.M. (2001) "On the distribution of the largest eigenvalue in principal components analysis", Annals of Statistics, 29, 295–327

[15] Makarov, I. and D. Papanikolaou (2007). "Sources of systematic risk", London Business School, manuscript.

[16] Mehta, M (2004) Random Matrices. Elsevier, Academic Press.

[17] Olver, F. (1974) Asymptotics and Special Functions. New York: Academic Press.

[18] Onatski, A. (2005) "Determining the number of factors from the empirical distribution of eigenvalues", manuscript, Columbia University.

[19] Onatski, A. (2007) "A formal statistical test for the number of factors in the approximate factor models", manuscript, Columbia University.

[20] Ratnarajah, T. and Vaillanourt, R. (2005) "Complex Singular Wishart Matrices and Applications", working paper No 3185, Centre de recherches mathematiques of the Universite de Montreal.

[21] Roll, R., and Ross S. (1980) "An empirical investigation of the arbitrage pricing theory", Journal of Finance, 5, pp 1073-1103

[22] Ross, S. (1976) "The arbitrage theory of capital asset pricing", Journal of Economic Theory, 13, pp.341-360.

[23] Rudin, W. (1980) Function Theory in the Unit Ball of C^n. Springer-Verlag. New York, Heidelberg, Berlin.

[24] Simon, B. (2005). Trace Ideals and Their Applications. 2nd edition. American Mathematical Society, Providence.

[25] Soshnikov, A. (2000). "Determinantal random point fields", Rassian Mathematical Surveys, 55:5, pp. 923-975.

[26] Soshnikov, A. (2002) "A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices", Journal of Statistical Physics, 108, No. 5/6, 1033-1056.

[27] Stock, J. and Watson, M. (2002) "Forecasting using principal components from a large number of predictors", Journal of the American Statistical Association, 97, 460, pp. 1167-1179

[28] Tracy C. A. and H. Widom (1994) "Level Spacing Distributions and the Airy Kernel", Communications in Mathematical Physics, 159, pp151-174

[29] Tracy C. A. and H. Widom (1998) "Correlation functions, cluster functions, and spacing distributions for random matrices", Journal of Statistical Physics, Vol. 92, No. 5/6, 809-835.

[30] Trzcinka, C. (1986) "On the number of factors in the arbitrage pricing model", The Journal of Finance, 41, pp 347-368.