A formal statistical test for the number of factors in the approximate factor models

Alexei Onatski

Economics Department, Columbia University

September 26, 2006

Abstract

In this paper we study i.i.d. sequences of $n \times 1$ vectors X_t, $t = 1, ..., T$ such that, for each t, X_t has an approximate factor structure with correlated Gaussian idiosyncratic terms. We develop a test of the null of r factors vs. the alternative that the number of factors is larger than k but smaller than $k_{\max}+1$, where $k_{\max}$ is an a priori maximum number of factors. Our test statistic is equal to the ratio of $\gamma_{k+1} - \gamma_{k_{\max}+1}$ to $\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}$, where γ_i is the i-th largest eigenvalue of the sample covariance matrix $\frac{1}{T}\sum_{t=1}^{T} X_t X_t'$. We describe the asymptotic distribution of the test statistic as n and T go to infinity proportionally as a function of the Tracy-Widom distribution. We tabulate the critical values of the test corresponding to k and $k_{\max}$ relevant for financial and macroeconomic applications. As an application, we test different hypotheses about the number of factors in arbitrage pricing theory. We reject the nulls of no or a single factor against alternatives of more than one factors by the test of 5%-level. We cannot reject the null of 5 factors by the 5%-level test.

1

1 Introduction

Approximate factor analysis of high-dimensional data has attracted a lot of recent attention from researchers in macroeconomics and finance. Factors non-trivially influencing thousands of stock returns observed over hundreds of time periods have been used to study the pricing of financial assets, to evaluate the performance of financial portfolios, and to test arbitrage pricing theory. In macroeconomics, pervasive factors extracted from half a century of data on hundreds of macroeconomic indicators have been used to monitor business cycles, to forecast individual macroeconomic time series, and to augment vector autoregressions used for monetary policy analysis.

An important question to be addressed by any study which uses factor analysis is how many factors there are.

Although there have been many recent studies which develop consistent estimators of the number of factors in the empirically relevant context of large and comparable time and cross-sectional dimensions of the data (see, for example, Forni et al (2000), Bai and Ng (2002, 2005), Stock and Watson (2005), Hallin and Liska (2005), Onatski (2005), and Watson and Amengual (2006)), the corresponding estimates of the number of factors driving stock returns and macroeconomic time series often considerably disagree. Such a disagreement indicates that there exists a large amount of statistical uncertainty about the point estimates of the number of factors. Unfortunately, none of the above studies proposes a formal statistical test of different hypotheses about the number of factors that would quantify the amount of the uncertainty. The purpose of this paper is to develop such a test.

In this paper, we study i.i.d. observations of $n$dimensional vectors X_t, $t = 1, ..., T$, of data which admit the approximate factor structure (see Chamberlain and Rothschild, 1983) with correlated complex Gaussian idiosyncratic terms. If, as in the majority of the empirical applications, the original data is real, we construct X by adding the first half of the original sample and the product of the imaginary unit and the second half of the original sample. We develop a test of the null of k factors vs. the alternative that the number of factors is larger

2

than k but smaller than $k_{\max} + 1$, where $k_{\max}$ is an *a priori* maximum number of factors. Our test statistics is equal to the ratio of $\gamma_{k+1} - \gamma_{k_{\max}+1}$ to $\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}$, where γ_i is the i-th largest eigenvalue of the sample covariance matrix $\frac{1}{T} \sum_{t=1}^{T} X_t X_t'$. We describe the asymptotic distribution of the test statistics as n and T go to infinity proportionally as a function of the Tracy-Widom distribution (see Tracy and Widom, 1994), and tabulate the critical values of the test corresponding to k and $k_{\max}$ relevant for financial and macroeconomic applications.

The logic of our test is based on the standard identification assumption (see Chamberlain and Rothschild, 1983) that the portion of the data's variation per cross-sectional unit explained by the factors remains non-trivial as n tends to infinity, whereas the portion of the variation per cross-sectional unit due to any idiosyncratic influence tends to zero. This assumption is equivalent to the requirement that the first r eigenvalues of the data's covariance matrix (where r is the true number of factors) rise proportionally to n, whereas the rest of the eigenvalues stay bounded. If $k < r$, we would, therefore, expect that $\gamma_{k+1} - \gamma_{k_{\max}+1}$ is much larger than $\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}$. In contrast, when $k = r$, the eigenvalues $\gamma_{k+1}, \gamma_{k_{\max}+1}$, and $\gamma_{k_{\max}+2}$ remain bounded as n tends to infinity, and we do not have a reason to expect the ratio of $\gamma_{k+1} - \gamma_{k_{\max}+1}$ to $\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}$ to be large.

In this paper, we establish the fact that, as both n and T tend to infinity, the joint distribution of the centered and scaled eigenvalues $\sigma^{-1} \left(\gamma_{r+1} - \mu \right), ..., \sigma^{-1} \left(\gamma_{k_{\max}+1} - \mu \right)$, where the centering and scaling constants μ and σ depend on n, T, and the covariance structure of the idiosyncratic terms, converges to the same limit as the joint distribution of the similarly centered and scaled $k_{\max} - r + 1$ largest eigenvalues of the sample covariance matrix of the idiosyncratic terms. Further, we extend recent results of El Karoui (2006) to show that the latter joint distribution converges to the distribution described by Tracy and Widom (1994), which does not depend on the parameters of the data generating process. We conclude that, under the null hypothesis that $k = r$, the asymptotic distribution of the ratio of $\sigma^{-1} \left(\gamma_{k+1} - \mu \right) - \sigma^{-1} \left(\gamma_{k_{\max}+1} - \mu \right)$ to $\sigma^{-1} \left(\gamma_{k_{\max}+1} - \mu \right) - \sigma^{-1} \left(\gamma_{k_{\max}+2} - \mu \right)$ is a function of the Tracy-Widom distribution. Finally, we note that the latter ratio is equal to the ratio of

3

$\gamma_{k+1} - \gamma_{k_{\max}+1}$ to $\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}$. Therefore, its computation does not require the knowledge of the centering and the scaling constants, and hence, the knowledge of the covariance structure of the idiosyncratic terms. We take the ratio of $\gamma_{k+1} - \gamma_{k_{\max}+1}$ to $\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}$ as our test statistics and tabulate the corresponding critical values.

We study the finite sample performance of our test by running several Monte Carlo experiments. We find that the test has correct size and good power for our Monte Carlo design and samples as small as $n = 50$ and $T = 50$. Moreover, the test seems to be reasonably robust with respect to different distributional assumptions about the idiosyncratic terms. Using Monte Carlo experiments we also study the performance of an analog of our test which does not require transforming the original real data into the complex form. Although we are unable to formally establish the asymptotic distribution of the corresponding test statistics, we conjecture the form of the asymptotic distribution and find by simulation that the test works as well as our test designed for the complex data.

As an application, we test different hypotheses about the number of factors in arbitrage pricing theory. Our test rejects the nulls of zero or only one factors against alternatives of more factors at 5% levels. We also can reject the nulls of 2, 3 and 4 factors at 5% level at least against some alternatives. We cannot reject the null of 5 factors against alternatives of more factors at 5% level.

Although this paper is the first to develop a formal statistical test of hypotheses about the number of factors in a situation when n and T tend to infinity simultaneously, Connor and Korajczyk (1993) were the first to develop another test of similar hypotheses under a sequential asymptotics when first n goes to infinity and then T goes to infinity. The Connor-Korajczyk test is not directly based on the eigenvalue-separation idea which forms the basis for our test, although their test's logic can be traced to this idea. We compare the performance of the Connor-Korajczyk test and our test using simulated data. We find that in samples of various sizes our test has a much less distorted size than the Connor-Korajczyk test. At the same time, the power of our test is similar to that of the Connor-Korajczyk

test.

The logic of our test (and that of Connor and Korajzyk, 1993) differs from the logic of the standard likelihood ratio test (see Anderson (1984), chapter 14), which is based on the fact that the idiosyncratic covariance matrix is diagonal under the classical k-factor structure if the true number of factors is less than or equal to k. The approximate factor structure that we are concerned with in this paper does not require that all the cross-sectional correlation be due to the factors as is the case for the classical factor structure. Since the identification of factors is fundamentally different in the classical and the approximate factor model cases, it is not natural, in general, to compare our test and the classical test.

The rest of the paper is organized as follows. In the next section we state our assumptions and develop the test. Section 3 contains Monte Carlo experiments. Section 4 tests different hypotheses about the number of factors in arbitrage pricing theory. Section 5 concludes. Technical proofs are contained in the Appendix.

2   The number of factors test

We consider a sequence of approximate factor models indexed by n :

$$X^{(n)} = L^{(n)} F^{(n)\prime} + e^{(n)} \tag{1}$$

where $X^{(n)}$ is an $n \times T^{(n)}$ matrix of data; $F^{(n)}$ is a $T^{(n)} \times r$ matrix of $T^{(n)}$ observations of r factors, where r does not depend on n; $L^{(n)}$ is an $n \times r$ deterministic matrix of factor loadings; and $e^{(n)}$ is an $n \times T^{(n)}$ noise matrix with i.i.d. $N_{\mathbb{C}}(0, \Sigma^{(n)})$ columns independent from the elements of $F^{(n)}$. Here $N_{\mathbb{C}}(0, \Sigma^{(n)})$ denotes a complex normal distribution, which is the distribution of a complex random variable whose real and imaginary parts are independent and identically distributed normals $N(0, \Sigma^{(n)}/2)$. We assume that (1) satisfies Assumptions 1 and 2 formulated below. In what follows we will omit superscript (n) over the variables that change with the dimensionality of data to simplify notations.

Assumption 1. *There exist a positive definite matrix B and a positive number b such that $(L'L/n)(F'F/T) \xrightarrow{p} B$ as n and T tend to infinity so that $n/T \to b$.*

Assumption 1 can be thought of as a part of the identification restriction which allows us to identify the systematic component LF' of the data. The rest of the identification restriction is given by the first inequality of Assumption 2 formulated below. If we further normalize factors so that $F'F/T = I_r$, then the assumption can be used to separately identify factors and factor loadings and may be interpreted as a requirement that the effects of the factors per cross-sectional unit measured by $L'L/n$ remain non-trivial as n tends to infinity. In this paper we do not impose any separate restrictions on the convergence or divergence of $L'L/n$ and $F'F/T$ because we are not going to separately identify factors and factor loadings. One consequence of putting no separate restrictions on factors and factor loadings is that much room is left for modeling factors. For example, they are allowed to be deterministic, or random and stationary, or random and non-stationary.

The proportional n and T asymptotics differs from the asymptotic assumptions made in the previous literature. Connor and Korajczyk (1993) develop their test using sequential limit asymptotics when, first, n tends to infinity and, then, T tends to infinity. Bai and Ng (2002) allow n and T to go to infinity without any restrictions on the relative growth rates. The proportional asymptotics allows us to use the machinery of random matrix theory to establish the asymptotic distribution of our test statistics. Note that the limit b may be any positive number so that the asymptotics is consistent with a variety of empirically relevant finite sample situations.

Our second assumption restricts the asymptotic behavior of the matrix of the cross-sectional covariance of the idiosyncratic terms Σ. Let $l_1 \geq ... \geq l_n$ be the eigenvalues of Σ. Denote by H the spectral distribution of Σ, that is $H(\lambda) = 1 - \frac{1}{n}\#\{i \leq n : l_i > \lambda\}$, where $\#\{\cdot\}$ denotes the number of elements in the indicated set. Further, let c be the unique root in $[0, l_1^{-1})$ of the equation $\int (\lambda c/(1 - \lambda c))^2 \, dH(\lambda) = T/n$.

Assumption 2. *As n and T tend to infinity so that $n/T \to b$, $\limsup l_1 < \infty$,*

$\liminf l_n > 0$, *and* $\limsup l_1 c < 1$.

Note that l_1 is the variance of the most variable or, loosely speaking, the most influential weighted average of the idiosyncratic terms. Hence, the first of the three inequalities of Assumption 2 requires that the strongest cumulative idiosyncratic influence on the cross-sectional units remains bounded as the number of the units tends to infinity. This requirement taken together with Assumption 1 constitute a slight modification of the standard (see Chamberlain and Rothschild, 1983) identification restriction imposed on model (1). The second inequality of Assumption 2 requires that, whatever the dimensionality of the data is, there is no multicollinearity among the idiosyncratic terms. The third inequality is crucial for the analytic apparatus developed in El Karoui (2006) for the analysis (which we rely on in this paper) of the spectral distribution of large complex Wishart matrices. In a nutshell, the inequality requires that the right tail of the spectral distribution of Σ is not too thin.

More precisely, it is not difficult to see that for the inequality $\limsup l_1 c < 1$ to hold, it is sufficient that H weakly converges to a distribution H_∞ with the upper boundary of support equal to $\limsup l_1$ and the density *bounded away from zero* in the vicinity of $\limsup l_1$. As is pointed out by El Karoui (2006), such a sufficient condition would hold, for example, for Σ that have a symmetric Toeplitz structure with parameters $a_0, a_1, ...$ such that $\sum k |a_k| < \infty$ and for Σ that have uniformly spaced eigenvalues on a given bounded segment.

In this version of the paper we assume, in addition to Assumptions 1 and 2, that $b < 1$. We will get rid of this additional assumption in the later versions of the paper.

Let $\lambda_1 \geq ... \geq \lambda_n$ be the eigenvalues of ee'/T. Define a centering constant μ and a scaling constant σ so that $\mu c = 1 + (n/T) \int (\lambda c) / (1 - \lambda c) \, dH(\lambda)$ and $(\sigma c)^3 = 1 + (n/T) \int (\lambda c)^3 / (1 - \lambda c)^3 \, dH(\lambda)$, and let $\tilde{\lambda}_i = T^{2/3} \sigma^{-1} (\lambda_i - \mu)$. Further, let W be a $T \times T$ Hermitian matrix with i.i.d. $N_{\mathbb{C}}(0, 1/T)$ lower triangular entries and (independent from them) i.i.d $N(0, 1/T)$ diagonal entries. The collection of such matrices is called Gaussian Unitary Ensemble. It plays an important role in random matrix theory (see Mehta, 2004). Let $d_1 \geq ... \geq d_T$ be the eigenvalues of W, and define $\tilde{d}_i = T^{2/3}(d_i - 2)$. We, first, establish

the following

Theorem 1. *Let Assumptions 1 and 2 hold and let $b < 1$. Then, as n and T go to infinity so that $n/T \to b$, for any finite positive integer j, the limiting joint distribution of $\tilde{\lambda}_1, ..., \tilde{\lambda}_j$ is equal to the limiting joint distribution of $\tilde{d}_1, ..., \tilde{d}_j$ described by Tracy and Widom (1994).*

Theorem 1 is, essentially, due to El Karoui (2006). Although he proves the theorem only for the case of $j = 1$, his derivations imply that it holds for any fixed positive integer j. An argument which establishes this fact is the same as that in Soshnikov (2002) which shows that Johnstone's (2001) result stated for the largest eigenvalue of the sample covariance matrix of uncorrelated data easily generalizes to the several largest eigenvalues. We will sketch the argument in the future version of the paper which will deal with the case $b \geq 1$.

Tracy and Widom (1994) studied the asymptotic distribution of a few largest eigenvalues of matrices from the Gaussian Unitary Ensemble when the dimensionality of the matrices tend to infinity. They showed that under appropriate centering and scaling the asymptotic distribution exists and can be expressed in terms of a solution of a system of partial differential equations. The system simplifies to a single ordinary differential equation $q''(s) = sq(s) + 2q^3(s)$ when we are interested in the asymptotic distribution of the largest eigenvalue only. In such a case, the asymptotic distribution is equal to $F(x) \equiv \exp\left(-\int_x^\infty (x - s)q^2(s)ds\right)$, where $q(s)$ is the solution of the above ordinary differential equation which is asymptotically equivalent to the Airy function $Ai(s)$ as $s \to \infty$.[1] We describe the Tracy-Widom distribution in more detail in the Monte Carlo section of our paper.

Our next theorem shows the equivalence of the joint asymptotic distribution of unobservable eigenvalues of ee'/T and that of observable eigenvalues of XX'/T. Let $\gamma_1 \geq ... \geq \gamma_n$ be the eigenvalues of the sample covariance matrix XX'/T of the data, and let $\tilde{\gamma}_i = T^{2/3}\sigma^{-1}(\gamma_i - \mu)$, where μ and σ are as defined above. Then, we have:

---

[1] For the definition and properties of the Airy function see Olver (1974).

Theorem 2. *Let Assumptions 1 and 2 hold. Then, as n and T go to infinity so that $n/T \to b$, for any finite positive integer j, the limiting joint distribution of $\tilde{\gamma}_{r+1}, ..., \tilde{\gamma}_{r+j}$ is equal to the limiting joint distribution of $\tilde{\lambda}_1, ..., \tilde{\lambda}_j$.*

A proof of the theorem is given in the Appendix.

Theorem 2 is the basis of our test of the number of factors. To test a hypothesis that the number of factors is k vs. an alternative that the number of factors is larger than k but smaller than $k_{\max} + 1$, we form a test statistics $\left(\gamma_{k+1} - \gamma_{k_{\max}+1}\right) / \left(\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}\right)$. Note that the test statistics is equal to the ratio of the normalized and centered eigenvalues $\left(\tilde{\gamma}_{k+1} - \tilde{\gamma}_{k_{\max}+1}\right) / \left(\tilde{\gamma}_{k_{\max}+1} - \tilde{\gamma}_{k_{\max}+2}\right)$. Therefore, by Theorem 2, the asymptotic distribution of our test statistics under the null of k factors is the same as the asymptotic distribution of the ratio $\left(\tilde{\lambda}_1 - \tilde{\lambda}_{k_{\max}-k+1}\right) / \left(\tilde{\lambda}_{k_{\max}-k+1} - \tilde{\lambda}_{k_{\max}-k+2}\right)$. By Theorem 1 and the continuous mapping theorem, the asymptotic distribution of the latter ratio is equal to the distribution of $(x_1 - x_{k_{\max}-k+1}) / (x_{k_{\max}-k+1} - x_{k_{\max}-k+2})$, where $x_1, ..., x_j$ are random variables with the Tracy-Widom joint distribution. The critical values of the test statistics can therefore be tabulated based on the knowledge of the Tracy-Widom law. We do such a tabulation in the Monte Carlo section of the paper. In contrast, when the alternative hypothesis is true, γ_{k+1} rises proportionally to n while $\gamma_{k_{\max}+1}$ and $\gamma_{k_{\max}+2}$ remain bounded. Therefore, under the alternative hypothesis our test statistics explodes. Theorem 3 summarizes the properties of our test.

Theorem 3. *Let Assumptions 1 and 2 hold and let $b < 1$. Then, if the true number of factors is equal to k, the distribution of the ratio $\left(\gamma_{k+1} - \gamma_{k_{\max}+1}\right) / \left(\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}\right)$ converges to the distribution of $(x_1 - x_{k_{\max}-k+1}) / (x_{k_{\max}-k+1} - x_{k_{\max}-k+2})$, where $x_1, ..., x_j$ are random variables with the Tracy-Widom joint distribution. The convergence takes place as n and T go to infinity so that $n/T \to b$. In contrast, when the true number of factors is larger than k but smaller than $k_{\max} + 1$, the ratio $\left(\gamma_{k+1} - \gamma_{k_{\max}+1}\right) / \left(\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}\right)$ diverges in probability to infinity.*

Theorem 3 is a simple consequence of Theorem 2.

3   Monte Carlo Study

In this section we use Monte Carlo simulations to describe in some detail the Tracy-Widom distribution, to tabulate the critical values of our test, and to study its finite sample properties. We approximate the Tracy-Widom distribution by the distribution of a few 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. The left panel of Figure 1 shows the empirical distribution function of the largest eigenvalue centered by 2 and scaled by $T^{2/3} = 1000^{2/3}$. The right panel of Figure 1 shows a scatterplot of 30,000 observations of the difference between the first and the second eigenvalues vs. the difference between the second and the third eigenvalues (both differences scaled by $T^{2/3}$).

Tracy and Widom (2002) report that the mean of their univariate distribution[2] is about -1.77, the standard deviation is close to 0.90, the skewness is slightly larger than 0.22, and the kurtosis is around 0.09. These characteristics are consistent with the left panel of Figure 1. The right panel of Figure 1 shows that it is not unlikely that $x_1 - x_2$ is substantially larger than $x_2 - x_3$, where x_1, x_2, and x_3 have joint Tracy-Widom distribution. This observation suggests that ad hoc methods of the determination of the number of factors based on visual inspection of the eigenvalues of the sample covariance matrix, and their separation into a group of "large" and a group of "small" eigenvalues may be misleading. It may happen, for example, that even though data have no factors, the first eigenvalue of the sample covariance matrix is substantially larger than the second one, while the second eigenvalue is not much different from the third one.

Table 1 contains approximate percentiles of the distributions of random variables $(x_1 - x_{j-1}) / (x_{j-1} - x_j)$ for $j = 3, 4, ..., 10$, where $x_1, ..., x_j$ have the joint Tracy-Widom distribution. The approximate percentiles were obtained as the Jackknifed sectioning es-

---

[2] They denote the distribution as F_2 to distinguish it from the distributions F_1 and F_4 that correspond to the limiting distributions of the largest eigenvalues of matrices from the so called Gaussian Orthogonal and Simplectic Ensenbles, respectively.
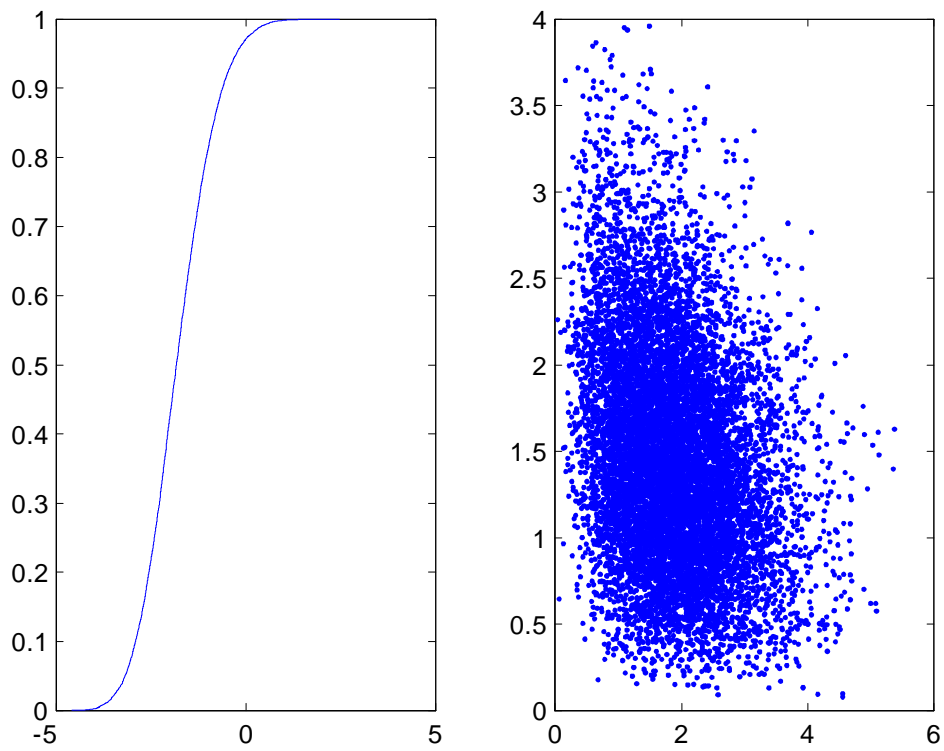
Figure 1: Left panel: a univariate Tracy-Widom distribution. Right panel: 30,000 draws from the joint distribution of $x_1 - x_2$ and $x_2 - x_3$, where x_1, x_2, x_3 have the joint Tracy-Widom distribution

	$j = k_{\max} - k + 2$							
	3	4	5	6	7	8	9	10
99%	8.81 (7.77,9.84)	15.10 (13.93,16.28)	21.30 (19.38,23.22)	28.10 (23.73,32.47)	34.61 (30.26,38.96)	42.57 (37.87,47.28)	45.84 (43.54,48.13)	54.58 (51.97,57.17)
95%	4.52 (4.43,4.61)	8.08 (7.89,8.28)	11.90 (11.67,12.14)	15.66 (15.42,15.90)	19.09 (18.69,19.49)	22.90 (22.18,23.61)	26.72 (26.22,27.21)	30.54 (30.03,31.04)
90%	3.33 (3.28,3.38)	6.18 (6.09,6.27)	9.02 (8.91,9.12)	12.04 (11.81,12.27)	14.72 (14.62,14.83)	17.81 (17.45,18.18)	20.75 (20.48,21.02)	23.68 (23.41,23.96)

Table 1: Approximate percentiles of the test statistics for the tests of k factors vs. an alternative of more than k but less than kmax factors

timator (reference here) of the corresponding percentiles of the distribution of the ratio $(y_1 - y_{j-1}) / (y_{j-1} - y_j)$, where $y_1, ..., y_j$ are the first j eigenvalues of 1000×1000 matrices from the Gaussian Unitary Ensemble. The sectioning estimator uses 5 equal-length sections of 10,000 i.i.d. draws of $(y_1 - y_{j-1}) / (y_{j-1} - y_j)$. The 90% confidence intervals for the true percentiles of the distribution of $(y_1 - y_{j-1}) / (y_{j-1} - y_j)$ are given in the parenthesis.

The confidence intervals reported in Table 1 underestimate the uncertainty about the true percentiles of the asymptotic distribution of our test statistics because $y_1, ..., y_j$ are distributed according the joint Tracy-Widom law only asymptotically, as the dimensionality of the matrix of which y_i's are eigenvalues increases to infinity. We hope that the amount of the additional uncertainty due to the fact that we used 1000×1000 matrices to simulate y_i's is small and do not try to improve the uncertainty estimates. Our hope is supported by El Karoui's (2006a) finding that the cumulative distribution function of y_1 converges to that of the univariate Tracy-Widom law at the very fast rate of $n^{-2/3}$, where n is the dimensionality of the corresponding matrix.

Table 1 can be used to find critical values of our test that the number of factors is k vs. an alternative that it is more than k but less than $k_{\max} + 1$. According to Theorem 3, to determine, say, 5% critical value of the test, we need to consult Table 1 and find the 95-th percentile of the distribution of $(x_1 - x_{j-1}) / (x_{j-1} - x_j)$, where $j = k_{\max} - k + 2$. For example, the 5% critical value of our test of the hypothesis that there are 3 factors vs. an alternative that there are more than 3 factors but less than 7 factors is equal to 11.90.

To study the finite sample properties of our test, we simulated 1000 data sets having

	$k_{\max}$							
	4	5	6	7	8	9	10	11
99%	0.008	0.010	0.011	0.009	0.012	0.011	0.008	0.013
95%	0.044	0.058	0.050	0.056	0.048	0.058	0.053	0.070
90%	0.090	0.112	0.111	0.102	0.097	0.103	0.125	0.126

Table 2: Empirical size of the tests of 3 factors vs. different alternatives, n=T=100

factor structure $X = \Lambda F + e$, where Λ is an $n \times r$ matrix with i.i.d. standard normal entries, F is a $r \times \tilde{T}$ matrix with standard normal entries (hence, the true number of factors is r), and e is an $n \times \tilde{T}$ matrix with i.i.d. $N(0, \Sigma)$ columns, where $\Sigma_{ij} = \rho^{|i-j|}$. We considered three different choices of n and $\tilde{T}$: $\left(n, \tilde{T}\right) = (100, 200), (50, 100)$, and $(25, 50)$; ten different choices of $\rho : \rho = 0, 0.1, 0.2, ..., 0.9$; and nine different choices of $r : r = 3, 4, ..., 11$..To create complex-valued data sets we added the first $\tilde{T}/2$ columns of X and $\sqrt{-1}$ times the last $\tilde{T}/2$ columns of X. Hence, the dimensionality of our complex data sets were $n \times T$, where $T = \tilde{T}/2$. Using the simulated data we tested a hypothesis that there are 3 factors vs. alternatives that there are more than 3, but less than 5, 6,..., 12 factors.

Table 2 reports the empirical size of the performed tests for $n = 100, \tilde{T} = 200$, and $\rho = 0$. The last 3 rows of the table correspond to the tests with theoretical size equal to 1%, 5%, and 10%. Different columns of the table correspond to the tests that share the null of 3 factors but have different alternatives indexed by $k_{\max}$. We see that the size distortions due to the small sample size are small.

Figure 2 shows contour plots of the empirical size and power of our tests (with theoretical size 5%) for different choices of ρ and $\left(n, \tilde{T}\right)$. The contour plots are drawn in the space of ρ (horizontal axis) and $k_{\max} - 1$ (vertical axis). The left panel of Figure 2 shows the tests' size, and the right panel shows the tests' power when the true number of factors is equal to $k_{\max}$. The first row of the figure corresponds to the sample size $n = 100$, $\tilde{T} = 200$; the second row corresponds to $n = 50$, $\tilde{T} = 100$, and the third row corresponds to $n = 25$, $\tilde{T} = 50$. We see that the size of our tests deteriorates when the amount of dependence in the idiosyncratic terms rises. This is especially noticeable for tests corresponding to large $k_{\max}$ and for $n = 25$.
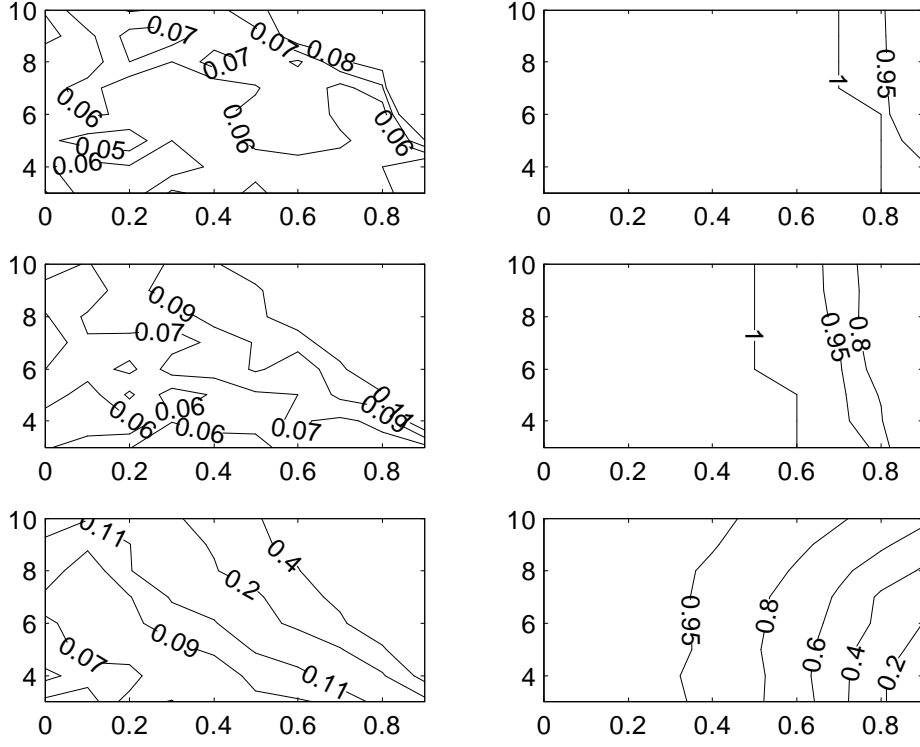
Figure 2: Contour plots of the empirical size and power of the tests of 3 factors vs. alternatives of less than $k_{\max} + 1$ factors. Horizontal axis: ρ. Vertical axis: $k_{\max} - 1$. Left panel: size. Right panel: power when the true number of fators is $k_{\max}$. First row: $n = T = 100$, second row: $n = T = 50$, third row: $n = T = 25$.

For $n \geq 50$ and $k_{\max} = 4$, the size distortions are small for all ρ. For $n \geq 50$ and relatively large $k_{\max}$, the size distortions are larger. The power of the tests is very good for $n \geq 50$ and virtually all ρ and $k_{\max}$, but becomes small for $n = 25$, large ρ, and small $k_{\max}$.

The fact that the power of our test deteriorates as the amount of the idiosyncratic dependence rises and the sample size falls accords with the following intuition. When the true number of factors is $k_{\max}$, the eigenvalue $\gamma_{k_{\max}+1}$ measures the largest contribution of the idiosyncratic terms into the data's variance. As the amount of the dependence in the idiosyncratic terms rises, this contribution increases so that it may become comparable to the factors' shares of the data's variance measured by $\gamma_1, ..., \gamma_{k_{\max}}$. Furthermore, since factors

14

are common to all idiosyncratic units, their share in the data's variance rises proportionally to the number of the cross-sectional units. Therefore, eigenvalues $\gamma_1, ..., \gamma_{k_{\max}}$ would be relatively small when the sample size is small and the gap between γ_k and $\gamma_{k_{\max}+1}$ can become small even for relatively small amount of the idiosyncratic dependence. Hence, for a large amount of the idiosyncratic dependence or for small sample sizes, the numerator of our test statistics $\left(\gamma_{k+1} - \gamma_{k_{\max}+1} \right) / \left(\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2} \right)$ may become relatively small even when the true number of factors is $k_{\max}$. This explains deterioration of the power of our test when the amount of the idiosyncratic dependence rises or the sample size falls.

To check the robustness of our test to violations of Gaussianity and time independence of the idiosyncratic terms, we perform the following three experiments. First, we use an AR(1) process $e_{i,t} = \rho e_{i-1,t} + \varepsilon_{it}$, where ε_{it} are i.i.d. random variables having Student's t distribution with 5 degrees of freedom,[3] to generate relatively heavy-tailed idiosyncratic terms. Second, we use the same AR(1) process but assume that ε_{it} are i.i.d. random variables having $\chi^2(1)$ distribution centered so that its mean is equal to zero. Hence, in the second experiment the idiosyncratic terms are skewed to the right. Finally, we simulate idiosyncratic terms that satisfy a version of the constant conditional correlation multivariate GARCH model of Bollerslev (1990): $e_{i,t} = \sqrt{h_{i,t}} u_{i,t}$, $u_{i,t} = \rho u_{i-1,t} + \varepsilon_{it}$, $\varepsilon_{i,t}$ are i.i.d. $N\left(0, 1 - \rho^2 \right)$, and $h_{it} = 0.1 + 0.8 h_{i,t-1} + 0.1 u_{i,t-1}^2$. In the first two experiments, we simulate data using the idiosyncratic terms after normalizing them so that they have unit sample variance. In the last experiment, the theoretical unconditional variance of e_{it} is equal to 1, so we do not normalize the idiosyncratic terms. An equivalent of Figure 2 for the simulated data looks very similar to Figure 2 for all three experiments and it is not reported here. Instead, Figure 3 reports the size and the power of the test of the null of 3 factors vs. the alternative that the number of factors is 4 for the case $N = 50$, $\tilde{T} = 100$.

The left panel of Figure 3 shows the size of the test and the right panel shows the

---

[3] We chose to consider Student's t distribution with 5 degrees of freedom because we wanted the distribution to have 4 moments, which is a necessary condition for many Large Random Matrix theory results to hold (see Bai, 1999).
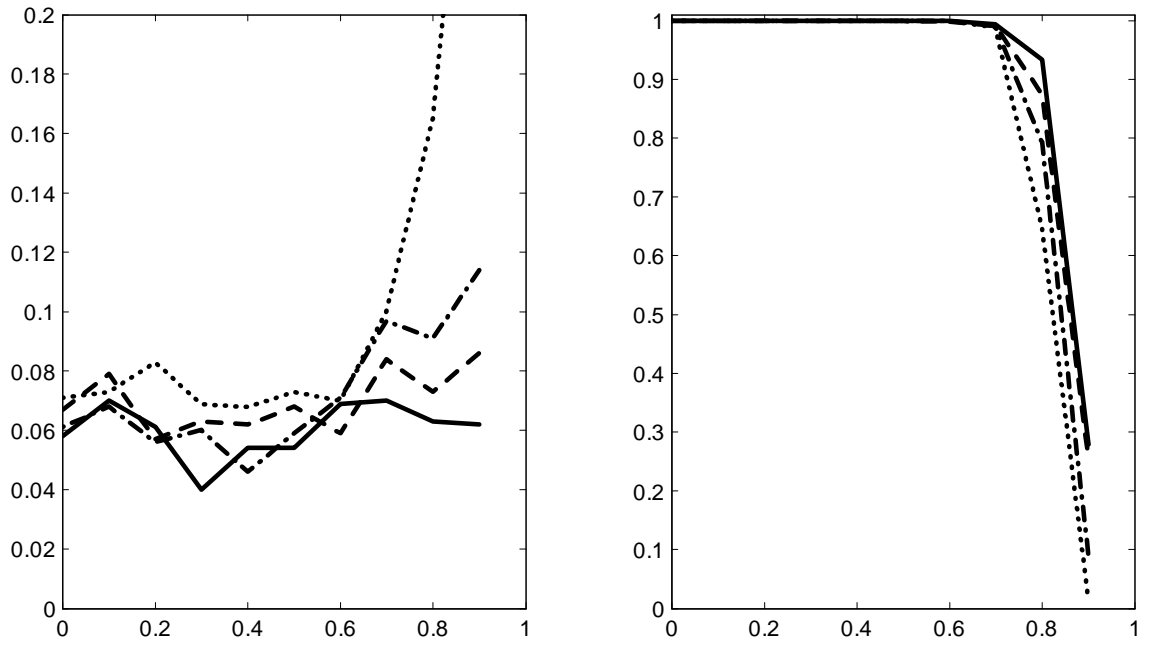
Figure 3: Size and power of our test of 3 factors vs. 4 factors. Left panel: size. Right panel: power. Horizontal axis: ρ. $N = 50, \tilde{T} = 100$. Solid line: no violations of the assumptions. Dashed line: multivarite GARCH model for the idiosyncratic terms. Dash-dot line: skewed idiosyncratic terms. Dotted line: heavy tails.

power. The solid line corresponds to the benchmark simulations when no assumptions of the paper were violated. The dashed line corresponds to the idiosyncratic terms satisfying the multivariate GARCH model. The dash-dot line corresponds to the idiosyncratic terms heavily skewed to the right. The dotted line corresponds to the idiosyncratic terms which have heavy tails. We see that the small sample properties of our test are reasonably robust to different violations of our assumptions. The worst distortion of the size and the power happens when heavy tails are assumed for the idiosyncratic process. However, the distortions do not become prohibitively large until the amount of the cross-sectional correlation in the idiosyncratic terms measured by ρ becomes as large as 0.7 even in the heavy-tail case.

Our next task is to compare the finite sample properties of our test and that proposed by Connor and Korajczyk (1993). The procedure of the Connor-Korajczyk test of a null of k factors vs. an alternative of $k+1$ factors is as follows. First, a k-factor model and a $k+1$-factor model are estimated from the data, say, by principle components estimator. Then, the average over the cross-sectional units or the squared residuals for each of the models and each time period are obtained. Let us denote these averages as $\mu_{1,t}$ and $\mu_{2,t}$, where the first subscript is equal to 1 if the average corresponds to the k-factor model and it is equal to 2 if the average corresponds to the $k+1$-factor model. After that, a vector Δ with components $\mu_{1,2\tau-1} - \mu_{2,2\tau}$, $\tau = 1,...,T/2$ is formed. Connor and Korajzcyk (1993) show that, under the null, when n tends to infinity and T remains fixed, the asymptotic joint distribution of the components of Δ is zero mean Gaussian with some unspecified covariance matrix. In contrast, under the alternative, we would expect the residuals from the $k+1$-factor model to be much smaller than those from the k-factor model. Therefore, the components of Δ should be large. Hence, Connor and Korajzcyk propose as their test statistics a ratio of the mean of the components of Δ and the heteroskedasticity and autocorrelation robust estimate of its standard error. Under the null, such a ratio should be asymptotically distributed as a standard normal random variable as first n goes to infinity and then T goes to infinity.

We implemented the Connor-Korajzcyk test of 3 factors relative to the alternative of 4

factors using the same 1000 simulations of the real data that we used to study the finite sample properties of our test. We focused on those simulations for which the true number of factors equals 3 or 4. Figure 4 shows the size and the power of our test (thick lines) and that of the Connor-Korajzcyk test (thin lines) for different amounts of the idiosyncratic dependence. The left panel of the figure shows the size and the right panel of the figure shows the power of the tests. The solid lines correspond to the sample size $n = 100$, $\tilde{T} = 200$; the solid lines with dot markers on them correspond to the sample size $n = 50$, $\tilde{T} = 100$; and the dashed lines correspond to the smallest sample size $n = 25$, $\tilde{T} = 50$. We see that the Connor-Korajzcyk test has much too large a size (the theoretical size is equal to 5%) for all ρ (horizontal axis). The size of the test becomes extremely distorted even for moderate amounts of the idiosyncratic dependence. The empirical size of our test behaves strikingly better than that of the Connor-Korajzcyk test. The power of both tests is very good for small amounts of the idiosyncratic dependence. For relatively large amounts of the idiosyncratic dependence and for relatively small sample sizes, the power of our test becomes much worse than that of the Connor-Korajzcyk test. However, given the huge size distortions of the latter, such a power advantage cannot exploited. Overall, the properties of our test are much better than those of the Connor-Korajzcyk at least in our Monte Carlo experiments.

Before we turn to an application of our test, we would like formulate a conjecture which is an analog Theorem 1 for the case of real data.

Conjecture 1: *Assume that the columns of the matrix of idiosyncratic terms $e^{(n)}$ are multivariate real random variables with covariance matrix Σ. Let Assumptions 1 and 2 hold. Then, as n and T go to infinity so that $n/T \to b$, for any finite positive integer j, the limiting joint distribution of $\tilde{\lambda}_1, ..., \tilde{\lambda}_j$ is equal to the limiting joint distribution of $T^{2/3} (d_i - 2)$, $i = 1, ..., j$, where d_i is the i-th largest eigenvalue of a matrix from the Gaussian Orthogonal Ensemble described by Tracy and Widom (1996) (denoted as TW1)..*

Conjecture 1 has been supported by numerical simulations in El Karoui (2006). If the conjecture is correct, then we have:
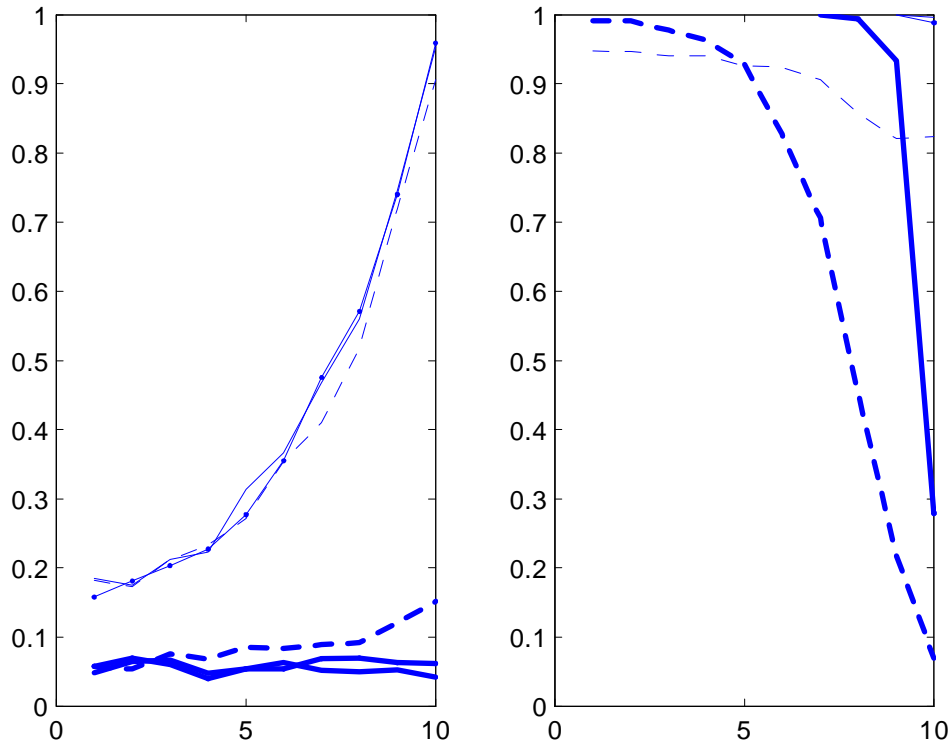
Figure 4: Empirical size and power of our test of 3 vs. 4 factors (thick lines) and the Connor-Korajzcyk test of 3 vs. 4 factors (thin lines). Left panel: size. Right panel: power. Solid lines: $n = 100$, $\tilde{T} = 200$. Solid lines with dot markers: $n = 50$, $\tilde{T} = 100$. Dashed lines: $n = 25$, $\tilde{T} = 50$. Horizontla axis: ρ.

	$j = k_{\max} - k + 2$							
	3	4	5	6	7	8	9	10
99%	16.56 (14.79,18.33)	28.75 (25.97,31.53)	41.57 (36.02,47.12)	55.07 (48.49,61.65)	67.53 (61.28,73.78)	79.13 (72.87,85.40)	91.90 (74.13,109.67)	106.01 (89.03,122.99)
95%	6.89 (6.64,71.5)	12.41 (12.07,12.75)	18.16 (17.49,18.83)	23.99 (23.26,24.73)	29.41 (28.33,30.50)	35.05 (34.30,35.79)	39.89 (38.27,41.52)	47.35 (45.49,49.21)
90%	4.54 (4.45,4.63)	8.53 (8.44,8.62)	12.46 (12.13,12.78)	16.40 (16.18,16.62)	20.29 (20.03,20.55)	24.34 (23.98,24.71)	27.42 (26.37,28.46)	32.62 (31.87,33.37)

Table 3: Approximate percentiles of the test statistics for the tests of k factors vs. an alternative of more than k but less than kmax factors. Real data.

Conjecture 2. *Assume that the columns of the matrix of idiosyncratic terms $e^{(n)}$ are multivariate real random variables with covariance matrix Σ. Let Assumptions 1 and 2 hold. Then, if the true number of factors is equal to k, the distribution of the ratio $\left(\gamma_{k+1} - \gamma_{k_{\max}+1}\right)/\left(\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}\right)$ converges to the distribution of $(x_1 - x_{k_{\max}-k+1})/(x_{k_{\max}-k+1} - x_{k_{\max}-k+2})$, where $x_1, ..., x_j$ are random variables with the Tracy-Widom (TW1) joint distribution. The convergence takes place as n and T go to infinity so that $n/T \to b$. In contrast, when the true number of factors is larger than k but smaller than $k_{\max} + 1$, the ratio $\left(\gamma_{k+1} - \gamma_{k_{\max}+1}\right)/\left(\gamma_{k_{\max}+1} - \gamma_{k_{\max}+2}\right)$ diverges in probability to infinity.*

Table 3 below is the analog of Table 1 for the test based on Conjecture 2.

4 The number of factors in asset returns

As an application of our test, we would like to test different hypotheses about the number of pervasive factors driving stock returns. There is a large amount of controversy about this number in the literature. For example, Connor and Korajcyk (1993) find evidence for between one and six pervasive factors in the stock returns. Trzcinka (1986) finds some support to the existence of five pervasive factors. Five seems also to be a preferred number for Roll and Ross (1980) and Reinganum (1981). A study by Brown and Weinstein (1983) also suggested that the number of factors is unlikely to be greater than five.[4] Huang and Jo (1995), Bai and Ng (2002), and Onatski (2005) identify only two common factors

---

[4]Dhrymes, Friend, and Gultekin (1984) find that the estimated number of factors grows with the sample size. However, their setting was the classical factor model as opposed to the approximate factor model.

To test hypotheses about the number of factors in stock returns we use CRSP data on monthly returns of 171 company for a period from January 1960 to December 2005. Our 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 compute the excess returns we used monthly returns on 6-month Treasury bills provided by the Board of Governors of the Federal Reserve System.

The cross-sectional dimensionality of our data is $n = 171$, and the time series dimensionality is $\tilde{T} = 552$. To get a complex valued data set, we divide the real-valued data set into two periods: the first containing all observations from January 1960 to December 1982, and the second containing all observations from January 1983 to December 2005. Then we add the data from the first time period and square root from -1 times the data from the second period. The dimensionality of the obtained complex-valued data set is $n = 171$ and $T = 276$.

Using the obtained data, we test hypotheses that the number of factors is 0,1,...,7 vs. alternatives that the number of factors is more than the null-hypothesis number but no more than up to 8 factors. the resulting test statistics are reported in Table 4. The numbers in parentheses given below the values of the test statistics are approximate p-values of the test. The approximate p-values were computed as the proportion of 30,000 simulated draws from the null distribution that fall above the corresponding test statistics.

We see that the null hypothesis of no factors vs. any of the alternatives of the form: positive number of factors no larger than $k_{\max}$ is overwhelmingly rejected by the data. The null of one factor vs. more factors is rejected at 5% level for all corresponding alternatives except the alternative that the number of factors is larger than 1 but no larger than 4. The null of 2 factors is not rejected at 5% level for all corresponding alternatives except the alternative that the number of factors is more than 2 but no larger than 5. The situation is the same for the null of 3 and 4 factors. The nulls of 5, 6, and 7 factors are not rejected by the data even at at 29% level.

To interpret these results it is useful to analyze the largest eigenvalues of the sample

21

$k_{\max}$	H_0 number of factors							
	0	1	2	3	4	5	6	7
1	9.72 (0.008)							
2	75.19 (0.000)	7.01 (0.012)						
3	126.71 (0.000)	13.33 (0.010)	1.66 (0.286)					
4	46.88 (0.002)	5.26 (0.279)	0.98 (0.944)	0.37 (0.923)				
5	592.27 (0.000)	77.44 (0.000)	24.46 (0.005)	16.91 (0.004)	12.37 (0.002)			
6	204.54 (0.000)	27.04 (0.016)	8.78 (0.155)	6.18 (0.171)	4.61 (0.132)	0.34 (0.928)		
7	257.62 (0.000)	35.15 (0.011)	12.57 (0.121)	8.99 (0.138)	7.03 (0.120)	1.69 (0.682)	1.25 (0.401)	
8	349.00 (0.000)	48.78 (0.007)	17.89 (0.072)	13.49 (0.087)	10.84 (0.080)	3.62 (0.449)	3.04 (0.295)	1.35 (0.360)

Table 4: Value of test statistics and approximate p-values for the tests of the number of factors in the second row vs. alternative of more factors but less than or equal to the number in the first column

covariance matrix of the data. Note that the eigenvalues can be interpreted as reductions in the mean square error which result from using an extra principal component to model systematic part of the data. The eigenvalues turned out to be (in percentage units relative to the largest eigenvalue):

$$100.00; 18.40; 10.00; 8.80; 8.09; 6.12; 5.96; 5.51; 5.14; 4.87; \text{ etc.}$$

The rejection of the null of no factors means that the gap between the explanatory power of the first principal component and any consequent principal components is too large relative to the gaps between the explanatory power of more distant components. It is too large in the sense that such a gap is not consistent with the null that all principal components, including the first one, represent just idiosyncratic influences. The first principal component must, therefore, contain a systematic or pervasive influence on the cross-sectional units. Similarly, the gap between the explanatory power of the second principal component and virtually all consequent principal components is too large (in relative terms) to be consistent with the idiosyncratic nature of the second component.

As to the nulls of 2, 3 and 4 factors, they are rejected at 5% level only vs. an alternative of no more than 5 factors. Technically, this happens because the gaps between the explanatory power of, say, the second principal component and that of the 6th, 7th, 8th, and 9th principal components do not seem large (at 5% level) relative to all gaps but the gap between the explanatory power of the 6th and the 7th components. An important question arises whether we should consider the closeness of the 6th and 7th eigenvalues as a fluke in the data, and, therefore, "accept" the null of only two factors, or we should interpret our results as suggesting that there are likely to be five factors.

We favor the latter interpretation. It is because, although we cannot reject the nulls of 2, 3 and 4 factors vs. alternatives of more factors but less than 7, 8, and 9 factors at 5% level, we reject these nulls vs. the alternatives at about 17% level, which is not a high level. Had the relative closeness of the 6th and 7th eigenvalues been just a fluke, we would not expect the p-values of the tests against the alternatives of less than 7, 8, and 9 factors to be consistently "at the low end". Note in this context, that the nulls of 5, 6 and 7 factors cannot be rejected with the corresponding p-values being "everywhere" in the range from 0.295 to 0.928.

To check our finding that there likely be 5 factors driving the stock returns, we run the real data tests (which are conjectured but not proven to be correct procedures) on the untransformed real data. We get very similar results to those reported in Table 4. To check robustness of the finding to the choice of the data, we run our complex-data test on the CRSP data for 1148 stocks traded on the NYSE, AMEX, and NASDAQ during the period from January 1983 to December 2003, which we use in our previous work (see Onatski (2005)). For these data, the nulls of zero and one factors are, again, overwhelmingly rejected. The nulls of 2, 3, and 4 factors are rejected (at 5% level) again only vs. the alternatives of no more than 5 factors. This time, however, the p-values corresponding the alternatives of less than 7, 8, and 9 factors are not consistently "small". They "span" a range between 0.225 and 0.885. Moreover, when we run the real-data test on the new data, we still overwhelmingly

reject hypotheses of zero and one factors, but no longer reject the hypotheses of 2, 3, etc. factors at any reasonable critical levels against any alternatives (the smallest p-value being 0.368 for the test of two factors vs. the alternative of more than two but no more than four factors).

The partial robustness of the our finding that there likely be 5 factors in the stock returns may be due either to the fact that much more stocks were included into the second data set, or to a possibility that the number of factors driving the stock returns decreased over time so that the second data set which spans more recent data period does not support the 5-factor hypothesis as strongly as the first data set. To check the latter possibility, we split the original data set into two periods and then run our tests separately for the complex data constructed from the first and the second periods. For both periods the nulls of zero and one factors are rejected, as usual. However, the nulls of 2 and 3 factors are now rejected at about 5% level against two alternatives: no more than 4 factors and no more than 7 factors for both time periods. The null of 4 factors is not rejected for any conventional critical levels against any of the studied alternatives for both time periods. Hence, the 5-factor phenomenon does not hold in both subsamples.

To summarize, we find that our tests overwhelmingly reject the hypotheses that there are either no factors or only one factor driving stock returns. There is some evidence that the number of factors may be five, but this evidence is only partially robust with respect to the choice of the stocks to be included to the data set and with respect to the different choice of time periods.

5   Conclusion

The new test shows good finite sample properties and uses simultaneous large n- large T asymptotics. It strongly outperforms the Connor-Korajczyk (1993) test for a variety of finite sample situations. The biggest weakness of the test is that it is developed for the situation

when the data are Gaussian and independent over time. The Monte Carlo experiments show, however, that the test is robust with respect to violations of the Gaussianity and the independence assumptions. We apply the test to test different hypotheses about the number of pervasive factors driving US stock returns. Our test rejects the nulls of zero or only one factors against alternatives of more factors at 5% levels. We also can reject the nulls of 2, 3 and 4 factors at 5% level at least against some alternatives. We cannot reject the null of 5 factors against alternatives of more factors at 5% level. The 5-factor finding turns out to be only partially robust with respect to different choices of the time interval and stocks to be included into the dataset. The rejection of zero and 1 factor hypotheses is very robust.

6 Appendix

Proof of Theorem 2:

We will first formulate and prove a lemma, which our proof of Theorem 2 will be based upon. Let $A^{(1)}$ be a symmetric non-negative definite $n \times n$ matrix and A be an $n \times n$ diagonal matrix of the form $A = \mathrm{diag}\,(a_1, ..., a_k, 0, 0, ..., 0)$, $a_1 \geq a_2 \geq ... \geq a_k > 0$. Note that A and $A^{(1)}$ can be interpreted as matrix representations of linear operators acting in the space R^n. For any linear bounded operator on R^n, B, we define its norm as $\|B\| = (\max \mathrm{eval}\,(B^*B))^{1/2}$, which is the operator norm induced by the standard Euclidean norm of vectors in R^n. We denote the j-th largest by absolute value eigenvalue of B as $\mu_j(B)$, and the j-th largest by absolute value eigenvalue of B restricted to its invariant subspace M as $\mu_i(B|M)$. Let M_0 be the invariant subspace of A corresponding to the eigenvalue 0 and let P_0 be the orthogonal projection on M_0. We have the following

Lemma 1: *Let* $A(\varkappa) = A + \varkappa A^{(1)}$ *and let* $r_0 = a_k/2$. *For real* $\varkappa$ *such that* $0 < \varkappa < r_0/\left\|A^{(1)}\right\|$ *we have:*

$$\left|\mu_{k+1}\left(A\left(\varkappa\right)\right) - \varkappa\mu_1\left(P_0 A^{(1)} P_0 | M_0\right)\right| \leq 3r_0 \frac{|\varkappa|^2 \left\|A^{(1)}\right\|^2}{\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^2}$$

Proof of Lemma 1: Let $R(z, \varkappa) = (A(\varkappa) - zI_n)^{-1}$ be the resolvent of $A(\varkappa)$ defined for all complex z not equal to any of the eigenvalues of $A(\varkappa)$. Then $R(z, 0)$ is the resolvent of A. We will denote $R(z, 0)$ as $R(z)$. Let Γ be a positively oriented circle in the complex plain with center at 0 and radius r_0. According to Kato (1980), pages 67-68, $R(z, \varkappa)$ can be expanded as $R(z, \varkappa) = R(z) + \sum_{j=1}^{\infty} (-\varkappa)^j R(z) \left(A^{(1)} R(z)\right)^j$, where the sum on the right hand side is uniformly convergent on Γ for $|\varkappa| < \min_{z \in \Gamma} \left(\left\|A^{(1)}\right\| \|R(z)\|\right)^{-1} = r_0/\left\|A^{(1)}\right\|$, where the last equality follows from the fact that

$$\|R(z)\| = r_0^{-1} \tag{2}$$

for any $z \in \Gamma$. Furthermore, for $|\varkappa| < r_0/\left\|A^{(1)}\right\|$, there are exactly $n - k$ eigenvalues (counted as many times as their algebraic multiplicity) inside the circle Γ. We will call these eigenvalues 0-group eigenvalues and denote them as $a^{(1)}(\varkappa) \geq ... \geq a^{(n-k)}(\varkappa)$.

Denote the direct sum of the invariant subspaces of $A(\varkappa)$ corresponding to the 0-group eigenvalues as $M_0(\varkappa)$. Let P_0 be the orthogonal projection on M_0 and $P_0(\varkappa)$ be a projection on $M_0(\varkappa)$ defined as $P_0(\varkappa) = -\frac{1}{2\pi i} \int_{\Gamma} R(z, \varkappa)\, dz$. We have:

$$P_0(\varkappa) = P_0 - \frac{1}{2\pi i} \sum_{j=1}^{\infty} (-\varkappa)^j \int_{\Gamma} R(z) \left(A^{(1)} R(z)\right)^j dz \tag{3}$$

where the series absolutely converges for $|\varkappa| < r_0/\left\|A^{(1)}\right\|$ (see Kato (1980), pages 75-76). Equations (3) and (2) imply, that

$$\|P_0(\varkappa) - P_0\| \leq \sum_{j=1}^{\infty} |\varkappa|^j \frac{\left\|A^{(1)}\right\|^j}{r_0^j} = \frac{|\varkappa| \left\|A^{(1)}\right\|}{r_0 - |\varkappa| \left\|A^{(1)}\right\|}. \tag{4}$$

Now consider an operator $\tilde{A}^{(1)}(\varkappa) \equiv \frac{1}{\varkappa} A(\varkappa) P_0(\varkappa)$. Since $M_0(\varkappa)$ is the invariant subspace of $A(\varkappa)$ corresponding to the 0-group eigenvalues we have:

$$\mu_1\left(\tilde{A}^{(1)}(\varkappa)\,|M_0(\varkappa)\right) = \frac{1}{\varkappa} a^{(1)}(\varkappa). \tag{5}$$

As explained in Kato (1980), see his formulae 2.17, 2.37, and 2.38 on pages 77 and 81,

$$\tilde{A}^{(1)}(\varkappa) = \frac{1}{2\pi i}\sum_{n=0}^{\infty}(-\varkappa)^n\int_\Gamma R(z)\left(A^{(1)}R(z)\right)^{n+1}zdz, \tag{6}$$

$$\tilde{A}^{(1)}(0) \equiv \tilde{A}^{(1)} = P_0 A^{(1)} P_0, \tag{7}$$

where the series is absolutely convergent for $|\varkappa| < r_0/\left\|A^{(1)}\right\|$. Equations (6) and (7) imply that

$$\left\|\tilde{A}^{(1)}(\varkappa)\right\| \le r_0\sum_{n=0}^{\infty}|\varkappa|^n\frac{\left\|A^{(1)}\right\|^{n+1}}{r_0^{n+1}} = \frac{r_0\left\|A^{(1)}\right\|}{r_0 - |\varkappa|\left\|A^{(1)}\right\|} \tag{8}$$

and

$$\left\|\tilde{A}^{(1)}(\varkappa) - \tilde{A}^{(1)}\right\| \le r_0\sum_{n=1}^{\infty}|\varkappa|^n\frac{\left\|A^{(1)}\right\|^{n+1}}{r_0^{n+1}} = \frac{|\varkappa|\left\|A^{(1)}\right\|^2}{r_0 - |\varkappa|\left\|A^{(1)}\right\|}. \tag{9}$$

Consider the discrepancy between $\mu_m\left(\tilde{A}^{(1)}(\varkappa)|M_0(\varkappa)\right)$ and $\mu_m\left(\tilde{A}^{(1)}|M_0\right)$ for $m = 1,...,n-k$. For real positive $\varkappa$, we have:

$$\left|\mu_m\left(\tilde{A}^{(1)}(\varkappa)|M_0(\varkappa)\right) - \mu_m\left(\tilde{A}^{(1)}|M_0\right)\right| \le \sup_{i=1,...,n}\left|\mu_i\left(P_0(\varkappa)\tilde{A}^{(1)}(\varkappa)P_0(\varkappa)\right) - \mu_i\left(\tilde{A}^{(1)}\right)\right|. \tag{10}$$

Indeed, clearly, $M_0(\varkappa)$ is an invariant subspace of $P_0(\varkappa)\tilde{A}^{(1)}(\varkappa)P_0(\varkappa)$ and $\mu_m\left(\tilde{A}^{(1)}(\varkappa)|M_0(\varkappa)\right)$, $m = 1,...,n-k$ are eigenvalues of $P_0(\varkappa)\tilde{A}^{(1)}(\varkappa)P_0(\varkappa)$. Note that since we assume that $\varkappa$ is positive and since, by assumption of the lemma A and $A^{(1)}$ are symmetric non-negative definite, eigenvalues $\mu_m\left(\tilde{A}^{(1)}(\varkappa)|M_0(\varkappa)\right)$, $m = 1,...,n-k$ are non-negative. The rest of the eigenvalues of $P_0(\varkappa)\tilde{A}^{(1)}(\varkappa)P_0(\varkappa)$ must be zero because $M_0^\perp(\varkappa)$ is the null-space of $P_0(\varkappa)\tilde{A}^{(1)}(\varkappa)P_0(\varkappa)$. Similarly, (7) implies that $\mu_m\left(\tilde{A}^{(1)}|M_0\right)$, $m = 1,...,n-k$, are non-negative eigenvalues of $\tilde{A}^{(1)}$ and the rest of the eigenvalues are zero. To summarize, $\mu_m\left(\tilde{A}^{(1)}(\varkappa)|M_0(\varkappa)\right) = \mu_m\left(P_0(\varkappa)\tilde{A}^{(1)}(\varkappa)P_0(\varkappa)\right)$ and $\mu_m\left(\tilde{A}^{(1)}|M_0\right) = \mu_m\left(\tilde{A}^{(1)}\right)$. These two equalities imply (10).

Inequality (10) and Weyl's inequalities (see Theorem 4.3.1 in Horn and Johnson (1985))

imply that

$$\left| \mu_m \left(\tilde{A}^{(1)} \left(\varkappa \right) | M_0 \left(\varkappa \right) \right) - \mu_m \left(\tilde{A}^{(1)} | M_0 \right) \right| \leq \left\| P_0 \left(\varkappa \right) \tilde{A}^{(1)} \left(\varkappa \right) P_0 \left(\varkappa \right) - \tilde{A}^{(1)} \right\|.$$

We have:

$$\left\| P_0 \left(\varkappa \right) \tilde{A}^{(1)} \left(\varkappa \right) P_0 \left(\varkappa \right) - \tilde{A}^{(1)} \right\| \leq \left\| \left(P_0 \left(\varkappa \right) - P_0 \right) \tilde{A}^{(1)} \left(\varkappa \right) P_0 \left(\varkappa \right) \right\| \tag{11}$$
$$+ \left\| P_0 \tilde{A}^{(1)} \left(\varkappa \right) \left(P_0 \left(\varkappa \right) - P_0 \right) \right\|$$
$$+ \left\| P_0 \left(\tilde{A}^{(1)} \left(\varkappa \right) - \tilde{A}^{(1)} \right) P_0 \right\|$$

Note that, for real $\varkappa$, P_0 and $P_0 \left(\varkappa \right)$ are orthogonal projections because A and $A \left(\varkappa \right)$ are self-adjoint. Therefore,

$$\| P_0 \| \leq 1, \text{ and } \| P_0 \left(\varkappa \right) \| \leq 1. \tag{12}$$

Using (4), (8), (9), and (12) to estimate right hand side of (11), we get:

$$\left\| P_0 \left(\varkappa \right) \tilde{A}^{(1)} \left(\varkappa \right) P_0 \left(\varkappa \right) - \tilde{A}^{(1)} \right\| \leq \left(2r_0 + r_0 - |\varkappa| \left\| A^{(1)} \right\| \right) \frac{|\varkappa| \left\| A^{(1)} \right\|^2}{\left(r_0 - |\varkappa| \left\| A^{(1)} \right\| \right)^2}.$$

Finally, combining the latter inequality with (10) and (5) and recalling that $\tilde{A}^{(1)} = P A^{(1)} P$ so that $\mu_m \left(\tilde{A}^{(1)} | M_0 \right) = \mu_m \left(P A^{(1)} P | M_0 \right)$, we get the statement of Lemma 1. $\square$

Now, we are ready to prove Theorem 2. Let us assume that the $T \times r$ matrix of factors F has all but the first r rows zero. There is no loss of generality in such an assumption because we can always find a unitary matrix Q such that QF has the above form. Multiplying the data matrix $X = \Lambda F' + e$ from the right by Q' does not change neither the eigenvalues of XX'/T nor the joint distribution of the elements of e. The latter fact follows from the complex Gaussianity of e, from the mutual independence of its columns, and from the independence of e from F.[5]

---

[5] Our assumption of the independence of F and e is used only at this point in the paper. It is possible to relax the assumption at the cost of more complicated technical proofs.

Denote the matrix that consists of the nonzero rows of F as $\tilde{F}$, the matrix of the first r columns of e as e_1, and the matrix of the last $T - r$ columns of e as e_2. Then we can decompose XX'/T into a sum of two terms: $\left(\Lambda\tilde{F}' + e_1\right)\left(\Lambda\tilde{F}' + e_1\right)'/T$ and $e_2 e_2'/T$. Now, let $R'AR$ be a spectral decomposition of the first term. Since the first term has rank r, diagonal matrix A can be chosen so that it has all but first r nonzero diagonal elements. We will denote these elements as $a_1 \geq a_2 \geq ... \geq a_r > 0$. Note that Assumptions 1 and 2 imply that $\liminf a_r/n \equiv 2r_0 > 0$ as n and T go to infinity so that $n/T \rightarrow b$.

Let us denote $R \cdot e_2$ as $\tilde{e}_2$. Note that $\tilde{e}_2$ has complex normal columns with Hermitian covariance matrix $R\Sigma R'$. Further, let us denote the matrix of the first r rows of $\tilde{e}_2$ as $\tilde{e}_{21}$, and the matrix of the last $n - r$ rows of $\tilde{e}_2$ as $\tilde{e}_{22}$. Then, we have: $RXX'R'/nT = A/n + (1/n)\left(\tilde{e}_2\tilde{e}_2'/T\right)$, and therefore, by Lemma 1:

$$\left|\lambda_{r+i}\left(RXX'R'/nT\right) - (1/n)\,\lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right)\right| \leq 3r_0\frac{(1/n)^2\left\|\tilde{e}_{22}\tilde{e}_{22}'/T\right\|^2}{\left(r_0 - (1/n)\left\|\tilde{e}_{22}\tilde{e}_{22}'/T\right\|\right)^2}. \qquad (13)$$

Now, Theorem 1 of El Karoui (2006) implies that $\|\tilde{e}_{22}\tilde{e}_{22}'/T\| \equiv \lambda_1\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right)$ converges in probability to a finite number. Therefore, (13) implies that $\left|\lambda_{r+i}\left(XX'/T\right) - \lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right)\right| = O_p\left(1/n\right)$. It remains to show that $T^{2/3}\sigma^{-1}\left(\lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right) - \mu\right)$, $i = 1, ..., j$ has the Tracy-Widom asymptotic joint distribution. Indeed, in such a case, since the difference between $\lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right)$ and $\lambda_{r+i}\left(XX'/T\right)$ is of order $1/n$ and n is asymptotically proportional to T, the random variables $T^{2/3}\sigma^{-1}\left(\lambda_{r+i}\left(XX'/T\right) - \mu\right)$, $i = 1, ..., j$ will have the same asymptotic joint distribution as $T^{2/3}\sigma^{-1}\left(\lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right) - \mu\right)$, $i = 1, ..., j$.

To prove that $T^{2/3}\sigma^{-1}\left(\lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/T\right) - \mu\right)$, $i = 1, ..., j$ has the Tracy-Widom asymptotic joint distribution, we will use Theorem 1 of El Karoui (2006). By the theorem (check conditions), $(T - r)^{2/3}\tilde{\sigma}^{-1}\left(\lambda_i\left(\tilde{e}_{22}\tilde{e}_{22}'/(T - r)\right) - \tilde{\mu}\right)$, $i = 1, ..., j$ has the Tracy-Widom asymptotic distribution, where $\tilde{\mu}$, $\tilde{\sigma}$, and $\tilde{c}$ are defined relative to the covariance matrix $\tilde{\Sigma}$, which consists of the intersection of the last $n - r$ columns and rows of $R\Sigma R'$, in the same way as μ, σ, and c are defined relative to Σ. Since the spectral distribution $\tilde{H}$ of

matrix $\tilde{\Sigma}$ is asymptotically equivalent to the spectral distribution H of matrix Σ, variables $(T-r)^{2/3} \tilde{\sigma}^{-1} \left(\lambda_i \left(\tilde{e}_{22} \tilde{e}'_{22} / (T-r) \right) - \tilde{\mu} \right)$, $i = 1, ..., j$ and $T^{2/3} \sigma^{-1} \left(\lambda_i \left(\tilde{e}_{22} \tilde{e}'_{22} / T \right) - \mu \right)$, $i = 1, ..., j$ have the same asymptotic distribution.$\square$

References

[1] Anderson, T.W. (1984) An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore.

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

[3] Bai, J., and S. Ng (2005) "Confidence intervals for diffusion index forecasts and inference for factor augmented regressions", *Econometrica*, forthcoming.

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

[5] Bollerslev, T. (1990) "Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model", Review of Economics and Statistics, 72, pp. 498-505.

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

[7] 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

[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. (2006) "Tracy-Widom Limit for the Largest Eigenvalue of a Large Class of Complex Wishart Matrices", manuscript, Stanford University

[10] Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000) "The generalized dynamic-factor model: identification and estimation", The Review of Economics and Statistics, 82, pp 540-554.

[11] Hallin, M. and Liska,.R. (2005) "The generalized dynamic factor model: determining the number Amen of factors" manuscript.

[12] Horn, R.A. and C. R. Johnson (1985) Matrix Analysis, Cambridge University Press, Cambridge, New York.

[13] 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.

[14] Johnstone, I (2001). "On the distribution of the largest eigenvalue in principal component analysis", Annals of Statistics, Vol. 29, pp 295-327.

[15] Kato, T (1980) Perturbation theory for linear operators, Springer-Verlag, Berlin, New York.

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

[17] Reinganum, M., (1981) "The arbitrage pricing theory: some empirical results", Journal of Finance, 36, pp 313-321.

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

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

[20] Soshnikov, A. (2002) "A Note on Universality of the Distribution of the Largest Eigenvalues in Certain Sample Covariance Matrices", Journal of Statistical Physics, Vol. 108 (5/6), pp. 1033-56

[21] Stock, J., and M. Watson (2005) "Implications of Dynamic Factor Models for VAR Analysis", manuscript available at http:// www.wws.princeton.edu/ mwatson/ wp.html

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

[23] Tracy C. A. and H. Widom (2002) "Distribution Functions for Largest Eigenvalues and Their Applications", ICM, Vol 1, pp.587-596.

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

[25] Watson, M. and Amengual, D. (2006) "Consistent estimators of the number of dynamic factors in a large N and T panel", fourthcoming Journal of Business and Economic Statistics.