Testing hypotheses about the number of factors in large factor models.

Alexei Onatski, Columbia University

(Preliminary and Incomplete)

November 14, 2007

Abstract

In this paper we study high-dimensional time series that have the generalized dynamic factor structure. We develop a test of the null that the number of factors k equals k_0 versus an alternative that $k \neq k_0$ while $0 \leq k \leq k_{\max}$, where $k_{\max}$ is an *a priori* maximum number of factors. Our test is based on the statistic $\max_{k_0 < k \leq k_{\max}} \left(\gamma_k - \gamma_{k+1}\right) / \left(\gamma_{k+1} - \gamma_{k+2}\right)$, where γ_i is the i-th largest eigenvalue of the smoothed periodogram estimate of the spectral density matrix of data at a pre-specified frequency. We describe the asymptotic distribution of the statistic, as the dimensionality and the number of observations rise, as a function of the Tracy-Widom distribution and tabulate the critical values of the test. As an application, we construct asymptotic 95% confidence sets for the number of dynamic factors in macroeconomic time series and for the number of dynamic factors driving excess stock returns.

KEYWORDS: Generalized dynamic factor model, approximate factor model, number of factors, hypothesis test, Tracy-Widom distribution.

1 Introduction

High-dimensional factor models with correlated idiosyncratic terms have been extensively used in recent research in finance and macroeconomics. In finance, they form the basis of the Chamberlain-Rothschild (1983) extension of the Arbitrage Pricing Theory. They have been used in portfolio performance evaluation (Connor and Korajczyk, 1986), in the analysis of the profitability of trading strategies (Yao, 2006), in testing implications of the Arbitrage Pricing Theory (Zhang, 2006), and in the analysis of bond risk premia (Ludvigson and Ng, 2005). In macroeconomics, such models have been used in business cycle analysis, in forecasting, in monitoring economic activity, in construction of inflation indexes and in monetary policy analysis (see Breitung and Eickmeier (2005) for a survey of work in these areas). More recent macroeconomic applications include the analysis of international risk sharing (Giannone and Lenza, 2004), the identification of global shocks (Forni et al. 2005), the analysis of price dynamics (Boivin et al 2006), the instrumental variable estimation (Bai and Ng, 2006) and the estimation of the dynamic stochastic general equilibrium models (Boivin and Giannoni 2006).

An important question to be addressed by any study which uses factor analysis is how many factors there are. The number of factors is needed in the implementation of various estimation and forecasting procedures. Moreover, it often has interesting economic interpretations and important theoretical consequences. In finance and macroeconomics, it can be interpreted as the number of the sources of non-diversifiable risk, and the number of the fundamental shocks driving the macroeconomic dynamics, respectively. In consumer demand theory, the number of factors in budget share data provides crucial information about the demand system (see Lewbel, 1991).[1] For

[1] We thank an anonimous referee for pointing out this fact.

example, the number of factors must be exactly two for aggregate demands to exhibit the weak axiom of revealed preference.

Although there have been many studies which develop consistent estimators of the number of factors (for recent work in this area see Bai and Ng (2002, 2005), Stock and Watson (2005), Hallin and Liska (2005), Kapetanios (2004), 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. In finance, the estimated number of factors ranges from two to more than ten. In macroeconomics, there is an ongoing debate (see Stock and Watson, 2005) whether the number of factors is only two or, perhaps, larger than seven. The purpose of this paper is to develop formal statistical tests of various hypotheses about the number of factors in large factor models. Such tests can be used, for example, to decide between competing point estimates or to provide confidence intervals for the number of factors.

We consider T observations $X_1, ..., X_T$ of n-dimensional vectors that have the generalized dynamic factor structure introduced by Forni et al. (2000). In particular, $X_t = \Lambda(L)F_t + e_t$, where $\Lambda(L)$ is an $n \times k$ matrix of possibly infinite polynomials in the lag operator L, F_t is a k-dimensional vector of factors at time t, and e_t is an n-dimensional vector of correlated stationary idiosyncratic terms. We develop the following test of the null that there are $k = k_0$ factors at a particular frequency of interest ω_0, say a business cycle frequency, vs. the alternative that $k \neq k_0$ while $0 \leq k \leq k_{\max}$, where $k_{\max}$ is an *a priori* maximum number of factors.

- First, compute the finite Fourier transforms $\hat{X}_j \equiv \sum_{t=1}^{T} X_t \exp\{-i\omega_j t\}$ of the data at frequencies $\omega_1 \equiv 2\pi s_1/T, .., \omega_m \equiv 2\pi s_m/T$ approximating ω_0, where $s_1, ..., s_m$ are integers such that $s_j \pm s_k \neq 0 \,(\mathrm{mod}\, T)$ for $j \neq k$ and

$\max_j |\omega_j - \omega_0| \leq 2\pi m/T$.

- Next, compute two statistics: $R \equiv \max_{k_0 < i \leq k_{\max}} \frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$ and $R_1 \equiv \frac{\gamma_{k_0} - \gamma_{k_0+1}}{\gamma_{k_0+1} - \gamma_{k_0+2}}$, where γ_i is the i-th largest eigenvalue of the smoothed periodogram estimate $\frac{1}{2\pi m} \sum_{j=1}^{m} \hat{X}_j \hat{X}_j'$ of the spectral density of the data at frequency ω_0. Here and throughout the paper the "prime" over a complex-valued matrix denotes the conjugate-complex transpose of the matrix.

- Finally, reject the null if and only if either of the following two inequalities hold: R is above a critical value given in Table 1, or R_1 is below a cutoff level $\bar{c}(n)$ such that $\bar{c}(n) \to \infty$ and $\bar{c}(n)n^{-2/3} \to 0$ as $n \to \infty$. For $n \leq 150$, we recommend using $\bar{c}(n) = 2$ although this choice does not influence the asymptotic properties of the test.

Table 1: Critical values of the test statistic R.

The test's size	$k_{\max} - k_0$							
%	1	2	3	4	5	6	7	8
15	2.75	3.62	4.15	4.54	4.89	5.20	5.45	5.70
10	3.33	4.31	4.91	5.40	5.77	6.13	6.42	6.66
9	3.50	4.49	5.13	5.62	6.03	6.39	6.67	6.92
8	3.69	4.72	5.37	5.91	6.31	6.68	6.95	7.25
7	3.92	4.99	5.66	6.24	6.62	7.00	7.32	7.59
6	4.20	5.31	6.03	6.57	7.00	7.41	7.74	8.04
5	4.52	5.73	6.46	7.01	7.50	7.95	8.29	8.59
4	5.02	6.26	6.97	7.63	8.16	8.61	9.06	9.36
3	5.62	6.91	7.79	8.48	9.06	9.64	10.11	10.44
2	6.55	8.15	9.06	9.93	10.47	11.27	11.75	12.13
1	8.74	10.52	11.67	12.56	13.42	14.26	14.88	15.25

For example, the 5% critical value for statistic R of the test of $k = k_0$ vs. $k \neq k_0$ while $0 \leq k \leq k_{\max}$ where $k_0 = 3$ and $k_{\max} = 10$ is in the 5th row (counting from the bottom up) and 2nd column (counting from the right) of the table. It equals 8.29.

We prove that as n, m and T go to infinity so that n/m remains in a compact subset of $(0, \infty)$ and T grows sufficiently faster than n, statistics R and R_1 behave as follows. Under the null, statistic R is asymptotically pivotal whereas $\Pr(R_1 < \bar{c}(n)) \to 0$. The asymptotic distribution of R is a function of the Tracy-Widom distribution (see Tracy and Widom, 1994), which refers to the asymptotic joint distribution of a few of the largest eigenvalues of a particular Hermitian random matrix as the dimensionality of the matrix tends to infinity. Under the alternative, either R explodes, or $\Pr(R_1 < \bar{c}(n)) \to 1$, or both. Our Monte Carlo analysis shows that the size and power of the test remain very good even for T comparable to n as long as m remains relatively large.

The main idea behind our test is as follows. Suppose there are k dynamic factors in X_t. Then there will be k static factors in $\hat{X}_j$. Consider the sample covariance matrix of $\hat{X}_j$. Compute its eigenvalues and leave away the largest k. What is left over should have the same asymptotic distribution as the largest eigenvalues in the zero-factor case. Now, the zero-factor case corresponds to $\hat{X}_j$ being asymptotically complex normal, independent finite Fourier transforms of X_t (see Theorem 4.4.1 of Brillinger 1981). The asymptotic distribution of the scaled and centered largest eigenvalues of the corresponding sample covariance matrix is known to be Tracy-Widom (see El Karoui (2007) and Onatski (2007)). However, the common asymptotic centering and scaling for the eigenvalues do depend on the unknown details of the correlation between the entries of vector $\hat{X}_j$. Our statistic R gets rid of both the unknown centering and scaling parameters which makes it asymptotically pivotal under the null.

Further, under the null, $R_1 = \frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}}$, and therefore it explodes because the denominator of $\frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}}$ is bounded whereas the numerator explodes. Under the alternative, when $k > k_0$, $R \geq \frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}}$, which, as we have just explained, explodes.

When $k < k_0$, R_1 converges in law to a non-degenerate distribution and, hence, $\Pr(R_1 < \bar{c}(n)) \to 1$.

Our test procedure can be interpreted as formalizing the widely used empirical method of the number of 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 appears. Our R and R_1 statistics effectively measure the curvature of the frequency-domain scree plot at a would-be break point under the alternative when $k > k_0$ and under the null, respectively. When $k > k_0$, the curvature measured by R asymptotically goes to infinity. In contrast, under the null, this curvature has a non-degenerate asymptotic distribution that does not depend on the model's parameters. Quite the opposite is true for the curvature measured by R_1. It explodes under the null, but converges to a non-degenerate distribution under "the other part" of the alternative hypothesis: $k < k_0$.

In an important special case when $\Lambda(L)$ does not depend on L but factors are not necessarily white noise, model (1) reduces to an approximate factor model of Chamberlain and Rothschild (1983). If, in addition, the idiosyncratic terms e_t follow an i.i.d. Gaussian process, we can test hypotheses about the number of factors using a modified procedure which does not rely on the frequency domain transformation of the data. First, we transform the original data X_t to the independent complex Gaussian form $\tilde{X}_j$ by splitting the sample into two time periods of equal length, multiplying the second half by the imaginary unit $\sqrt{-1}$, and adding to the first half. Then the second and the third steps of the above testing procedure follow without any changes.

The rest of the paper is organized as follows. Section 2 describes the model and states our assumptions. In Section 3 we develop the test. Section 4 considers the special case of the approximate factor model. Section 5 contains Monte Carlo experiments and comparisons with procedures proposed in the previous literature. Section 6 applies our test to macroeconomic and financial data. Section 7 concludes. Technical proofs are contained in the Appendix.

2 The model and assumptions

Suppose that our data $\{X_{it}; i \leq n, t \leq T\}$ come from a double sequence of random variables $\{X_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ which admits a version of the generalized dynamic k-factor structure of Forni et al. (2000). Precisely, for any i and t :

$$X_{it} = \Lambda_{i1}(L)F_{1t} + ... + \Lambda_{ik}(L)F_{kt} + e_{it}, \tag{1}$$

where $\Lambda_{ij}(L) \equiv \sum_{u=0}^{\infty} \Lambda_{ij}^{(u)} L^u$, and factor loadings $\Lambda_{ij}^{(u)}$, factors F_{jt}, and idiosyncratic terms e_{it} satisfy Assumptions 1, 2, 3, and 4 stated below.

To make the assumptions more compact, let us introduce the following notations. Let $X_t(n) \equiv (X_{1t}, ..., X_{nt})'$ denote the data vector at time t; $e_t(n) \equiv (e_{1t}, ..., e_{nt})'$ denote its idiosyncratic part; $\chi_t(n) \equiv X_t(n) - e_t(n)$ denote its systematic part; $S_n^e(\omega)$ and $S_n^\chi(\omega)$ denote the spectral density matrices of $e_t(n)$ and $\chi_t(n)$, respectively; $F_t \equiv (F_{1t}, ..., F_{kt})'$ denote the vector of factors at time t; $\Lambda^{(u)}(n)$ denote the $n \times k$ matrix of factor loadings $\Lambda_{ij}^{(u)}$ with $i = 1, ..., n$ and $j = 1, ..., k$; $\Lambda(n, L) \equiv \sum_{u=0}^{\infty} \Lambda^{(u)}(n) L^u$ denote the $n \times k$ matrix of the corresponding dynamic loadings; finally, let $\hat{X}_s(n) \equiv \sum_{t=1}^{T} X_t(n) e^{-i\omega_s t}/\sqrt{T}$, $\hat{F}_s \equiv \sum_{t=1}^{T} F_t e^{-i\omega_s t}/\sqrt{T}$ and $\hat{e}_s(n) \equiv \sum_{t=1}^{T} e_t(n) e^{-i\omega_s t}/\sqrt{T}$ denote the finite Fourier transforms of the processes $X_t(n)$, F_t and $e_t(n)$, respectively, at the frequencies ω_s, $s = 1, ..., m$, defined in the Introduction.

Assumption 1: i) *The factors F_t follow an orthonormal white noise process.*

ii) *For each n, vector $e_t(n)$ is independent from F_s at all lags and leads and follows a stationary zero-mean process.*

This assumption is standard. It is somewhat stronger than Assumption 1 in Forni et al. (2000) which requires only orthogonality of $e_t(n)$ and F_s.

Assumption 2: i) T*he factor loadings are such*[2] *that* $\sum_{u=0}^{\infty} \left\| \Lambda^{(u)}(n) \right\| (1+u) = O\left(n^{1/2}\right)$. **ii)** *The idiosyncratic terms e_{it}* are jointly Gaussian with *autocovariances* $c_{ij}(u) \equiv E e_{it} e_{j,t-u}$ *satisfying* $\sum_{u=0}^{\infty} (1+u) |c_{ij}(u)| < \infty$ *uniformly in* $i, j \in \mathbb{N}$.

To obtain results without referring to Gaussianity of e_{it}, we will need a stronger version of Assumption 2ii:

Assumption 2: iia) *For each n, $e_t(n)$ follows a strictly stationary process with slowly growing cumulants so that* $C_l \equiv \sup_{j_1,...,j_l} \sum_{u_1,...,u_{l-1}} \left| cum\left(e_{j_1 t_1+u}, e_{j_2 t_2+u}, ..., e_{j_{l-1} t_{l-1}+u}, e_{j_l u}\right)\right| < \infty$ *for* $l = 0, 1, 2, ...$ *and* $\sum_{k=0}^{\infty} C_{k+3} z^k / k! < \infty$ *for z in a neighborhood of* 0.

Assumptions similar to Assumption 2iia) have been often used to obtain strong convergence results for the periodogram based estimates of spectral density matrices (see Brillinger (1981), Chapter XXX). Assumption 2i guarantees that $\Lambda(n, e^{-i\omega_s})$ is well approximated by $\Lambda(n, e^{-i\omega_0})$ for T sufficiently larger than n and m. We need Assumption 2ii (or 2iia) to show that, for T sufficiently larger than n and m, the real and imaginary parts of vector $\hat{e}_s(n)$ are well approximated by Gaussian vectors which are independent across $s = 1, ..., m$ and have variance-covariance matrices proportional to $S_n^e(\omega_0)$.

Our next assumption concerns the asymptotic behavior of $S_n^e(\omega_0)$. Let $l_{1n} \geq ... \geq$

[2]For any matrix or vector A, $\|A\|$ denotes the standard Euclidean norm of A, that is $\|A\|^2$ equals the maximum eigenvalue of $A'A$.

l_{nn} be the eigenvalues of $S_n^e(\omega_0)$. Denote by H_n the spectral distribution of $S_n^e(\omega_0)$, that is $H_n(\lambda) = 1 - \frac{1}{n}\#\left\{i \leq n : l_{in} > \lambda\right\}$, where $\#\left\{\cdot\right\}$ denotes the number of elements in the indicated set. Further, let $c_{m,n}$ be the unique root in $\left[0, l_{1n}^{-1}\right)$ of the equation $\int \left(\lambda c_{m,n} / \left(1 - \lambda c_{m,n}\right)\right)^2 dH_n(\lambda) = m/n$.

Assumption 3: *As* n and m *tend to infinity so that* n/m *remains in a compact subset of* $(0, \infty)$*,* $\limsup l_{1n} < \infty$, $\liminf l_{nn} > 0$, *and* $\limsup l_{1n} c_{m,n} < 1$.

The inequalities of Assumption 3 have the following meaning. The inequality $\limsup l_{1n} < \infty$ guarantees that the cumulative effects of the idiosyncratic causes of variation at frequency ω_0 on the n observed cross-sectional units remain bounded as $n \to \infty$. It relaxes Assumption 3 of Forni et al. (2000) that the largest eigenvalue of $S_n^e(\omega)$ is *uniformly bounded* over $\omega \in [-\pi, \pi]$ which is a crucial identification requirement for generalized dynamic factor models. Our focus on a single frequency allows us not to worry about the identification at frequences other than ω_0. In particular, we allow the number of factors at other frequences to be different from k, a situation which may occur, for example, if data are preprocessed by a band-pass filter.

The inequality $\liminf l_{nn} > 0$ requires that the distribution of the ω_0-frequency component of the stationary process $e_t(n)$ does not become degenerate as $n \to \infty$. The inequality $\limsup l_{1n} c_{m,n} < 1$ is crucial for Onatski (2007) which we rely on in our further analysis. For this inequality to hold, it is sufficient that H_n weakly converges to a distribution H_∞ with density *bounded away from zero* in the vicinity of the upper boundary of support $\limsup l_{1n}$. Hence, the inequality essentially requires that relatively large eigenvalues of $S_n^e(\omega_0)$ do not scatter too much as n goes to infinity. Intuitively, such a requirement rules out situations when a few weighted averages of the idiosyncratic terms cause unusually large variation at frequency ω_0 so that they can be misinterpreted as common dynamic factors.

Assumption 4: *The k-th largest eigenvalue of $S_n^x(\omega_0)$ diverges to infinity faster than $n^{2/3}$.*

Note that k-th largest eigenvalue of $S_n^x(\omega_0)$ can be interpreted as measuring the cumulative explanatory power of "the least influential linear combination of factors" on the observed cross-sectional units at frequency ω_0. Hence, Assumption 4 insures that the factor explanatory power at frequency ω_0 rises sufficiently fast as n goes to infinity. It relaxes a standard requirement that factors' cumulative effects on the cross-sectional units rises linearly in n (see, for example, assumption A4 in Hallin and Liska, 2007).

Allowing for a slower than linear in n cumulative factor effect is important because it better corresponds to the current practice of using larger and larger macroeconomic data sets created by going to higher and higher levels of disaggregation. For example, suppose one disaggregates a normalized macroeconomic indicator $x = (y+z)/\sqrt{Var(y+z)}$ into two indicators: $y = F+\xi$ and $z = F+\eta$, where F, ξ, and η are independent and have variance 1/2. It is not difficult to see that $x = 2F/\sqrt{3}+\zeta$, where ζ is independent from F. Hence, the factor F explains 2/3 of the variation in x, but its cumulative explanatory power does not double increasing only to 1 for the disaggregated series y and z.

3 The test

Before we explain the workings of our test, let us recall the notions of the complex Gaussian and complex Wishart distributions (see, for example, Brillinger (1981), p. 89). We say that an n-dimensional vector Y has a complex Gaussian distribution $N_n^{\mathbb{C}}(\beta, \Sigma)$ if and only if the $2n$-dimensional vector $Z \equiv (\operatorname{Re} Y', \operatorname{Im} Y')'$ which stacks the real and imaginary parts of Y has a usual Gaussian distribution

$N_{2n}\left(\left(\begin{array}{c}\operatorname{Re}\beta\\ \operatorname{Im}\beta\end{array}\right),\frac{1}{2}\left(\begin{array}{cc}\operatorname{Re}\Sigma & -\operatorname{Im}\Sigma\\ \operatorname{Im}\Sigma & \operatorname{Re}\Sigma\end{array}\right)\right)$. Further, if $Y_1,...,Y_m$ are independent n-dimensional $N_n^{\mathbb{C}}(0,\Sigma)$ variates, then we say that the $n\times n$ matrix-valued random variable $\sum_{j=1}^{m} Y_j Y_j'$ has a complex Wishart distribution $W_n^{\mathbb{C}}(m,\Sigma)$ of dimension n and degrees of freedom m.

Our test is based on the following three observations. First, Assumption 2i implies that the vectors of the finite Fourier transform of our data admit an approximate factor structure in the sense of Chamberlain and Rothschild (1983). That is, $\hat{X}_s(n) = \Lambda(n,e^{-i\omega_0})\hat{F}_s + \hat{e}_s(n) + R_s(n)$, where, as Lemma 4 in the Appendix shows, $R_s(n)$ with $s=1,...,m$ can be made uniformly arbitrarily small for T which is larger enough than n and m.

Second, as is well known (see, for example, Theorem 4.4.1 of Brillinger, 2001), for fixed n and m, Assumption 2ii (or 2iia) implies that $\hat{e}_1,...,\hat{e}_m$ converge in distribution to m independent $N_n^{\mathbb{C}}(0,2\pi S_n^e(\omega_0))$ vectors and hence, the smoothed periodogram estimate $\hat{S}_n^e(\omega_0)\equiv\sum_{s=1}^{m}\hat{e}_s\hat{e}_s'/2\pi m$ of $S_n^e(\omega_0)$ converges in distribution to a complex Wishart $W_n^{\mathbb{C}}(m,S_n^e(\omega_0)/m)$ random matrix. As Lemma 5 in the Appendix shows, the complex Wishart approximation of $\hat{S}_n^e(\omega_0)$ remains good even if n and m are not fixed as long as T grows sufficiently faster than n and m.

Finally, let $\gamma_1\geq ...\geq\gamma_n$ be the eigenvalues of $\sum_{s=1}^{m}\hat{X}_s\hat{X}_s'/2\pi m$. Then, since, as have been observed above, $\hat{X}_s$ have an approximate k-factor structure asymptotically, the eigenvalues $\gamma_1,...,\gamma_k$ must explode as $n,m\to\infty$ sufficiently slower than T while the rest of the eigenvalues must approach the eigenvalues of $\hat{S}_n^e(\omega_0)$, whose distribution becomes arbitrarily close to the distribution of the eigenvalues of a complex Wishart $W_n^{\mathbb{C}}(m,S_n^e(\omega_0)/m)$ matrix.

The last observation implies that as $n,m\to\infty$ sufficiently slower than T, our statistic $R\equiv\max_{k_0<i\leq k_{\max}}\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$ explodes under the alternative of $k>k_0$ but becomes well approximated in distribution by $\max_{0<i\leq k_{\max}-k_0} \frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1}-\lambda_{i+2}}$ under the null, where λ_i is the i-th largest eigenvalue of a complex Wishart $W_n^{\mathbb{C}}\left(m, \frac{1}{m}S_n^e(\omega_0)\right)$ matrix. On the contrary, $R_1 \equiv \frac{\gamma_{k_0}-\gamma_{k_0+1}}{\gamma_{k_0+1}-\gamma_{k_0+2}}$ explodes under the null but becomes well approximated in distribution by $\frac{\lambda_{k_0-k}-\lambda_{k_0-k+1}}{\lambda_{k_0-k+1}-\lambda_{k_0-k+2}}$ under the alternative $k < k_0$.

The marginal distribution of the largest eigenvalue of $W_n^{\mathbb{C}}(m, \Sigma)$ as both n and m tend to infinity so that n/m remains in a compact subset of $(0, 1)$ is studied by El Karoui (2007). Onatski (2007) extends his results to the case of the joint distribution of a few of the largest eigenvalues of $W_n^{\mathbb{C}}(m, \Sigma)$ as both n and m tend to infinity so that n/m remains in a compact subset of $(0, \infty)$. We have:

Lemma 1: (Onatski, 2007) *Let Assumption 3 hold. Define*

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

Then, for any positive integer r, as m and n tend to infinity so that n/m remains in a compact subset of $(0, \infty)$, the joint distribution of the first r centered and scaled eigenvalues $\sigma_{m,n}^{-1}\left(\lambda_1 - \mu_{m,n}\right), ..., \sigma_{m,n}^{-1}\left(\lambda_r - \mu_{m,n}\right)$ of matrix $W_n^{\mathbb{C}}(m, S_n^e(\omega_0)/m)$ weakly converges to the r-dimensional joint Tracy-Widom distribution of type 2.

The Tracy-Widom law of type 2 (denoted as TW_2 in what follows) refers to a distribution with a 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 $q''(s) = sq(s) + 2q^3(s)$, which is asymptotically equivalent to the Airy function $Ai(s)$ as $s \to \infty$.[3] It plays an important role in large random matrix theory (see Mehta, 2004) because it is the asymptotic distribution of the scaled and normalized largest eigenvalue of a matrix

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

from the so called Gaussian Unitary Ensemble (GUE) as the size of the matrix tends to infinity.

The GUE is the collection of all $n \times n$ Hermitian matrices with i.i.d. complex Gaussian $N_1^{\mathbb{C}}(0, 1/n)$ lower triangular entries and (independent from them) i.i.d. real Gaussian $N_1(0, 1/n)$ 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 of the largest eigenvalues of matrices from GUE when $n \to \infty$. They described the asymptotic marginal distributions of $\tilde{d}_1, ..., \tilde{d}_r$ where r is any fixed positive integer, in terms of a solution of a completely integrable system of partial differential equations. If we are interested in the asymptotic distribution of the largest eigenvalue only, the system simplifies to the single ordinary differential equation given above. The joint asymptotic distribution of $\tilde{d}_1, ..., \tilde{d}_r$ is called a joint TW_2 distribution.

Lemma 1 implies that as long as the distribution of $\gamma_{k+1}, ..., \gamma_{k+r}$ is well approximated by the distribution of $\lambda_1, ..., \lambda_r$, we can test our null hypothesis $k = k_0$ by checking whether the scaled and centered eigenvalues $\gamma_{k_0+1}, ..., \gamma_{k_0+r}$ come from the joint TW_2 distribution and, simultaneously, whether γ_{k_0} is too large for $\gamma_{k_0}, \gamma_{k_0+1}$ and γ_{k_0+2} to be a part of a Tracy-Widom-distributed vector. Our test statistics R and R_1 are designed so as to get rid of the unknown scale and center parameters $\sigma_{m,n}$ and $\mu_{m,n}$ which makes such a testing strategy feasible.

Theorem 1: *Let Assumptions 1,2i,3 and 4 hold, and n, m and T go to infinity so that n/m remain in a compact subset of $(0, \infty)$. Then if either Assumption 2ii holds and $m = o\left(T^{3/8}\right)$, or Assumption 2iia holds and $m = o\left(T^{3/25}\right)$, there exist sequences of scale and center parameters $\tilde{\sigma}_{m,n} \tilde{} m^{-2/3}$ and $\tilde{\mu}_{m,n}$ such that for any positive integer r, the joint distribution of $\tilde{\sigma}_{m,n}^{-1}\left(\gamma_{k_0+1} - \tilde{\mu}_{m,n}\right), ..., \tilde{\sigma}_{m,n}^{-1}\left(\gamma_{k_0+r} - \tilde{\mu}_{m,n}\right)$ weakly converges to the r-dimensional joint TW_2 distribution.*

The proof of the theorem is given in the Appendix. As the Monte Carlo analysis of the next section suggests, the rates required by Theorem 1 are sufficient but not necessary for the theorem to hold. Our test has good finite sample power and size properties even for m comparable to T. A reason for this is that the Tracy-Widom limit appears to be universal for a class of random matrices much wider than the class of complex Wishart matrices. First important universality results were recently obtained by Soshnikov (2002). Further development of such results remains a challenge for mathematicians.

Our next theorem formally states the properties of our test.

Theorem 2: *Under conditions of Theorem 1, if* $k = k_0$*, statistic* R *converges in distribution to* $\max_{0<i\leq k_{\max}-k_0} \frac{\lambda_i-\lambda_{i+1}}{\lambda_{i+1}-\lambda_{i+2}}$*, where* $\lambda_1, ..., \lambda_j$ *are random variables with joint* TW_2 *distribution, and statistic* R_1 *diverges to infinity in probability faster than* $n^{2/3}$*. In contrast, when* $k \neq k_0$ *and* $k_0 < k \leq k_{\max}$*,* R *diverges to infinity in probability, and when* $k \neq k_0$ *and* $k < k_0$*,* R_1 *converges in distribution to* $\frac{\lambda_{k_0-k}-\lambda_{k_0-k+1}}{\lambda_{k_0-k+1}-\lambda_{k_0-k+2}}$*. Such a behavior of* R *and* R_1 *implies that our test of the null of* $k = k_0$ *vs. the alternative of* $k \neq k_0$ *while* $0 \leq k \leq k_{\max}$ *is consistent and has asymptotically correct size.*

A proof of the theorem is in the Appendix.

4 A test for the number of approximate factors

In an important special case when $\Lambda(L)$ does not depend on L but factors are not necessarily white noise, model (1) reduces to an approximate factor model of Chamberlain and Rothschild (1983). As is well known, a dynamic factor model with uniformly finite number of lags in $\Lambda_{ij}(L)$ can always be reduced to an approximate factor model by including the lags of dynamic factors into the set of approximate factors. For the special case when $\Lambda(L)$ does not depend on L but factors are not necessarily white noise, the Introduction describes an alternative testing procedure (for the null of $k = k_0$ *approximate* as opposed to *dynamic* factors) which is not related to the frequency domain transformation of the original data. It can be shown that this test procedure is valid if the following modifications of Assumptions 1 through 4 hold:

Assumption 1m: **i)** *The factors F_t follow a fourth-order zero-mean stationary process with non-degenerate variance, autocovariances $\Gamma_{ij}(u) \equiv EF_{it}F_{j,t+u}$ decaying to zero as u tends to infinity, and 4-th order cumulants cum $(F_{it_1}, F_{jt_2}, F_{rt_3}, F_{l0})$ decaying to zero as t_1, t_2, and t_3 tend to infinity.*

ii) *For each n, vector $e_t(n)$ is independent from F_s at all lags and leads and follows a stationary Gaussian zero-mean process.*

Assumption 2m: **i)** $\Lambda_{ij}^{(u)} = 0$ *for any* $u \neq 0$, **ii)** $c_{ij}(u) \equiv Ee_{it}e_{j,t-u}$ *are zero for any* $u \neq 0$.

Assumption 3m: *As n, and T tend to infinity so that n/T remains in a compact subset of $(0, \infty)$, $\limsup l_{1n} < \infty$, $\liminf l_{nn} > 0$, and $\limsup l_{1n}c_{T/2,n} < 1$, where $l_{1n}, ..., l_{nn}$ refer to the eigenvalues of $Ee_t(n)e_t'(n)$ and $c_{T/2,n}$ is defined as in Assumption 3 with H_n replaced by the spectral distribution of $Ee_t(n)e_t'(n)$.*

Assumption 4m: *The k-th largest eigenvalue of $\Lambda^{(0)}(n)\Lambda^{(0)}(n)'$ diverges to infinity faster than $n^{2/3}$*

Define $\tilde{X}_j(n) \equiv X_j(n) + \sqrt{-1}X_{j+T/2}(n)$ and let $\tilde{\gamma}_1 \geq \tilde{\gamma}_2 \geq ...$ be the eigenvalues of $\frac{2}{T}\sum_{j=1}^{T/2} \tilde{X}_j\tilde{X}_j'$. Then our statistics $\tilde{R}$ and $\tilde{R}_1$ for the approximate factor test procedure are defined in the same way as statistics R and R_1 but with γ_i replaced by $\tilde{\gamma}_i$. We have the following theorem.

Theorem 3: *Under Assumptions 1m through 4m, statistics $\tilde{R}$ and $\tilde{R}_1$ behave in the same way as statistics R and R_1 in Theorem 2. Our test of the null of $k = k_0$ approximate factors vs. the alternative of $k \neq k_0$ while $0 \leq k \leq k_{\max}$ is consistent* *and has asymptotically correct size.*

A proof of the theorem is similar to the proof of Theorem 2 and is available from the author upon a request. In the next section we explore the finite sample properties of our test.

5 A Monte Carlo Study

We approximate the joint 10-dimensional Tracy-Widom distribution of type 2 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. The left panel of Figure 1 shows the empirical distribution function of the largest eigenvalue centered by 2 and scaled by $n^{2/3} = 1000^{2/3}$. It approximates the univariate Tracy-Widom distribution of type 2. Tracy and Widom (2002) report that the mean of their univariate distribution[4] 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 the empirical distribution function of the ratio $(x_1 - x_2) / (x_2 - x_3)$, where x_i denotes the i-th largest eigenvalue of a matrix from GUE. It approximates the asymptotic cumulative distribution function of our test statistic $\left(\gamma_{k_0+1} - \gamma_{k_0+2}\right) / \left(\gamma_{k_0+2} - \gamma_{k_0+3}\right)$ for the test of the null of k_0 factors against an alternative that the number of factors is $k_0 + 1$. The graph reveals that it is not uncommon to see large values of the statistic under the null. This observation suggests that *ad hoc* methods of the determination of the number of factors based on visual

[4] They denote the distribution as F_2 (type 2) to distinguish it from the (type 1 and type 4) 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.

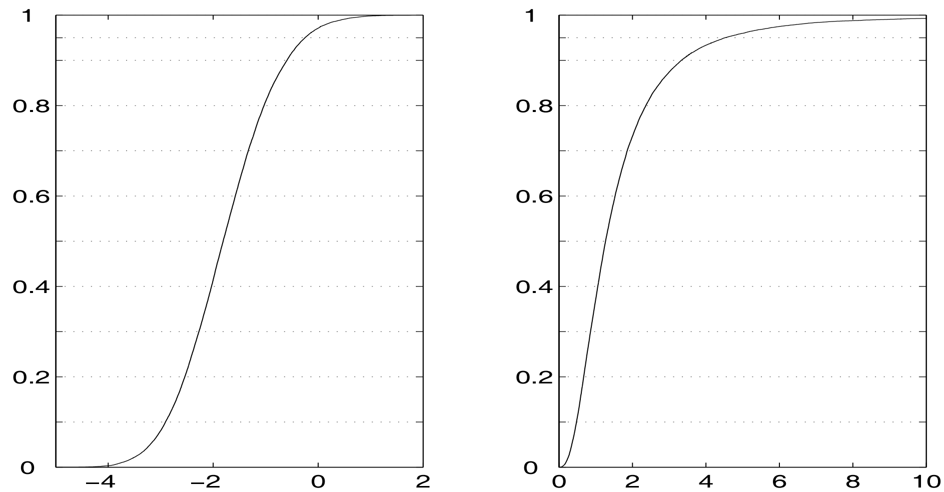

Figure 1: Left panel: CDF of the 1-dimensional Tracy-Widom distribution of type 2. Right panel: CDF of the statistic $(x_1 - x_2) / (x_2 - x_3)$, where x_1, x_2, and x_3 have joint 3-dimensional Tracy-Widom distribution of type 2.

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.

According to Theorems 2 and 3, the approximate asymptotic critical values of our test of $k = k_0$ versus $k \neq k_0$ while $0 \leq k \leq k_{\max}$ equal the corresponding percentiles of the empirical distribution of $\max_{0<i\leq k_{\max}-k_0} (x_i - x_{i+1}) / (x_{i+1} - x_{i+2})$. Table 1 given in the Introduction contains such percentiles for $k_{\max} - k_0 = 1, 2, ..., 8$.

5.1 Size-power properties of the dynamic factor test

To study the finite-sample properties of our dynamic factor test, we generate data from model (1) as follows. The k-dimensional factor vectors F_t are i.i.d. $N(0, I_k)$. The idiosyncratic components e_{it} follow AR(1) processes both cross-sectionally and over time: $e_{it} = \rho_i e_{it-1} + v_{it}$, where ρ_i are i.i.d. $U[-.8, .8]$, $v_{it} = 0.2 v_{i-1t} + u_{it}$ and u_{it} are i.i.d. $N(0, 1)$. The filters $\Lambda_{ij}(L)$ are randomly generated (independently from F_t's and e_{it}'s) by one of the two devices proposed by Hallin and Liska (2007):

- MA loadings: $\Lambda_{ij}(L) = \Lambda_{ij}^{(0)} + \Lambda_{ij}^{(1)} L + \Lambda_{ij}^{(2)} L^2$ with iid and mutually independent coefficients $\left(\Lambda_{ij}^{(0)}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\right) \sim N(0, I_3)$,
- AR loadings: $\Lambda_{ij}(L) = b_{ij}^{(0)} \left(1 - b_{ij}^{(1)} L\right)^{-1} \left(1 - b_{ij}^{(2)} L\right)^{-1}$, with iid and mutually independent coefficients $b_{ij}^{(0)} \sim N(0, 1)$, $b_{ij}^{(1)} \sim U[.8, .9]$, and $b_{ij}^{(2)} \sim U[.5, .6]$.

We normalize the systematic components $\sum_{j=1}^{k} \Lambda_{ij}(L) F_{jt}$ and the idiosyncratic components e_{it} so that their variances equal 1/2.

Solid lines on Figure 2 show the p-value discrepancy plots and size-power curves for our dynamic factor test. On the horizontal axis of a p-value discrepancy plot, we have the nominal size of a test, on the vertical axis, we have the difference between the finite-sample size and the nominal size. A size-power curve is a plot of the power against the *true* size of a test. We consider the test of the null of $k = 2$ factors at frequency zero against the alternative of $k \neq 2$ while $k \leq 8$. The graphs are based on 10,000 replications of data with $(n, T) = (100, 120)$. The frequencies ω_j "approximating" the zero frequency are $2\pi/T, ..., 2\pi 40/T$ so that $m = 40$. The reported power of the test corresponds to the true number of factors equal to 3.

For AR loadings, the finite-sample size of our test is somewhat larger than the nominal size. However, the discrepancy is small. For MA loadings, the size properties of the test are much worse. One reason for the better performance of our tests in the former case is that, in contrast to the MA loadings, the AR loadings are designed so that their Fourier transforms are relatively large at low frequences. Therefore, our choice of the zero frequency as the frequency of interest makes the tests' task easier. The power of the test in both AR and MA cases is very good.

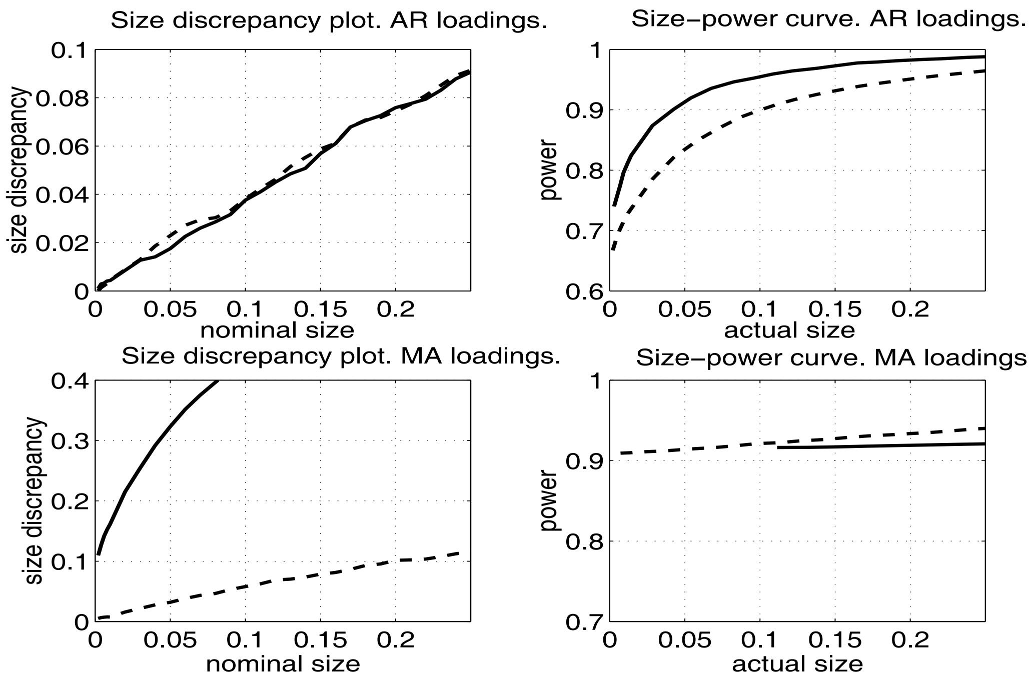

Figure 2: Size-power properties of the test when $(n, m, T) = (100, 40, 120)$. Dash line: test based on the weighted periodogram.

We tried several modifications of the benchmark test. Dashed lines on Figure 2 show the size-power performance of the best modification found. The modification is based on the weighted periodogram estimate of the spectral density matrix $\frac{1}{\pi m}\sum_{j=1}^{m}\left(1-\frac{j-1}{m}\right)\hat{X}_j\hat{X}_j'$ as opposed to the unweighted estimate $\frac{1}{2\pi m}\sum_{j=1}^{m}\hat{X}_j\hat{X}_j'$ used by the benchmark test. As is well known (see, for example, Section 5.5 of Brillinger, 2001), weighting reduces the bias of the periodogram estimate. This fact may explain better size properties of the weighted periodogram test relative to the unweighted periodogram test. Since the size properties of the weighted test are considerably better than those of the unweighted test and, at the same time, the power properties of the two versions of the test are not very different, we recommend to use in empirical applications the weighted test .

To check the robustness of our size-power analysis with respect to the non-Gaussianity of the simulated data, we repeated our Monte Carlo experiments altering the distribution of the idisoycratic terms' innovations u_{it}. We consider two alternative distributions: the centered chi-squared distribution $\chi^2(1) - 1$ and the Student's $t(5)$ distribution. The former distribution is an example of a skewed distribution, and the latter one is an example of a distribution with relatively fat tails. For both the chi-squared and Student's $t(5)$ cases the size-power properties of our tests are similar to the Gaussian case both quantitatively and qualitatively.

[Choice of the cutoff for R_1 statistics and choice of m here:]

5.2 Size-power properties of the approximate factor test

To study the finite-sample properties of our approximate factor test we generate data from model (1) with $\Lambda_{ij}(L) = \Lambda_{ij}^{(0)}$ as follows. The k-dimensional factor vectors F_t follow VAR(1) process $F_t = AF_{t-1} + \varepsilon_t$, where A is a diagonal matrix with i.i.d. diagonal elements $A_{ii} \sim U[0,1]$ and $\varepsilon_t \sim N(0, I_k - A^2)$. The idiosyncratic components e_{it} are independent from F_s's and follow an AR(1) process cross-sectionally. That is, $e_{it} = 0.2e_{i-1t} + v_{it}$, where v_{it} are i.i.d $N(0,1)$. The factor loadings $\Lambda_{ij}^{(0)}$ are $N(0,1)$ random variables independent from F_s's and e_{rs}'s. We normalize the systematic com-

Table 2: Influence of the choice of the cutoff for R_1 on the test's size-power properties

cutoff	1.27	1.39	1.53	1.68	1.86	2.09	2.37	2.75	3.33	4.52
Pr($TW_2 >$ cutoff)	.5	.55	.6	.65	.7	.75	.8	.85	.9	.95
AR, size	7.29	7.29	7.29	7.29	7.29	7.29	7.29	7.30	7.31	7.41
MA, size	7.96	7.98	7.98	8.04	8.09	8.25	8.59	9.50	11.4	19.0
AR, power, $k = k_0 + 1$	76.1	77.1	78.2	79.4	80.9	83.1	85.2	88.0	91.1	95.3
MA, power, $k = k_0 + 1$	75.6	79.8	83.5	86.9	89.8	92.3	94.8	96.7	98.3	99.4
AR, power, $k = k_0 - 1$	41.3	46.3	51.7	57.0	62.5	68.1	73.7	80.2	86.1	92.7
MA, power, $k = k_0 - 1$	38.6	43.4	47.9	52.8	57.7	63.8	69.8	76.7	83.9	91.4

ponents $\sum_{j=1}^{k} \Lambda_{ij}^{(0)} F_{jt}$ and the idiosyncratic components e_{it} so that their variances equal 1/2. To create complex-valued data sets we add the first $T/2$ simulations of data and $\sqrt{-1}$ times the last $T/2$ simulations of data.

Figure 4 explores the size-power properties of the test of 2 vs. 3 factors for different combinations of n and T. The left panel shows the equal-size-contours for the test with nominal size 0.05. The right panel shows the equal-power-contours for the test with true size 0.05. The solid lines correspond to the Monte Carlo setting described in the previous paragraph. The dash, dash-dot and dot lines correspond to the $\chi^2(1) - 1$-distributed innovations v_{it}, $t(5)$-distributed innovations v_{it}, and the two-way correlated idiosyncratic terms generated as in the Monte Carlo setup of Section 4.1, respectively. To facilitate reading, we report only 0.06-size-contours and 0.8-power-contours for the alternative Monte Carlo settings.

5.3 Comparison to other tests

Our next task is to compare the finite sample properties of our test and the tests proposed by Connor and Korajzcyk (1993), Kapetanios (2005), and Jacobs and Otter (2005). In contrast to our test, these tests require the dynamic factor loadings $\Lambda_{ij}(L)$ be lag polynomials of finite order r. Furthermore, without the knowledge of r, they would be testing joint hypotheses about the number of dynamic factors k and the lag length r. In the comparisons below, we, therefore, will assume that r is finite and known. Then, the Connor-Korajzcyk, Kapetanios, and Jacobs-Otter procedures can be used to test the null of $k = k_0$ against the alternative $k_0 < k$ which is more narrow than the alternative $k \neq k_0$ handled by our test.

Let us briefly describe the alternative tests. Connor and Korajczyk (1993) test the null of $p = p_0$ *approximate* factors versus the alternative that $p = p_0+1$. Their test uses

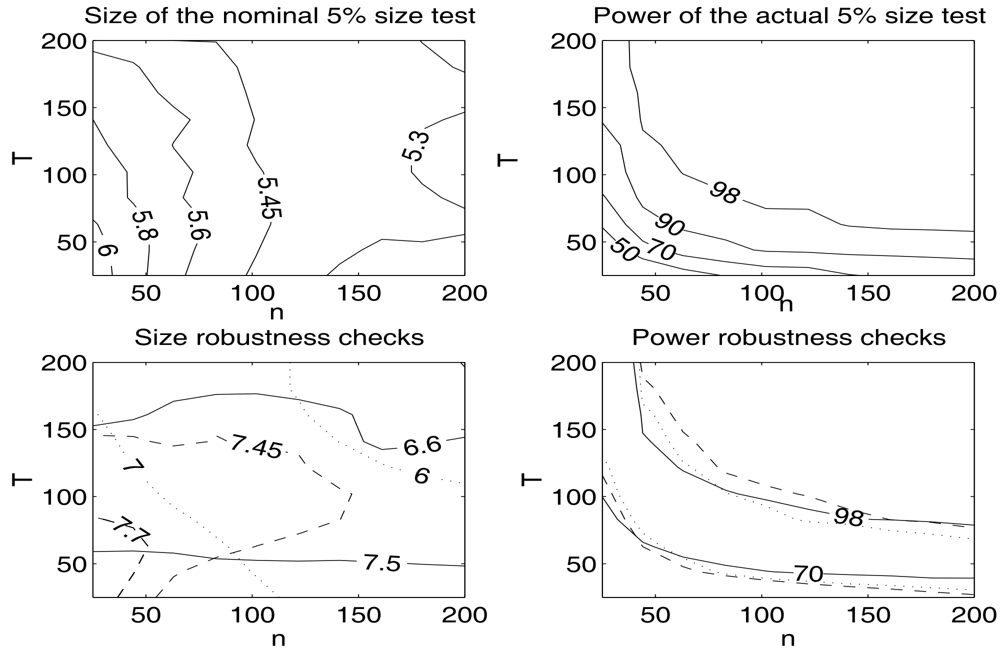

Figure 3: Upper pannel: benchmark. Lower panel: solid line- two-way correlation; dash line- t(5) innovations; dot line: $\chi^2(1) - 1$ innovations.

the fixed T, large n asymptotics and is based on the idea that the explanatory power of an extra $p_0 + 1$-th factor added to the model should be small under the null and large under the alternative. We can use the Connor-Korajczyk procedure to test the null of $k = k_0$ dynamic factors by transforming the dynamic factor model, via reinterpreting lags of dynamic factors as separate approximate factors, into an approximate factor model with $k(1+r)$ factors and testing the null of $k(1+r) = k_0(1+r)$ approximate factors.

Kapetanios (2005) tests the null of $p = p_0$ *approximate* factors vs. the alternative that $p_0 < p \leq p_{\max}$. He employs a subsampling method to approximate the asymptotic distribution of a test statistic $\lambda_{k_0+1} - \lambda_{k_{\max}+1}$, where λ_i is the i-th largest eigenvalue of the sample covariance matrix of the data $\frac{1}{T}\sum_{t=1}^{T} X_t X_t'$. He makes high-level assumptions about the existence of a scaling of $\lambda_{k_0+1} - \lambda_{k_{\max}+1}$ which converges in distribution to some unknown limit law, about properties of such a law, and about the functional form of the scaling constant. Kapetanios' test can be adapted to testing hypotheses about the number of dynamic factors in the same way as the Connor-Korajczyk test.

Jacobs and Otter (2005) propose a procedure to determine the number of *dynamic* factors based on formal statistical tests of hypotheses about singular values of certain combination of the sample auto-covariance matrices of the data. These tests can be used to test the null of $k = k_0$ dynamic factors versus the alternative of $k > k_0$. Jacobs and Otter's framework relies on the classical large T, small n asymptotics.

To compare the size-power properties of the tests we use the MA-loadings Monte Carlo setting of Section 5.1 with lag length $r = 2$. Table 2 reports the actual size of the nominal 5% size tests and the power of the actual 5%-size tests for the hypotheses of 0 vs. 1 dynamic factor and 2 vs. 3 dynamic factors.

Columns "O1" correspond to our dynamic factor test with $\omega_0 = 0$ and m equal to the integer part of $\min\left(4\sqrt{T}, T/2\right)$. Columns "O2" correspond to our static factor test of the hypotheses of 0 vs. 1-3 approximate factors, and 6 vs. 0-5 or 7-9 approximate factors. Columns "K" correspond to Kapetanios' test of 0 vs. 1-3 and 6 vs. 7-9 approximate factors. Columns denoted "CK" correspond to the Connor-Korajczyk test of 0 vs. 1 and 6 vs. 7 approximate factors. We do not report results for the Jacobs-Otter procedure because it always rejected the nulls in all of our Monte Carlo experiments.[Note here] The empty spaces in Table 2 correspond to situations when the minimal size detectable by our Monte Carlo experiment exceeded 5% so that the power of the actual 5% size test could not be evaluated.

Overall, our tests are much better sized than the alternative tests. The size of our tests become particularly good when $\min(n, T) \geq 100$. The alternative tests have very large size distortions which do not disappear when the sample size grows. Our approximate factor test and Kapetanios' test are the most powerful of the four

Table 3: Size-power properties of the alternative tests

n	T	O1 k_0=0	O1 k_0=2	O2 k_0=0	O2 k_0=2	K k_0=0	K k_0=2	CK k_0=0	CK k_0=2
		Size of the nominal 5%-size test							
50	100	8.9	14.8	6.7	15.6	32.3	24.2	25.8	22.9
100	50	8.4	41.4	6.6	34.2	41.5	36.3	27.5	23.0
100	100	7.5	11.4	6.4	6.8	30.9	31.6	27.0	23.9
100	200	7.3	7.3	5.8	5.8	23.6	22.5	26.3	25.0
200	100	6.6	11.4	6.5	5.8	30.5	33.2	27.3	25.0
200	200	6.5	8.4	6.0	5.4	24.5	27.7	27.4	25.7
		Power of the actual 5%-size test							
50	100	59.4		97.9			99.6	80.6	15.6
100	50	13.8		89.9				46.0	10.7
100	100	57.6	86.3	99.9	86.8			85.2	27.6
100	200	100	98.6	100	98.6	100	100	98.8	59.4
200	100	58.2	90.3	100	91.0			86.4	38.2
200	200	100	99.8	100	99.9	100		99.1	77.0

compared tests. The Connor-Korajczyk test is the least powerful for the test of 2 vs. 3 dynamic factors. Although our dynamic factor test is disadvantaged relative to the other tests because it does not use the additional information that $r = 2$, its power is comparable with the power of the alternative procedures, especially for the 2 vs. 3 dynamic factors case.

5.4 Using the tests to determine the number of factors

Using our test versus Hallin-Liska and Bai and Ng criteria.

6 Application

We apply our test to construct 95% confidence sets for the number of dynamic factors in macroeconomic and in financial data.

6.1 Macroeconomic factors

The literature is full of controversy about the number of dynamic factors driving macroeconomic data. Stock and Watson (2005) estimate seven dynamic factors in their data set. Giannone et al. (2005) find evidence supporting existence of only two dynamic factors. Two factors is also a preferred number in the older literature that uses factor models to describe the US economy.

The macroeconomic dataset we use is the same as in Stock and Watson (2002). It includes 215 monthly time series for the United States from 1959:1 to 1998:12 ($T = 480$). The variables in the dataset were transformed, standardized and screened for outliers as described in Stock and Watson (2002). Our analysis is based on the data subset of the transformed and screened 148 variables available for the full sample period ($n = 148$).

I maintain the assumption that the true number of dynamic factors is larger than zero but no larger than seven. My choice of one and seven as the lower and upper bounds is consistent with the conclusions of the majority of the previous studies. If a 5%-asymptotic-size test of the null of $k = k_0$ factors, versus the alternative of $k \neq k_0$, does not reject the null, I will include k_0 in the asymptotic 95% confidence set for the number of factors.

I set $m = 40$ and $\omega_1 = 2\pi/480, ..., \omega_{40} = 2\pi 40/480$ so that the range of the low and business cycle frequences is covered. I consider two versions of my test: one based on the weighted periodogram $\frac{1}{\pi m}\sum_{j=1}^{m}\left(1 - \frac{j-1}{m}\right)\hat{X}_j\hat{X}_j'$ and the other based on the weighted periodogram $\frac{1}{\pi m}\sum_{j=1}^{m}\frac{j}{m}\hat{X}_j\hat{X}_j'$. The first choice of weights emphasizes low freqeunces, and $\omega_0 = 0$ in particular. The second choice emphasizes ces close to the business cycle frequences, and $\omega_0 = 2\pi/12$ in particular. I set the cutoff level for statistic R_1 at 2.

Table 4 contains the first ten eigenvalues $\gamma_1, ..., \gamma_{10}$ of the two weighted periodograms normalized so that $\gamma_1 = 100$ and the corresponding ratios $(\gamma_i - \gamma_{i+1}) / (\gamma_{i+1} - \gamma_{i+2})$, $i = 1, ..., 8$. For the test emphasizing low frequencies, I reject all the tested nulls but the nulls of $1, 2$ and 7 dynamic factors, which means that the 95% confidence set for the number of pervasive dynamic factors at low frequences is $\{1, 2, 7\}$. For the test emphasizing business-cycle frequencess, I reject all the tested nulls but the nulls of $1, 2$ and 6 dynamic factor, which implies that the 95% confidence set for the number of pervasive dynamic factors at business-cycle frequences is $\{1, 2, 6\}$. The two 95% confidence sets would have changed[5] into $\{2\}$ for low frequences and $\{1\}$ for business-cycle frequences had we used the 95% quantile of the distribution of $(\lambda_1 - \lambda_2)/(\lambda_2 - \lambda_3)$, where $\lambda_1, \lambda_2, \lambda_3$ are joitly TW_2, as the cutoff level for R_1.

6.2 Financial factors

Similarly to the situation with the macroeocnomic factors, the previous studies of factors driving excess stock returns often do not agree and report from 1 to 6 such factors (see Onatski (2007) for a brief review of available results). The focus of the majority of these studies is on the number of the approximate factors. Therefore, below I construct 95% confidence sets both for the number of dynamic and the number of approximate factors driving the excess returns.

[5] Of course, although the asymptotic coverage rate of a confidence interval based on our test does not depend on the choice of the cutoff level, the finite sample coverage rate may change.

Table 4: Eigenvalue statistics for the Stock-Watson data.

ω_0	i	1	2	3	4	5	6	7	8	9	10
0	γ_i	100	33.3	13.9	11.4	8.88	5.31	3.45	2.43	2.00	1.61
0	$\frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$	3.44	7.78	1.00	0.70	1.91	1.82	2.39	1.08		
$2\pi/12$	γ_i	100	17.4	11.0	9.09	7.88	6.92	5.49	4.82	3.86	3.54
$2\pi/12$	$\frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$	12.9	3.32	1.59	1.27	0.67	2.13	0.70	2.99		

To construct the sets, I use the data provided by the Center for Research in Security Prices (CRSP) on monthly returns on $p = 972$ stocks traded on the NYSE, AMEX, and NASDAQ during the period from January 1983 to December 2006. 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 the 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 data, which leaves me with $T = 264$ time observations of real-valued data.

I maintain the assumption that the true number of dynamic factors is larger than zero but no larger than seven. I set $m = 40$ and consider the following three different frequency ranges. Range 1: $\omega_1 = 2\pi/264, ..., \omega_{40} = 2\pi 40/264$; Range 2: $\omega_1 = 2\pi 41/264, ..., \omega_{40} = 2\pi 80/264$; and Range 3: $\omega_1 = 2\pi 81/264, ..., \omega_{40} = 2\pi 120/264$. For Range 1, I consider the same two versions of my test as in the study of macroeconomic factors. For Ranges 2 and 3, I use the test based on the weighted periodogram $\frac{1}{\pi m}\sum_{j=1}^{m}\left|1 - 2\frac{j-20}{m}\right|\hat{X}_j\hat{X}_j'$. Hence, the two tests for Range 1 emphasize frequency $\omega_0 = 0$ and frequency $\omega_0 = 2\pi 40/264$ which corresponds to 1 cycle per 6.5 months. The test for Range 2 emphasizes frequency $\omega_0 = 2\pi 60/264$ which corresponds to 1 cycle per 4 months. The test for Range 3 emphasizes frequency $\omega_0 = 2\pi 100/264$ which corresponds to 1 cycle per 2.5 months.

As can be seen from Table 5, for the test corresponding to $\omega_0 = 0$, I do not reject the nulls of 1,2,3 and 6 factors. Hence, the 95% confidence set for the number of dynamic factors at low frequences is $\{1, 2, 3, 6\}$. For the test corresponding to $\omega_0 = \frac{2\pi 40}{264}$, only the nulls of 1,2 and 3 dynamic factors is not rejected. Hence, the 95% confidence set for the number of dynamic factors at the 1-cycle-per-half-a-year frequency is $\{1, 2, 3\}$. For the test corresponding to $\omega_0 = \frac{2\pi 60}{264}$, the only non-rejected

Table 5: Eigenvalue statistics for the excess stock return data.

ω_0	i	1	2	3	4	5	6	7	8	9	10
0	γ_i	100	53.4	31.2	25.2	23.6	19.8	17.9	17.4	15.6	14.8
	$\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$	2.10	3.69	3.81	0.42	1.97	3.41	0.32	2.10		
$\frac{2\pi 40}{264}$	γ_i	100	27.9	18.1	15.3	14.2	13.6	12.9	11.9	10.9	10.2
	$\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$	7.31	3.61	2.48	1.63	0.93	0.75	0.93	1.51		
$\frac{2\pi 60}{264}$	γ_i	100	56.4	33.4	29.3	25.6	23.7	22.2	21.3	20.1	18.2
	$\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$	1.89	5.62	1.10	2.00	1.21	1.75	0.75	0.59		
$\frac{2\pi 80}{264}$	γ_i	100	44.5	37.1	31.2	29.3	27.3	25.1	23.3	22.4	20.8
	$\frac{\gamma_i-\gamma_{i+1}}{\gamma_{i+1}-\gamma_{i+2}}$	7.49	1.24	3.22	0.94	0.91	1.17	2.16	0.54		

null is the null of 2 dynamic factors. Thus, the 95% confidence set is a singleton $\{2\}$. For the last test, which corresponds to relatively high frequencies, I do not reject the nulls of $1, 3$ and 7 dynamic factors. Hence, the 95% confidence set for the number of dynamic factors at 1-cycle-per-2-months frequency is $\{1, 3, 7\}$. there is a 95% chance that there are no common factors explaining low frequency movements of the excess stock returns.

To construct the 95% set for the number of approximate factors, I use my test developed for approximate factor models. I set the cutoff level as in the dynamic factor tests considered above. To get a complex-valued data set which is needed for the test, 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 to the product of the imaginary unit and the data from the second sub-period.

I maintain the assumption that the true number of approximate factors is larger than zero but no larger than nine, which is an upper bound on the estimates reported in Stock and Watson (2005). As can be seen from Table 6, I cannot reject the nulls of 1,2,4,6 and 8 approximate factors, so that the 95% confidence set for the number of approximate factors in excess stock returns is $\{1, 2, 4, 6, 8\}$. Since the approximate factors may correspond to the different lags of the same dynamic factor, such a 95% confidence set is not inconsistent with the 95% confidence sets for the number of dynamic factors reported above.

7 Discussion and Conclusion

TBA

8 Appendix

We first formulate and prove auxiliary Lemmas 1 through 5. Denote the i-th largest singular value of a matrix A with possibly complex entries, defined as the square root of the i-th largest eigenvalue of AA', as $\sigma_i(A)$.

Lemma 2 *For any matrix A, $\sigma_1^2(A) \leq \sum_{i,j} |A_{ij}|^2$.*

This is a well known result. See, for example, Horn and Johnson (1985), p.421.

Lemma 3 *Let n and m go to infinity at the same rate so that n/m remains in a compact subset of $(0, \infty)$ and let $A^{(n,m)}$ and $B^{(n,m)}$ be two sequences of random $n \times m$ matrices such that $\sigma_1^2\left(A^{(n,m)} - B^{(n,m)}\right) = o_p(n^{-1/3})$ and $\sigma_1^2\left(B^{(n,m)}\right) = O_p(n)$. Then $\left|\sigma_k^2\left(A^{(n,m)}\right) - \sigma_k^2\left(B^{(n,m)}\right)\right| = o_p\left(n^{1/3}\right)$ uniformly over k.*

The statement of Lemma 2 easily follows from Weyl's inequalities for singular values of a sum of two matrices (see Horn and Johnson (1985), p.423). Let $\hat{\chi}$, $\Lambda(L)$ and $\hat{F}$ be $n \times m$, $n \times k$ and $k \times m$ matrices with j-th row and s-th column entries

Table 6: Eigenvalue statistics for the excess stock return data. The test for the number of approximate factors.l

i	1	2	3	4	5	6	7	8	9	10	11
γ_i	100	30.1	19.9	17.2	12.9	12.4	10.4	9.57	8.91	8.77	8.59
$\frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$	6.84	3.77	0.63	8.00	0.27	2.53	1.19	4.90	0.76		

$\hat{\chi}_{js} = \sum_{t=1}^{T} (X_{jt} - e_{jt}) e^{-i\omega_s t}/\sqrt{T}$, $\Lambda_{js}(L)$ and $\hat{F}_{js} = \sum_{t=1}^{T} F_{jt} e^{-i\omega_s t}/\sqrt{T}$, respectively.

Lemma 4 *Let n, m and T go to infinity so that $m = o\left(T^{3/7}\right)$ and n/m remains in a compact subset of $(0, \infty)$. Then, under Assumptions 1i and 2ii $\sigma_1^2 \left(\hat{\chi} - \Lambda(e^{-i\omega_0})\hat{F}\right) = o_p\left(n^{-1/3}\right)$.*

Proof: Let R and Q an $n \times m$ matrices with j, s-th entries $\hat{\chi}_{js} - \sum_{r=1}^{k} \Lambda_{jr}(e^{-i\omega_s})\hat{F}_{rs}$ and $\sum_{r=1}^{k} (\Lambda_{jr}(e^{-i\omega_s}) - \Lambda_{jr}(e^{-i\omega_0})) \hat{F}_{rs}$, respectively. A simple extension of Theorem 1 of Hannan (1970), p.248 which takes into account that n goes to infinity, implies that $E \sum_{j,s} |R_{js}|^2 = o\left(n^{-1/3}\right)$. Using Markov's inequality and Lemma 1, we get: $\sigma_1(R) = o_p\left(n^{-1/6}\right)$. Further, using Assumption 1i and Markov's inequality, it is easy to show that $\sum_{j,s} \left|\hat{F}_{js}\right|^2 = O_p(m)$. In addition, Assumption 2ii together with the assumption about the relative rates of growth of n, m and T and the definition of the frequencies ω_s, imply that $\|\Lambda(e^{-i\omega_s}) - \Lambda(e^{-i\omega_0})\| = o\left(n^{-5/6}\right)$ uniformly in s. Therefore, using Lemma 1, we obatin: $\sigma_1(Q) = o_p\left(n^{-1/3}\right)$. Now the statement of the lemma follows from the fact that $\sigma_1\left(\hat{\chi} - \Lambda(e^{-i\omega_0})\hat{F}\right) \leq \sigma_1(R) + \sigma_1(Q)$. QED

Let $\hat{e}$ be an $n \times m$ matrix with entries $\hat{e}_{js} = \sum_{t=1}^{T} e_{jt} e^{-i\omega_s t}/\sqrt{T}$ and let $\hat{e}_s$ be.the s-th column of $\hat{e}$.

Lemma 5 *Let n, m and T go to infinity so that n/m remains in a compact subset of $(0, \infty)$ and Assumptions 1ii, 2i, 3 hold. Then under either additional assumption i) or ii) of Theorem 1, there exists an $n \times m$ matrix $\tilde{e}$ with independent $N_C(0, 2\pi S_n^e(\omega_0))$ columns, independent from $\hat{F}$, and such that $\sigma_1^2(\hat{e} - \tilde{e}) = o_p\left(n^{-1/3}\right)$.*

Proof: Suppose additional assumption i) holds. Define a real zero-mean Gaussian vector $\eta \equiv \left((\operatorname{Re}\hat{e}_1)', (\operatorname{Im}\hat{e}_1)', \ ..., (\operatorname{Re}\hat{e}_m)', (\operatorname{Im}\hat{e}_m)'\right)'$. By Theorem 4.3.2 of Brillinger (2001) and by the fact, following from Assumption 2i, that $[S_n^e(\omega_s)]_{jr} = [S_n^e(\omega_0)]_{jr} + O(m/T)$ uniformly in j and r, we have: $E\eta\eta' = I_m \otimes \Omega + R$, where R is an $2nm \times 2nm$ matrix with $2n \times 2n$ blocks R_{ij} such that $R_{ij} = O(T^{-1})$ if $i \neq j$ and $R_{ij} =$

$O(m/T)+O(T^{-1})$ if $i=j$, and $\Omega=\begin{pmatrix}\pi\,\mathrm{Re}\,S_n^e(\omega_0) & -\pi\,\mathrm{Im}\,S_n^e(\omega_0)\\ \pi\,\mathrm{Im}\,S_n^e(\omega_0) & \pi\,\mathrm{Re}\,S_n^e(\omega_0)\end{pmatrix}$. Construct $\tilde{\eta}=(I_m\otimes\Omega)^{1/2}(I_m\otimes\Omega+R)^{-1/2}\eta$ and define an $n\times m$ matrix $\tilde{e}$ with s-th column $\tilde{e}_s$ so that $\left((\mathrm{Re}\,\tilde{e}_1)',(\mathrm{Im}\,\tilde{e}_1)',...,(\mathrm{Re}\,\tilde{e}_m)',(\mathrm{Im}\,\tilde{e}_m)'\right)'=\tilde{\eta}$. It is not difficult to show that $E(\tilde{\eta}-\eta)'(\tilde{\eta}-\eta)=\mathrm{tr}\left((I_m\otimes\Omega+R)^{1/2}-(I_m\otimes\Omega)^{1/2}\right)^2=o\left(n^{-1/3}\right)$, which together with Lemma 1 and Markov's inequality implies that $\sigma_1^2(\hat{e}-\tilde{e})=o_p\left(n^{-1/3}\right)$.

Now, suppose additional assumption ii) holds. Although vector η is no longer Gaussian, we still have $Var(\eta)=I_m\otimes\Omega+R$. Lemma 4i then implies that to prove part ii), it is enough to find a zero-mean Gaussian vector $\hat{\eta}$ such that $Var(\hat{\eta})=Var(\eta)$ and $\|\eta-\tilde{\eta}\|^2=o_p\left(n^{-1/3}\right)$.

According to Strassen's theorem (see Theorem 8, p.243 of Pollard, 2002), such a vector exists if and only if , for any positive ε and δ, $\pi\left(F,G;\varepsilon^{1/2}n^{-1/6}\right)\leq\delta$ for large enough n. Here F is the distribution of η, G is $N(0,I_m\otimes\Omega+R)$, and $\pi(F,G;\lambda)\equiv\sup_{A\in\mathcal{B}^r}\max\left\{F(A)-G\left(A^\lambda\right),G(A)-F\left(A^\lambda\right)\right\}$, where $A^\lambda=\{y\in R^r:\inf_{x\in A}\|x-y\|<\lambda\}$ and $\mathcal{B}^r$ is the Borel sigma-algebra on R^r ($r\equiv 2mn$ is the dimensionality of η).

Theorem 1.1 of Zaitsev (1987) provides an upper bound on $\pi\left(F,G;\varepsilon^{1/2}n^{-1/6}\right)$. Under the conditions of Lemma 4ii, this bound converges to zero as n,m and T diverge to infinity which completes the proof. The proof of the convergence of Zaitsev's bound to zero is tedious but straightforward. It is available from the Detailed Appendix to the paper posted at the author's website. QED

Let $A^{(1)}\equiv\begin{pmatrix}A_{11}^{(1)} & A_{12}^{(1)}\\ A_{21}^{(1)} & A_{22}^{(1)}\end{pmatrix}$ be a symmetric non-negative definite $n\times n$ matrix and let $A_{11}^{(1)}$ be its $k\times k$ block. Further, let $A\equiv\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}$ be an $n\times n$ matrix partitioned similarly to $A^{(1)}$ and such that A_{21},A_{12}, and A_{22} are zero matrices and $A_{11}=\mathrm{diag}(a_1,...,a_k)$, $a_1\geq a_2\geq...\geq a_k>0.$

Lemma 6 *Let* $A(\varkappa) = A + \varkappa A^{(1)}$ *and let* $r_0 = a_k/2$. *For real* $\varkappa$ *such that* $0 < \varkappa < r_0/\left\|A^{(1)}\right\|$ *and for any* $i = 1, 2, ..., n-k$, *we have*:

$$\left|\lambda_{k+i}(A(\varkappa)) - \varkappa\lambda_i\left(A_{22}^{(1)}\right)\right| \leq 3r_0 \frac{|\varkappa|^2 \left\|A^{(1)}\right\|^2}{(r_0 - |\varkappa| \left\|A^{(1)}\right\|)^2}$$

The proof of this lemma is rather technical. It is based on the properties of the expansion of the resolvent $(A(\varkappa) - zI_n)^{-1}$ in the power series of $\varkappa$. To save space, we allot the proof to the Detailed Appendix which is available from the author's website. Now, we are ready to prove Theorem 1.

Proof of Theorem 1:

Denote $\left[\hat{X}_1, ..., \hat{X}_s\right]$ as $\hat{X}$ and $\Lambda(e^{-i\omega_0})$ as $\hat{\Lambda}$. Lemmas 2, 3 and 4 imply that there exists matrix $\tilde{X} \equiv \hat{\Lambda}\hat{F} + \tilde{e}$, where $\tilde{e}$ and $\hat{F}$ are independent and $\tilde{e}$ has independent $N_C(0, 2\pi S_n^e(\omega_0))$ columns, and such that $\left|\lambda_j\left(\tilde{X}\tilde{X}'/m\right) - \lambda_j\left(\hat{X}\hat{X}'/m\right)\right| = o_p\left(m^{-2/3}\right)$ uniformly over j. Therefore, the asymptotic joint distribution of interest in Theorem 1 is the same as that of $\left\{\tilde{\sigma}_{m,n}^{-1}\left(\lambda_{k_0+i}\left(\tilde{X}\tilde{X}'/m\right) - \tilde{\mu}_{m,n}\right), i = 1, ..., r\right\}$. We will prove that this distribution is Tracy-Widom of type 2.

Assume that the $k \times m$ matrix of factors $\hat{F}$ has all but the first k columns zero. There is no loss of generality in such an assumption because we can always find a unitary Q such that $\hat{F}Q$ has the above form. Multiplying $\tilde{X} = \hat{\Lambda}\hat{F} + \tilde{e}$ from the right by Q does not change neither the eigenvalues of $\tilde{X}\tilde{X}'/m$ nor the joint complex Gaussian distribution of the elements of $\tilde{e}$.

Denote the matrix of the first k columns of $\hat{F}$ as $\hat{F}_{1:k}$, the matrix of the first k columns of $\tilde{e}$ as $\tilde{e}_1$, and the matrix of the last $m-k$ columns of $\tilde{e}$ as $\tilde{e}_2$. Then we can decompose $\tilde{X}\tilde{X}'/m$ into a sum of two terms: $\left(\hat{\Lambda}\hat{F}_{1:k} + \tilde{e}_1\right)\left(\hat{\Lambda}\hat{F}_{1:k} + \tilde{e}_1\right)'/m$ and $\tilde{e}_2\tilde{e}_2'/m$. Let $R'AR$ be a spectral decomposition of the first term. Since the first term has rank k, the diagonal matrix A can be chosen so that it has all but first k nonzero diagonal elements. We will denote these elements as $a_1 \geq a_2 \geq ... \geq a_k > 0$.

Define r_0 as $r_0 = a_k/(2n)$. Note that Assumptions 1 and 4 imply that $r_0^{-1} = o_p\left(n^{1/3}\right)$. Indeed, using Weyl's inequalities for singular values (see Horn and Johnson, 1985, p. 423), we have: $a_k^{1/2} \geq \lambda_k^{1/2}\left(\hat{\Lambda}\hat{F}\hat{F}'\hat{\Lambda}'/m\right) - \lambda_1^{1/2}\left(\tilde{e}_1\tilde{e}_1'/m\right)$. Further, $\lambda_1\left(\tilde{e}_1\tilde{e}_1'/m\right)$ is no larger than $\lambda_1\left(\tilde{e}\tilde{e}'/m\right)$, which by Lemma 1 is bounded in probability. On the contrary, $\lambda_k\left(\hat{\Lambda}\hat{F}\hat{F}'\hat{\Lambda}'/m\right)$ is bounded below by $\lambda_k\left(\hat{F}\hat{F}'/m\right)\lambda_k\left(\hat{\Lambda}'\hat{\Lambda}\right)$ which, by Assumptions 1 and 4, is diverging to infinity in probability faster than $n^{2/3}$. Hence, $r_0^{-1} = 2n/a_k$ may diverge to infinity but slower than $n^{1/3}$.

Let us denote $R \cdot \tilde{e}_2$ as $\bar{e}_2$. Further, let us denote the matrix of the last $n-k$ rows of $\bar{e}_2$ as $\bar{e}_{22}$. Then, we have: $R\tilde{X}\tilde{X}'R'/nm = A/n + (1/n)\left(\bar{e}_2\bar{e}_2'/m\right)$, and therefore, by Lemma 5:

$$\left|\lambda_{k+i}\left(R\tilde{X}\tilde{X}'R'/nm\right) - (1/n)\,\lambda_i\left(\bar{e}_{22}\bar{e}_{22}'/m\right)\right| \leq 3r_0\frac{(1/n)^2\left\|\bar{e}_2\bar{e}_2'/m\right\|^2}{\left(r_0 - (1/n)\left\|\bar{e}_2\bar{e}_2'/m\right\|\right)^2}. \tag{2}$$

Now, note that $\bar{e}_{22}\bar{e}_{22}'/m$ is a $W_n^{\mathbb{C}}\left(m-k, S_{n,22}^e(\omega_0)/m\right)$ matrix, where $S_{n,22}^e(\omega_0)$ is obtained from $S_n^e(\omega_0)$ by eliminating its first k rows and columns. It is straightforward to show that $S_{n,22}^e(\omega_0)$ satisfies an assumption analagous to Assumption 3 for $S_n^e(\omega_0)$ (a proof of this fact is available from the Detailed Appendix). Hence, by Lemma 1, there exist sequences of center and scale constants $\tilde{\mu}_{m,n}$ and $\tilde{\sigma}_{m,n} \sim m^{-2/3}$ (related to the spectral distribution of $S_{n,22}^e(\omega_0)$ in the same way as $\mu_{m,n}$ and $\sigma_{m,n}$ are related to the spectral distribution of $S_n^e(\omega_0)$) such that for any fixed r, $\tilde{\sigma}_{m,n}^{-1}\left(\lambda_i\left(\bar{e}_{22}\bar{e}_{22}'/m\right) - \tilde{\mu}_{m,n}\right)$ with $i = 1, ..., r$ jointly converge to the Tracy-Widom law of type 2.

Further, since $\bar{e}_2\bar{e}_2'/m$ is distributed as $W_n^{\mathbb{C}}\left(m-k, 2\pi S_n^e(\omega_0)/m\right)$, the norm $\left\|\bar{e}_2\bar{e}_2'/m\right\|$ is bounded in probability by Lemma 1. Finally, the fact that r_0^{-1} can diverge to infinity in probability only slower than $n^{1/3}$ and inequality (2) imply that $\left|\lambda_{k+i}\left(\tilde{X}\tilde{X}'/m\right) - \lambda_i\left(\bar{e}_{22}\bar{e}_{22}'/m\right)\right| = o_p\left(n^{-2/3}\right)$. Therefore, since $\tilde{\sigma}_{m,n} \sim m^{-2/3}$ and n/m remains in a compact subset of $(0, \infty)$, the random variables $\tilde{\sigma}_{m,n}^{-1}\left(\lambda_{k+i}\left(\tilde{X}\tilde{X}'/m\right) - \tilde{\mu}_{m,n}\right)$, $i = 1, ..., r$ have the same asymptotic joint distribution as $\tilde{\sigma}_{m,n}^{-1}\left(\lambda_i\left(\bar{e}_{22}\bar{e}_{22}'/m\right) - \tilde{\mu}_{m,n}\right)$, $i = 1, ..., r$. QED

Proof of Theorem 2:

The first part of Theorem 2 follows from Theorem 1. Therefore, we only need to show that $\max_{k_0 < i \leq k_{\max}} \left(\gamma_i - \gamma_{i+1}\right) / \left(\gamma_{i+1} - \gamma_{i+2}\right)$ diverges in probability to infinity when the true number of factors k satisfies $k_0 < k \leq k_{\max}$. Clearly, it is enough to show the divergence in probability of the ratio $\left(\gamma_k - \gamma_{k+1}\right) / \left(\gamma_{k+1} - \gamma_{k+2}\right)$.

Recall that $\gamma_i \equiv \lambda_i\left(\hat{X}\hat{X}'/m\right)$. As was shown in the proof of Theorem 1, $\left|\gamma_i - \lambda_i\left(\tilde{X}\tilde{X}'/m\right)\right| = o\left(n^{-2/3}\right)$ uniformly in i. Further, using Weyl's inequalities for singular values (see Horn and Johnson, 1985, p. 423), we have:

$$\lambda_i^{1/2}\left(\hat{\Lambda}\hat{F}\hat{F}'\hat{\Lambda}'/m\right) - \lambda_1^{1/2}\left(\tilde{e}\tilde{e}'/m\right) \leq \lambda_i^{1/2}\left(\tilde{X}\tilde{X}'/m\right) \leq \lambda_i^{1/2}\left(\hat{\Lambda}\hat{F}\hat{F}'\hat{\Lambda}'/m\right) + \lambda_1^{1/2}\left(\tilde{e}\tilde{e}'/m\right)$$

for $i = 1, ..., n$. First, take $i = k$ and consider the first inequality. By Lemma 1, $\lambda_1\left(\tilde{e}\tilde{e}'/m\right)$ is bounded in probability, and, as was shown in the proof of Theorem 1, $\lambda_k\left(\hat{\Lambda}\hat{F}\hat{F}'\hat{\Lambda}'/m\right)$ diverges to infinity in probability at least as fast as $n^{2/3}$. Therefore, $\lambda_k\left(\tilde{X}\tilde{X}'/m\right)$, and hence γ_k, diverges to infinity in probability as n, m and T rise. Now, take $i > k$. Then $\lambda_i\left(\hat{\Lambda}\hat{F}\hat{F}'\hat{\Lambda}'/m\right) = 0$. Therefore, $\lambda_i\left(\tilde{X}\tilde{X}'/m\right)$, and hence γ_i, is bounded in probability because $\lambda_1\left(\tilde{e}\tilde{e}'/m\right)$ is bounded in probability. Summing up, the numerator of $\left(\gamma_k - \gamma_{k+1}\right) / \left(\gamma_{k+1} - \gamma_{k+2}\right)$ diverges to infinity in probability and the denominator stays bounded in probability. Hence, the ratio diverges to infinity in probability. QED

References

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

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

[3] Bai, J. and S. Ng (2006) "Instrumental Variable Estimation in a Data Rich Environment", mimeo, University of Michigan

[4] Bai, Z.D. (1999) "Methodologies in spectral analysis of large dimensional random matrices, a review", Statistica Sinica, 9, 611-677

[5] 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.

[6] Boivin, J., Giannoni, M., and I. Mihov (2006) "Sticky Prices and Monetary Policy: Evidence from Disaggregated U.S. Data" mimeo, Columbia University.

[7] Boivin, J. and M. Giannoni (2006) "DSGE Models in a Data-Rich Environment" mimeo, Columbia University

[8] 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.

[9] Breitung, J. and Eickmeier, S. (2005) "Dynamic Factor Models", Deutsche Bundesbank Discussion Paper 38/2005.

[10] Cattell, R. B. (1966) "The Scree Test for the Number of Factors", Multivariate Behavioral Research, vol. 1, 245-76.

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

[12] 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

[13] Connor, G. and R. A. Korajczyk, (1986) "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." Journal of Financial Economics, 15, 373-394.

[14] 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

[15] Forni, M., Giannone, D., Lippi, M., and L. Reichlin (2005) "Opening the black box: identifying shocks and propagation mechanisms in VAR and factor models", IAP technical Reports Series TR #0482, Universite Catholique de Louvain

[16] 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.

[17] Giannone, D., and M. Lenza (2004) "The Feldstein - Horioka Fact", CEPR Discussion Papers 4610, C.E.P.R. Discussion Papers.

[18] Hallin, M., and Liska R. (2007). The generalized dynamic factor model: determining the number of factors. Journal of the American Statistical Association 102, 603-617.

[19] Hannan, E.J. (1970) Multiple Time Series. John Wiley and Sons, Inc. New York, London, Sydney, Toronto.

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

[21] Jacobs, J.P.A.M. and P.W. Otter (2005) "Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach." Forthcoming in Econometric Reviews. Special Issue on Information and Entropy Econometrics.

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

[23] Kapetanios, G. (2004) "A new method for determining the number of factors in factor models with large datasets", Queen Mary University of London Working Paper No. 525.

[24] Kapetanios, G. (2005) "A testing procedure for determining the number of factors in approximate factor models with large datasets", mimeo, Queen Mary University of London.

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

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

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

[28] Onatski, A. (2007) "The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices", to appear in Annals of Applied Probability.

[29] Lewbel, A. (1991) "The rank of demand systems: theory and nonparametric estimation", Econometrica, 59, 3, pp. 711-730.

[30] Ludvigson, S. and S. Ng (2005) "Macro Factors in Bond Risk Premia", mimeo, University of Michigan

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

[32] 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

[33] 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

[34] Stock, J. and Watson, M. (2002a) "Macroeconomic Forecasting Using Diffusion Indexes", Journal of Business and Economic Statistics, 20, pp. 147-162.

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

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

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

[38] Yao, T. (2006) "Dynamic Factors and the Source of Momentum Profits", forthcoming in Journal of Business and Economic Statistics.

[39] Zhang, C. (2006) "Testing the APT with Maximum Sharpe Ratio of Extracted Factors", mimeo, The Hong Kong University of Science and Technology