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

Alexei Onatski*

Economics Department, Columbia University

April 5, 2009

Abstract

In this paper we study high-dimensional time series that have the generalized dynamic factor structure. We develop a test of the null of k_0 factors against the alternative that the number of factors is larger than k_0 but no larger than $k_1 > k_0$. Our test statistic equals $\max_{k_0<k\leq k_1} \left(\gamma_k - \gamma_{k+1}\right) / \left(\gamma_{k+1} - \gamma_{k+2}\right)$, 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 test different hypotheses about the number of dynamic factors in macroeconomic time series and about 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.

*I am very grateful to Harald Uhlig and three anonymous referees for helpful comments and suggestions.

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, in the analysis of the profitability of trading strategies, in testing implications of the Arbitrage Pricing Theory, and in the analysis of bond risk premia. 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, the identification of global shocks, the analysis of price dynamics and the estimation of the dynamic stochastic general equilibrium models.

An important question to be addressed by any study which uses factor analysis is how many factors there are. The number of factors must be known to implement various estimation and forecasting procedures. Moreover, that number 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). For 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 some recent work in this area see Bai and Ng (2007) and Hallin

and Liska (2007)), 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 one 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, as many as 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 $k_0 < k \leq k_1$.

- First, compute the discrete Fourier transforms (dft's) $\hat{X}_j \equiv \sum_{t=1}^{T} X_t e^{-i\omega_j t}/\sqrt{T}$ 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 $j \neq k$, $s_j \neq 0$, $s_j \neq T/2$ and $\max_j |\omega_j - \omega_0| \leq 2\pi(m+1)/T$.

- Next, compute a test statistic $R \equiv \max_{k_0 < i \leq k_1} \frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$, 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 R is above a critical value given in Table I.

Table I: Critical values of the test statistic R.

The test's size	$k_1 - k_0$							
%	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. $k_0 < k \leq k_1$, where $k_0 = 3$ and $k_1 = 10$, is in the 5th row (counting from the bottom up) and 2nd column (counting from the right) of the table. It equals 8.29. A matlab code dinamico.m, which implements the test, is available from the Econometrica supplementary materials website.

We prove that our test statistic is asymptotically pivotal under the null hypothesis and that it explodes under the alternative. We find that its asymptotic distribution, 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, is a function of the Tracy-Widom distribution. The Tracy-Widom distribution (see Tracy and Widom, 1994) 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.

Although our assumption that T grows faster than n is not in the spirit of the recent literature which lets n grow as fast or even faster than T, our Monte Carlo analysis shows that the test works well even when n is much larger than T. In Section 3, we discuss possible theoretical reasons for this Monte Carlo finding.

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 $\hat{X}_j$. Therefore, the k largest eigenvalues $\gamma_1, ...\gamma_k$ of the sample covariance matrix of $\hat{X}_j$ should explode, whereas the rest of the eigenvalues should have the same asymptotic distribution as the largest eigenvalues in the zero-factor case. Now, the zero-factor case corresponds to $\hat{X}_j$'s being asymptotically complex normal, independent dft's 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 (2008)). 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. Under the alternative, R is larger than or equal to the ratio $\frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}}$, which explodes because γ_k explodes while γ_{k+1} and γ_{k+2} stay bounded.

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

The rest of the paper is organized as follows. Section 2 describes the model and

states our assumptions. Section 3 develops 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 discusses the choice of our test statistic and the workings of the test under potential misspecifications of the factor model. Section 8 concludes. Technical proofs are contained in the Appendix.

2 The model and assumptions

Consider 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):

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

where $\Lambda_{ij}(L)$ equals $\sum_{u=0}^{\infty} \Lambda_{ij}^{(u)} L^u$, and factor loadings $\Lambda_{ij}^{(u)}$, factors F_{jt}, and idiosyncratic terms e_{it} satisfy Assumptions 1, 2, 3, and 4 stated below. Suppose that we observe X_{it} with $i = 1, ..., n$ and $t = 1, ..., T$. Let us denote the data vector at time t, $(X_{1t}, ..., X_{nt})'$, as $X_t(n)$, and its idiosyncratic part, $(e_{1t}, ..., e_{nt})'$, as $e_t(n)$. Further, let $\chi_t(n) \equiv X_t(n) - e_t(n)$ be the systematic part of $X_t(n)$. Then,

$$\begin{aligned} X_t(n) &= \chi_t(n) + e_t(n) \text{ and} \\ \chi_t(n) &= \sum\nolimits_{u=0}^{\infty} \Lambda^{(u)}(n) F_{t-u}, \end{aligned}$$

where $F_t \equiv (F_{1t}, ..., F_{kt})'$ denotes the vector of factors at time t and $\Lambda^{(u)}(n)$ denotes the $n \times k$ matrix with i, j-th entry $\Lambda_{ij}^{(u)}$.

Assumption 1: *The factors F_t follow an orthonormal white noise process. 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 the orthogonality of $e_t(n)$ and F_s.

Assumption 2: i) *The factor loadings are such that* $\sum_{u=0}^{\infty} \left\|\Lambda^{(u)}(n)\right\| (1+u) = O\left(n^{1/2}\right)$, *where* $\left\|\Lambda^{(u)}(n)\right\|$ *denotes the square root of the largest eigenvalue of* $\Lambda^{(u)}(n)'\Lambda^{(u)}(n)$. **ii)** *The idiosyncratic terms are jointly Gaussian with autocovariances* $c_{ij}(u) \equiv Ee_{i,t+u}e_{jt}$ *satisfying* $\sum_u (1+|u|)\,|c_{ij}(u)| < \infty$ *uniformly in* $i,j \in \mathbb{N}$.

Assumption 2 implies regularity of the dft's of $\chi_t(n)$ and $e_t(n)$ local to the frequency of interest ω_0. Let $\hat{\Lambda}_s$ denote $\sum_{u=0}^{\infty} \Lambda^{(u)}(n)e^{-iu\omega_s}$, where ω_s, $s = 0, 1, ..., m$, are the frequencies defined in the Introduction. Further, for any process y_t, let $\hat{y}_s \equiv \sum_{t=1}^{T} y_t e^{-i\omega_s t}/\sqrt{T}$ denote the dft's of y_t at the frequencies ω_s. For example, $\hat{X}_s(n), \hat{F}_s$ and $\hat{e}_s(n)$ are the dft's of $X_t(n), F_t$ and $e_t(n)$, respectively.

Assumption 2i guarantees that, for T sufficiently larger than $\max(n,m)$, $\hat{\chi}_s(n)$ is well approximated by $\hat{\Lambda}_0\hat{F}_s$ for all $s = 1, ..., m$ (see Lemma 4 in the Appendix). Note that in the special case when $\Lambda_{ij}(L)$ do not depend on L so that $\Lambda_{ij}(L) = \Lambda_{ij}^{(0)}$, Assumption 2i holds if the loadings are bounded functions of i and j. Indeed, then Lemma 2 in the Appendix implies that $\|\Lambda_0(n)\| \leq \left(\sum_{i=1}^{n}\sum_{j=1}^{k}\left(\Lambda_{ij}^{(0)}\right)^2\right)^{1/2} = O\left(n^{1/2}\right)$. In this special case, $\hat{\chi}_s(n)$ equals $\hat{\Lambda}_0\hat{F}_s$ exactly.

Assumption 2ii guarantees that, for T sufficiently larger than $\max(n,m)$, the real and imaginary parts of vectors $\hat{e}_s(n)$ with $s = 1, ..., m$ are well approximated by i.i.d. Gaussian vectors (see Lemma 5 in the Appendix). To establish such an approximation without assuming the Gaussianity of e_{it}, we will need:

Assumption 2: iia) *Let* ε_{jt}, $j \in \mathbb{N}, t \in \mathbb{Z}$, *be i.i.d. random variables with* $E\varepsilon_{jt} = 0, E\varepsilon_{jt}^2 = 1$ and $\mu_p \equiv E(|\varepsilon_{jt}|^p) < \infty$ *for some* $p > 2$. *Further, let* u_{jt}, $j \in \mathbb{N}$, *be linear*

filters of ε_{jt} *so that* $u_{jt} = C_j(L)\varepsilon_{jt} = \sum_{k=0}^{\infty} c_{jk}\varepsilon_{j,t-k}$ *with* $\sup_{j>0} \sum_{k=0}^{\infty} k\,|c_{jk}| < \infty$. *We assume that the idiosyncratic terms in (1) can be represented by square summable linear combinations of such filters. That is,* $e_{it} = \sum_{j=1}^{\infty} A_{ij}u_{jt}$, *where* $\sup_{i>0} \sum_{j=1}^{\infty} A_{ij}^2 < \infty$.

Assumption 2iia allows for rich patterns of cross-sectional and temporal dependence of the idiosyncratic terms. In addition, it allows for substantial departures from the Gaussianity. For example, the moments of order higher than p of the idiosyncratic terms may be infinite.

The remaining two assumptions concern the asymptotic behavior of the spectral density matrices at frequency ω_0 of $e_t(n)$ and $\chi_t(n)$, which we denote as $S_n^e(\omega_0)$ and $S_n^\chi(\omega_0)$, respectively. Let $l_{1n} \geq ... \geq l_{nn}$ be the eigenvalues of $S_n^e(\omega_0)$. Denote by H_n the spectral distribution of $S_n^e(\omega_0)$, that is, $H_n(\lambda) = 1 - \frac{1}{n}\#\{i \leq n : l_{in} > \lambda\}$, where $\#\{\cdot\}$ 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* m/n *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.

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 $n \to \infty$. It is needed in the technical parts of the proofs of Lemma 1, Lemma 5 and Theorem

1. The inequality $\limsup l_{1n} c_{m,n} < 1$ is crucial for Onatski (2008), 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 where 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* $S_n^\chi(\omega_0)$ *diverges to infinity faster than* $n^{2/3}$.

Assumption 4 relaxes the standard requirement that the factors' cumulative effects on the cross-sectional units rise linearly in n (see, for example, assumption A4 in Hallin and Liska, 2007). Allowing for cumulative factor effects which grow slower than linearly in n is important, because such a slower growth better corresponds to the current practice of creating larger macroeconomic data sets by going to higher levels of disaggregation instead of adding variables carrying genuinely new information.

3 The test

Before we explain the workings of our test, let us recall the notions of the complex Gaussian and complex Wishart distributions. 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 $N_{2n}\left(\begin{pmatrix} \operatorname{Re}\beta \\ \operatorname{Im}\beta \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \operatorname{Re}\Sigma & -\operatorname{Im}\Sigma \\ \operatorname{Im}\Sigma & \operatorname{Re}\Sigma \end{pmatrix}\right)$. 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 dft of our data admit an approximate factor structure in the sense of Chamberlain and Rothschild (1983). That is, $\hat{X}_s(n) = \hat{\Lambda}_0 \hat{F}_s + \hat{e}_s(n) + R_s(n)$, 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 $\max(n, m)$.

Second, as is well known (see, for example, Theorem 4.4.1 of Brillinger, 1981), for fixed n and m, Assumption 2ii (or 2iia) and our choice of the frequencies ω_s, $s = 1, ..., m$, as multiples of $2\pi/T$ imply 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 $\hat{X}_s$ 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 $R \equiv \max_{k_0 < i \leq k_1} \frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$ explodes under the alternative but becomes well approximated in distribution by $\max_{0 < i \leq k_1 - k_0} \frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1} - \lambda_{i+2}}$ under the null, where λ_i is the i-th largest eigenvalue of a complex Wishart $W_n^{\mathbb{C}}(m, S_n^e(\omega_0)/m)$ matrix.

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 (2008) 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 $(0, \infty)$.

Lemma 1: (Onatski, 2008) *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 $\sigma_{m,n}^{-1}(\lambda_1 - \mu_{m,n}), ..., \sigma_{m,n}^{-1}(\lambda_r - \mu_{m,n})$ of matrix $W_n^{\mathbb{C}}(m, S_n^e(\omega_0)/m)$ weakly converges to the r-dimensional Tracy-Widom distribution of type 2.

The univariate Tracy-Widom law of type 2 (henceforth denoted as TW_2) 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 $s \to \infty$. It plays an important role in large random matrix theory because it is the asymptotic distribution of the scaled and normalized largest eigenvalue of a matrix 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 multivariate 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 multivariate TW_2 distribution. Our test statistic R is 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, for any positive integer r, if either Assumption 2ii holds and $m = o\left(T^{3/8}\right)$, or Assumption 2iia holds and $m = o\left(T^{1/2-1/p} \log^{-1} T\right)^{6/13}$, the joint distribution of $\sigma_{m,n}^{-1}\left(\gamma_{k+1} - \mu_{m,n}\right), ..., \sigma_{m,n}^{-1}\left(\gamma_{k+r} - \mu_{m,n}\right)$ weakly converges to the r-dimensional TW_2 distribution.*

The proof of the theorem is given in the Appendix. As the Monte Carlo analysis in the next section suggests, the rates required by Theorem 1 are sufficient but not necessary for the theorem to hold. Our test works well even when n is much larger than T. Two possible theoretical reasons for such a good performance are as follows. First, Lemma 1 can, probably, be generalized to cases when $n/m \to \infty$. El Karoui (2006) obtains such a generalization in the special spherical Wishart case. Second, although making n compatible with T hurts the quality of the Gaussian approximation for dft's, the Gaussianity is likely unnecessary for the Tracy-Widom asymptotics to work. It is because the Tracy-Widom limit appears to be universal for a class of

random matrices much wider than the class of complex Wishart matrices. The first important universality results were recently obtained by Soshnikov (2002) and Péché (2009). Further development of such results remains a challenge for mathematicians.

Theorem 2 formally states the properties of our test. Its proof is in the Appendix.

Theorem 2: *Under conditions of Theorem 1, if* $k = k_0$, *statistic* R *converges in distribution to* $\max_{0<i\leq k_1-k_0} \frac{\lambda_i-\lambda_{i+1}}{\lambda_{i+1}-\lambda_{i+2}}$, *where* $\lambda_1, ..., \lambda_j$ *are random variables with joint multivariate* TW_2 *distribution. In contrast, when* $k_0 < k \leq k_1$, R *diverges to infinity in probability. Therefore, our test of the null of* $k = k_0$ *vs. the alternative of* $k_0 < k \leq k_1$ *is consistent and has asymptotically correct size.*

4 A test for the number of approximate factors

In an important special case when $\Lambda_{ij}(L)$ 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). In such a case, if the idiosyncratic terms are Gaussian and independent over time, we can test hypotheses about the number of *approximate,* as opposed to *dynamic,* factors using a modified procedure which does not rely on the frequency domain transformation of the data.

First, obtain transformations $\tilde{X}_j$ of the original data 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, that is $\tilde{X}_j = X_j + \sqrt{-1}X_{j+T/2}$. Next, compute a test statistic $\tilde{R} \equiv \max_{k_0<i\leq k_1} \frac{\tilde{\gamma}_i-\tilde{\gamma}_{i+1}}{\tilde{\gamma}_{i+1}-\tilde{\gamma}_{i+2}}$, where $\tilde{\gamma}_i$ is the i-th largest eigenvalue of matrix $\frac{2}{T}\sum_{j=1}^{T/2} \tilde{X}_j\tilde{X}_j'$. Finally, reject the null of $k = k_0$ approximate factors in favor of the alternative that $k_0 < k \leq k_1$ if and only if $\tilde{R}$ is above a critical value given in Table I. It can be shown that this test procedure is valid if the following modifications of Assumptions 1 through 4 hold:

Assumption 1m: *The factors F_t follow a fourth-order zero-mean stationary process with non-degenerate variance, with autocovariances $\Gamma_{ij}(u) \equiv EF_{it}F_{j,t+u}$ decaying to zero as u tends to infinity, and with 4-th order cumulants $cum\left(F_{it_1}, F_{jt_2}, F_{rt_3}, F_{l0}\right)$ decaying to zero as $\max\left(|t_1|, |t_2|, |t_3|\right)$ tends to infinity. 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 $n^{2/3}$.*

We have the following theorem.

Theorem 3: *Under Assumptions 1m through 4m, statistic $\tilde{R}$ behaves in the same way as statistic R in Theorem 2. Our test of the null of $k = k_0$ approximate factors vs. the alternative of $k_0 < k \leq k_1$ is consistent and has asymptotically correct size.*

Proof of this theorem is similar to the proof of Theorem 2. It is available from the Supplementary Appendix posted at Econometrica's supplementary materials website.

5 A Monte Carlo Study

We approximate the 10-dimensional TW_2 distribution 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 TW_2 distribution.

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 this statistic under the null. In particular, the first eigenvalue of the sample covariance matrix may be substantially larger than the other eigenvalues even when data have no factors at all. This observation suggests that *ad hoc* methods of determining the number of factors based on visual inspection and separation of the eigenvalues of the sample covariance matrix into a group of "large" and a group of "small" eigenvalues may be misleading.

According to Theorems 1 and 2, the approximate asymptotic critical values of our test of $k = k_0$ versus $k_0 < k \leq k_1$ equal the corresponding percentiles of the empirical distribution of $\max_{0<i\leq k_1-k_0} (x_i - x_{i+1}) / (x_{i+1} - x_{i+2})$. Table I in the Introduction contains such percentiles for $k_1 - k_0 = 1, 2, ..., 8$.

5.1 Size-power properties of the test

To study the finite-sample properties of our test, we generate data from model (1) as follows. The idiosyncratic components e_{it} follow AR(1) processes both cross-sectionally and over time: $e_{it} = \rho_i e_{it-1} + v_{it}$ and $v_{it} = \rho v_{i-1t} + u_{it}$, where ρ_i are

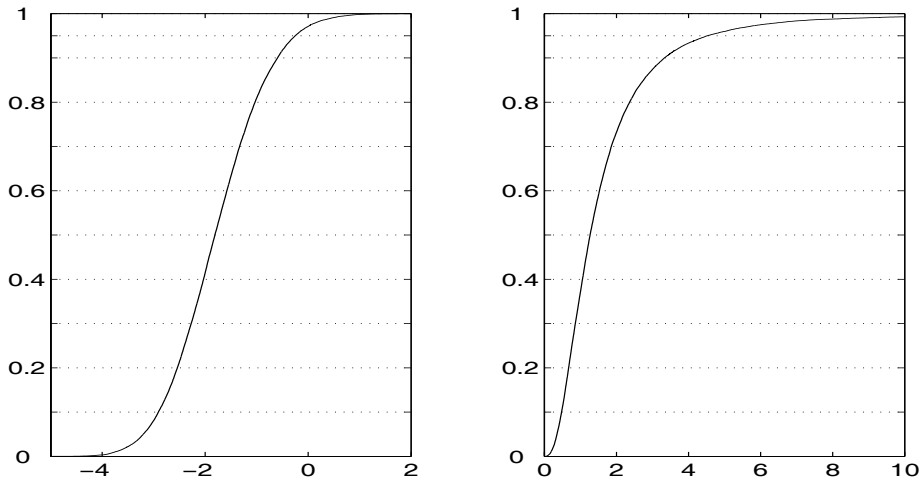

Figure 1: Left panel: CDF of the univariate TW_2 distribution. Right panel: CDF of the statistic $\frac{x_1-x_2}{x_2-x_3}$, where x_1, x_2, and x_3 have joint multivariate TW_2 distribution.

$iidU\left[-.8,.8\right]$, $\rho = 0.2$ and u_{it} are $iidN\left(0,1\right)$. The support $[-.8,.8]$ of the uniform distribution has been chosen to fit the range of the first-order autocorrelations of the estimated idiosyncratic components in the Stock and Watson (2002) dataset. The k-dimensional factor vectors F_t are $iidN\left(0,I_k\right)$. The filters $\Lambda_{ij}(L)$ are randomly generated (independently from F_t's and e_{it}'s) by each one of the following two devices:

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

We borrowed the AR loadings design, including the distributional assumptions on $b_{ij}^{(0)}, b_{ij}^{(1)}$ and $b_{ij}^{(2)}$ from Hallin and Liska (2007). An analysis of the partial autocorrelation function for the estimated systematic part of the Stock-Watson data shows that an AR(2) is a reasonable model for the systematic part of that data.

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 $0.4+0.05k$ and $1-(0.4+0.05k)$, respectively. Hence, for example, the factors explain 50% of the data variation in a 2-factor model and 75% of the data variation in a 7-factor model.

Figure 2 shows the p-value discrepancy plots and size-power curves[1] for our test of $H_0 : k = 2$ vs. $H_1 : 2 < k \leq 7$. The power of the test is computed when $k = 7$. The graphs are based on 8,000 replications of data with $(n, T) = (150, 500)$. Such a choice of n and T mimics the size of the Stock-Watson dataset. The dft's used by the test are computed at frequencies $\omega_i = 2\pi i/T$, $i = 1, ..., 65$, and therefore, $m = 65$. We checked that the figure does not change much for m on a grid 35:5:125. Note that, for our choice of $T = 500$ and $m = 65$, the interval $[\omega_1, \omega_m]$ covers a wide range of cycles: from $500/65 \approx 8$ periods per cycle, to 500 periods per cycle. Therefore, we cannot unambiguously name a specific frequency ω_0 which is approximated by $\omega_1, ..., \omega_m$. In the rest of the paper, instead of specifying a particular ω_0, we simply report the range of frequencies $\omega_1, ..., \omega_m$ used by the test.

As can be seen from Figure 2, the size of our test is only slightly larger than the nominal size and the power is around 100% for the test of actual sizes >1%. As a robustness check, we change the distribution of the idiosyncratic terms' innovations u_{it} from $N(0, 1)$ to the centered chi-squared distribution $\chi^2(1) - 1$ and to the Student's $t(5)$ distribution. For AR loadings, we find that the worst size discrepancy of the nominal 5% size test equals 0.027 and corresponds to the chi-squared case. For MA loadings, the worst size discrepancy of the nominal 5% size test equals 0.021 and corresponds to the $t(5)$ case. For both AR and MA loadings, we find no notable changes in the size-power curves.

[1] On the x-axis and y-axis of a p-value discrepancy plot, we have the nominal size of a test and the difference between the finite-sample size and the nominal size, respectively. A size-power curve is a plot of the power against the *true* size of a test.

Figure 2: Size-power properties of the test of $H_0 : k = 2$ vs. $H_1 : 2 < k \leq 7$. $(n, m, T) = (150, 65, 500)$.

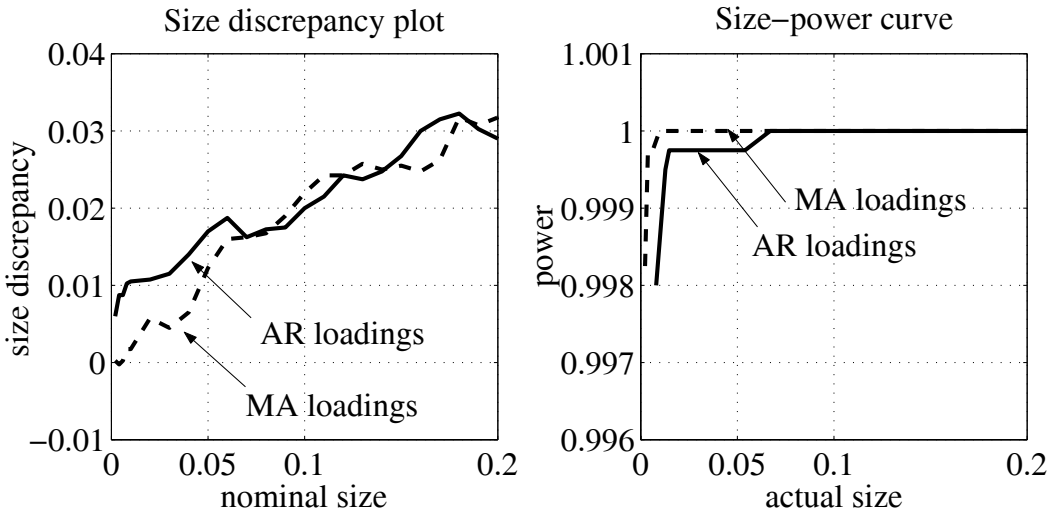

Table II describes the sensitivity to changes in the sample size and the sensitivity to the choice of the approximating frequencies of the size and power of the test of $H_0 : k = 2$ vs. $H_1 : 2 < k \leq 7$. Panels A and B of the table correspond to $\omega_s = 2\pi s/T$ and $\omega_s = 2\pi \left([T/2] + s - m - 1 \right) /T$, respectively. The first of the above choices of ω_s covers relatively low frequencies, while the second choice covers relatively high frequencies. Note that although the simulated data are scaled so that two factors explain 50% of the all-frequency-variation, the factors' explanatory power is larger at low frequencies and smaller at high frequencies. The raws of Table II which are labeled "% variation due to 2 factors" show how much of the variation of the simulated series X_{it} in the interval $[\omega_1, \omega_m]$ is due to its two-factor systematic component χ_{it}.[2]

As can be seen from the table, the size of the test remains good for most of the sample-size-frequency-range combinations. It remains good even in the cases when,

[2]More precisely, the numbers reported are: $100\% \cdot E\left(\int_{\omega_1}^{\omega_m} S^{\chi_{it}}(\omega)\, d\omega / \int_{\omega_1}^{\omega_m} S^{X_{it}}(\omega)\, d\omega \right)$, where the expectation is taken over the distribution of the parameters of our Monte Carlo setting.

Table II: The actual size of the nominal 5 percent size test and the power of the true 5 percent size test when k=7 for different sample sizes.

n	1000	250	150	500	100	250	70	150	70
T	250	1000	500	150	250	100	150	70	70
m	60	70	65	40	40	40	30	30	30
panel A: $\omega_s = 2\pi s/T$									
AR, size	6.4	6.1	6.7	6.8	7.6	8.2	8.9	8.4	8.7
AR, power	100	100	100	92	94	33	27	15	13
% variation due to 2 factors	67	83	76	63	73	51	69	45	46
MA, size	5.3	5.6	6.2	5.7	7.2	6.6	6.8	6.8	6.5
MA, power	100	100	100	100	100	99	86	81	51
% variation due to 2 factors	66	67	67	65	67	58	67	56	56
panel B: $\omega_s = 2\pi\left(\left[T/2\right] + s - m - 1\right)/T$									
AR, size	5.1	5.7	4.7	5.0	5.4	6.3	5.3	7.2	5.6
AR, power	4.4	4.5	4.7	4.9	4.9	5.6	5.1	5.8	6.4
% variation due to 2 factors	2.1	1.5	1.7	2.4	1.8	7.2	1.9	12	12
MA, size	5.6	4.7	5.6	5.4	5.9	6.7	6.2	7.2	7.1
MA, power	88	5.2	5.0	23	4.7	66	5.0	25	14
% variation due to 2 factors	26	9.6	14	31	17	46	21	48	48

contrary to the assumption of Theorem 1, n is larger than T. We discuss possible theoretical reasons for such a good performance in Section 3 above.

Not surprisingly, the power of the test strongly depends on the explanatory power of factors and on whether n is large or small. In the AR section of panel B, where factors explain only 2% of the data variation, the test does not have power at all. More important, as the table shows, the explanatory power of factors, and hence the power of the test, may strongly depend on the frequency range chosen. Such a dependence may be both advantageous and harmful for researchers.

On one hand, if there is *a priori* knowledge about the frequency content of the factors, the frequency range can be chosen so that the test is more powerful. Further, if the interest is in studying a particular frequency range, say, business cycle frequencies, the analysis will not be contaminated by factors operating at very dif-

ferent frequencies. On the other hand, the test may fail to detect the factors if the systematic variation happens to be low in the frequency range chosen.

One way to address this problem is to combine the tests at, say, N different frequencies. A natural combined test statistic would be $\max_{j=1,...,N} R_j$, where R_j are the test statistics of the separate tests. Assuming that N stays fixed as m, n and T go to infinity and that the different approximating frequency ranges converge to distinct frequencies, it is possible to show[3] that under the null, $\max_{j=1,...,N} R_j$ is asymptotically distributed as $\max_{j=1,...,N} \max_{0<i\leq k_1-k_0} \frac{\lambda_{ji}-\lambda_{j,i+1}}{\lambda_{j,i+1}-\lambda_{j,i+2}}$, where $\{\lambda_{j1}, ..., \lambda_{j,k_1-k_0+2}\}$, $j = 1, ..., N$, have independent joint multivariate TW_2 distributions.

Unfortunately, the choice of the N distinct frequencies for such a combined test remains *ad hoc*. To avoid arbitrary choices, it would be better to combine the tests at all frequencies. However, then the separate test statistics R_j will no longer be asymptotically independent and the asymptotic distribution of the combined test statistic will not be as above. We leave the formal analysis of combined test procedures for future research.

In the remaining two Monte Carlo experiments, we focus on the sample size $(n, T, m) = (150, 500, 65)$ and on the frequency range $\omega_s = 2\pi s/T$, $s = 1, ..., m$. First, we explore in more detail the sensitivity of the size and power of our test to changes in the proportion of the systematic variation in the data. We re-normalize the systematic and idiosyncratic components so that their variances equal $\alpha(0.4 + 0.05k)$ and $1 - \alpha(0.4 + 0.05k)$, respectively. For $k = 2$ factors, our benchmark choice $\alpha = 1$ corresponds to 50% of the all-frequency variation being systematic. For $\alpha = 0.2$, the proportion of the systematic variation drops to 10%. When α decreases from 1 to 0.8,0.6,0.4 and 0.2, the proportion of the systematic variation at the frequency range

[3] A proof would be based on the fact that the dft's of the idiosyncratic terms corresponding to the different frequency ranges can be simultaneously well approximated by *independent* complex Gaussian vectors.

$[\omega_1, \omega_m]$ decreases from 76 to 70, 61, 50 and 34% for AR loadings, and from 67 to 59, 50, 39 and 24% for MA loadings, respectively.

We find that the size of our test is not sensitive to the choice of α, but the power of the test is. For AR loadings and $\alpha = 1, 0.8, ..., 0.2$, the power of the test of actual size 5% equals 100, 100, 97, 58 and 8%, respectively. For MA loadings and $\alpha = 1, 0.8, ..., 0.2$, the power equals 100, 100, 96, 39 and 5%, respectively.

In our last Monte Carlo experiment, we study the effect of increasing cross-sectional correlation in idiosyncratic terms. We find that increasing parameter ρ in the equation $\upsilon_{it} = \rho\upsilon_{i-1,t} + u_{it}$ up to $\rho = 0.8$ does not notably affect neither size nor power of the test. However, for very large values of ρ, the size and the power deteriorate. For $\rho = 0.98$, the actual size of the nominal 5% size test equals 10.5 and 10.6% for AR and MA loadings, respectively, while the power of the actual 5% size test completely disappears. For $\rho = 0.95$, the actual size equals 8.6 and 7.8%, while the power equals 13.3 and 7.7% for AR and MA loadings, respectively.

In addition to the above Monte Carlo exercises, we did a Monte Carlo analysis of the approximate factor model version of our test. The analysis shows that its size and power properties are very good. To save space, we will not report these results here.

5.2 Comparison to other tests

To the best of our knowledge, there are no alternatives to our test in the general generalized dynamic factor setting.[4] However, in the special case of the approximate factor models, one can also use tests proposed by Connor and Korajczyk (1993) and Kapetanios (2005). Let us briefly describe these alternatives.

[4] Jacobs and Otter (2008) develop a test that can be used when the dynamic factor loadings $\Lambda_{ij}(L)$ are lag polynomials of known finite order r. This test uses fixed-n/large-T asymptotics. Although it works well for $n < 40$, it is very strongly oversized for $n \geq 70$.

Connor and Korajczyk (1993) test the null of $p = p_0$ approximate factors against the alternative that $p = p_0 + 1$. Their test uses 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.

Kapetanios (2005) tests the null of $p = p_0$ approximate factors against the alternative that $p_0 < p \leq p_1$. He employs a subsampling method to approximate the asymptotic distribution of a test statistic $\lambda_{p_0+1} - \lambda_{p_1+1}$, 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 $\lambda_{p_0+1} - \lambda_{p_1+1}$ which converges in distribution to some unknown limit law, about properties of such a law, and about the functional form of the scaling constant.

To compare our test with the Connor-Korajczyk and the Kapetanios tests[5], we simulate data from an approximate factor model $X_{it} = \sum_{j=1}^{k} \Lambda_{ij} F_{jt} + e_{it}$, where Λ_{ij} are i.i.d. $N(0,1)$; $F_{jt} = 0.85 F_{j,t-1} + \varepsilon_{jt}$ with ε_{jt} i.i.d. $N(0,1)$; e_{it} are as in Section 5.1 with $\rho_i \sim U\left[-0.8, 0.8\right]$, $\rho = 0.2$ or, alternatively, $\rho = 0.7$; and $\sum_{j=1}^{k} \Lambda_{ij} F_{jt}$ and e_{it} are scaled to have equal variances for each i. Table III reports the actual size of the nominal 5% size tests and the power of the actual 5% size tests for the hypothesis of 2 vs. 3 approximate factors. The size and power computations are based on 10,000 Monte Carlo replications. The row labeled "Onatski (dynamic)" corresponds to our dynamic factor test with $\omega_s = 2\pi s/T$, $s = 1, ..., 30$. The row labeled "Onatski (approximate)" corresponds to our approximate factor test described in Section 4.

As can be seen from the table, our tests are well-sized for both relatively weak ($\rho = 0.2$) and relatively strong ($\rho = 0.7$) cross-sectional correlation of the idiosyncratic terms. The power of our tests is always greater than 90%, except for the case of $n = T = 70$ and $\rho = 0.7$, when it drops to about 60%. The size and power properties

[5] We are grateful to George Kapetanios for sharing his codes with us.

Table III: Size-power properties of the alternative tests

	size (in %)			power (in%)		
n	70	150	70	70	150	70
T	70	70	150	70	70	150
	$\rho = 0.2$					
Onatski (dynamic)	7.21	8.37	7.09	94.9	99.2	100
Onatski (approximate)	6.05	6.21	6.21	96.7	100	100
Kapetanios	10.8	12.7	6.86	100	100	100
Connor-Korajczyk	22.5	24.0	23.2	76.7	89.5	99.6
	$\rho = 0.7$					
Onatski (dynamic)	7.01	7.09	7.03	57.5	93.1	94.9
Onatski (approximate)	6.52	6.44	5.90	61.0	94.3	94.5
Kapetanios	24.3	19.3	17.6	100	100	100
Connor-Korajczyk	44.5	37.4	62.4	33.6	59.8	69.5

of the Connor-Korajczyk test are clearly worse than those of our tests. The power of the Kapetanios test is excellent. However, the test is substantially more over-sized than our tests. The size distortion for the Kapetanios test becomes very large for the case of the relatively strong correlation of the idiosyncratic terms.

5.3 Using the test to determine the number of factors

Although our test's primary objective is testing hypotheses about the number of factors, it can also be used to determine the number of factors in a data set as follows. Suppose that it is known *a priori* that $k_1 \leq k \leq k_2$. Using our test of asymptotic size α, test $H_0 : k = k_1$ vs. $H_1 : k_1 < k \leq k_2$. If H_0 is not rejected, stop. The estimated number of factors is k_1. If H_0 is rejected, test $H_0 : k = k_1 + 1$ vs. $H_1 : k_1 + 1 < k \leq k_2$. Repeat the procedure until H_0 is not rejected and take the corresponding number of factors as the estimate. The estimate equals the true number of factors with probability approaching $1 - \alpha$ as n grows.

In the Monte Carlo simulations below, we choose $k_1 = 1$, $k_2 = 4$ and α equal to the

maximum of 0.01 and the p-value of the test of $H_0 : k = 0$ vs. $H_1 : 0 < k \leq k_1$. That is, α is calibrated so that our test has enough power to reject the *a priori* false null of $k = 0$. Our Monte Carlo design is the same as in Section 5.1, except we now magnify the idiosyncratic part of the simulated data by $\sigma \geq 1$, which makes the determination of the number of factors a more difficult problem.

Table IV: Percentage of 500 Monte Carlo replications resulting in estimates 1, 2, 3 or 4 of the true number of factors $k = 2$ according to different estimators. The estimators are constrained to be in the range from 1 to 4.

			Onatski				Hallin-Liska				Bai-Ng			
		$\hat{k} =$	1	**2**	3	4	1	**2**	3	4	1	**2**	3	4
n	T	σ^2	MA loadings											
70	70	1	0	**100**	0	0	0	**100**	0	0	1	**99**	0	0
70	70	2	0	**100**	0	0	0	**100**	0	0	1	**99**	0	0
70	70	4	12	**80**	6	2	85	**15**	0	0	52	**48**	0	0
100	120	1	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0
100	120	3	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0
100	120	6	3	**94**	3	0	31	**69**	0	0	56	**44**	0	0
150	500	1	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0
150	500	8	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0
150	500	16	2	**96**	2	0	49	**51**	0	0	100	**0**	0	0
n	T	σ^2	AR loadings											
70	70	1	16	**77**	6	1	12	**88**	0	0	21	**45**	32	2
70	70	2	29	**66**	4	1	18	**82**	0	0	26	**74**	0	0
70	70	4	62	**27**	7	4	88	**12**	0	0	81	**19**	0	0
100	120	1	0	**95**	4	1	0	**100**	0	0	79	**16**	4	1
100	120	3	10	**81**	7	2	4	**96**	0	0	11	**89**	0	0
100	120	6	35	**51**	8	6	42	**58**	0	0	70	**30**	0	0
150	500	1	0	**97**	2	1	0	**100**	0	0	88	**12**	0	0
150	500	8	0	**98**	1	1	0	**100**	0	0	1	**99**	0	0
150	500	16	5	**83**	7	5	6	**94**	0	0	96	**4**	0	0

We compare the above estimator to the Bai-Ng (2007) and the Hallin-Liska (2007) estimators with the following parameters (denoted as in the corresponding papers). For the Bai-Ng estimator, we use $\hat{D}_{1,k}$ statistic for the residuals of VAR(4), set the maximum number of static factors at 10 and consider $\delta = 0.1$ and $m = 2$.

For the Hallin-Liska estimator, we use the information criterion $IC^T_{2;n}$ with penalty $p_1(n,T)$, set the truncation parameter M_T at $\left[0.7\sqrt{T}\right]$ and consider the subsample sizes $(n_j, T_j) = (n - 10j, T - 10j)$ with $j = 0, 1, ..., 3$ so that the number of the subsamples is $J = 4$. We chose the penalty multiplier c on a grid $0.01 : 0.01 : 3$ using Hallin and Liska's second "stability interval" procedure.

To make our estimator comparable to the alternatives, which impose the correct restriction that the number of factors is the same across different frequencies, we compute our estimates for frequencies on a grid in the $[0, \pi]$ range.[6] We then weight these frequencies proportionally to their factor information content, as measured by the square of the sum of the four largest eigenvalues of the corresponding estimate of the data's spectral density. Our final estimate of the number of factors equals the estimate which agrees for most of the frequencies on the grid counted according to their weights.

Table IV reports the percentages of 500 Monte Carlo replications that deliver 1, 2, 3 and 4 estimated number of factors $\hat{k}$. The best possible estimator has $\hat{k} = 2$ for 100% of the replications. For AR loadings, the Hallin-Liska estimator performs the best. Our estimator is almost as good as the Hallin-Liska estimator and is much better than the Bai-Ng estimator, which tends to considerably underestimate the number of factors for the smallest and the largest σ^2 considered. For MA loadings and relatively less noisy data (relatively small σ^2), all estimators work very well. For very noisy data, however, our estimator outperforms both the Hallin-Liska and the Bai-Ng estimators, which tend to underestimate the number of factors.

The reason why the excellent performance of the Hallin-Liska estimator deterio-

[6] More precisely, we consider a linear grid of $\left[0.7\sqrt{T}\right]$ frequencies in $\left[0, \pi - \frac{2\pi m}{T}\right]$ range. Each frequency ω_0 on the grid is approximated by frequencies of the form $\omega_0 + \frac{2\pi}{T}j$ with $j = 1, ..., m$. We set $m = 30, 40$ and 65 for the data sizes $(n, T) = (70, 70), (100, 120)$ and $(150, 500)$, respectively.

rates as the data gets noisier is as follows. The Hallin-Liska automatic determination of the penalty multiplier utilizes the fact that, as long as the multiplier is too small or too large, the estimated number of factors should change from a wrong number of factors to the true number as n increases, whereas if the multiplier is appropriate, the estimate should remain equal to the true number of factors. However, for very noisy data, even if the multiplier is chosen correctly, the stabilization of the estimator at the true number of factors requires n be very large. Hence, choosing relatively small n in our MC experiments leads to the breakdown in the automatic determination procedure.

One possible remedy is to make the subsample sizes used by the Hallin-Liska estimator larger. For example, instead of $n_j = n - 10j$ and $T_j = n - 10j$, consider $n_j = n-3j$ and $T_j = n-3j$. Such a change indeed improves the quality of the estimator in the noisy data cases. For MA loadings and $n = T = 70$, the estimator's percentage of correct answers increases from 15% to 89% for $\sigma^2 = 4$. At the same time, the quality of the estimator for $\sigma^2 = 1$ and $\sigma^2 = 2$ remains very good. It overestimates the true number of factors only in 2% of the corresponding MC experiments.

6 Application

In this section, we test different hypotheses about the number of dynamic factors in macroeconomic time series and about the number of dynamic factors driving excess stock returns. The literature is full of controversy about the number of dynamic factors driving macroeconomic and financial data. Stock and Watson (2005) estimate seven dynamic factors in their data set. Giannone *et al.* (2005) find evidence supporting the existence of only two dynamic factors. Very recently, Uhlig (2008) in his discussion of Boivin et al. (2008) demonstrates that in the European data,

there might be no common factors at all, and that the high explanatory power of few principal components of the data reported by Boivin et al. (2008) may be an artifact of the high persistence of the individual time series in that data. Similarly, previous studies of factors driving excess stock returns often do not agree and report from 1 to 6 such factors (see Onatski (2008) for a brief review of available results).

In this section we consider two macroeconomic and one financial datasets. The first macroeconomic dataset is the same as in Stock and Watson (2002). It includes $n = 148$ monthly time series for the US 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). The second macroeconomic dataset is a subset of the data from Boivin et al. (2008). It includes $n = 243$ quarterly time series for Germany, France, Italy, Spain, the Netherlands and Belgium (33 series per country plus 12 international series) from 1987:Q1 to 2007:Q3 ($T = 83$). The series were regressed on current oil price inflation and short term interest rates as in Boivin et al. (2008) and the standardized residuals were taken as the data to which we apply our test. We are grateful to the authors of Stock and Watson (2002) and Boivin et al. (2008) for sharing their datasets with us.

Our financial dataset is provided by the Center for Research in Security Prices (CRSP). It includes $n = 972$ monthly excess returns on stocks traded on the NYSE, AMEX, and NASDAQ during the entire period from January 1983 to December 2006. Since previous empirical research suggests that the number of common risk factors may be different in January and non-January months, we drop January data, which leaves us with $T = 264$.

For our test, we use the dft's of the Stock-Watson data at frequencies $\omega_j = 2\pi s_j/480$ with $s_j \in \{4, ..., 40\}$; the dft's of the Boivin-Giannoni-Mojon data at frequencies $\omega_j = 2\pi s_j/83$ with $s_j \in \{2, ..., 30\}$; and the dft's of the stock return data

at frequencies $\omega_j = 2\pi s_j/264$ with $s_j \in \{1, ..., 60\}$. We used two criteria to make the above choices of the approximating frequencies. First, for the macroeconomic data we wanted to include the business cycle frequencies, but exclude cycles longer than 10 years. Second, we wanted to have at least 30 approximating frequencies, so that our asymptotic analysis may apply.

Table V reports the values of the ratio $\frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$ for different i. Using these values a reader can compute the R statistic for the test of her favorite null against her favorite alternative. Then, she can use the critical values in Table I to perform the test.

Table V: Values of the ratio $(\gamma_i - \gamma_{i+1})/(\gamma_{i+1} - \gamma_{i+2})$ for different i.

ω_0 (in years per cycle)	i						
	1	2	3	4	5	6	7
Stock-Watson data							
$\omega_0 \in [1, 10]$	9.90	5.01	0.75	3.44	1.78	0.75	1.28
Boivin-Giannoni-Mojon data							
$\omega_0 \in (\frac{8}{12}, 10\frac{4}{12}]$	3.76	1.64	1.14	2.17	0.74	5.44	0.39
Stock return data							
$\omega_0 \in (\frac{4}{12}, 22]$	7.43	3.20	1.97	0.96	2.00	1.00	2.37

For example, an interesting hypothesis to test for the stock return data is $H_0 : k = 1$ against $H_1 : 1 < k \leq 3$. The value of our R statistic for this test is the maximum of $\frac{\gamma_2 - \gamma_3}{\gamma_3 - \gamma_4}$ and $\frac{\gamma_3 - \gamma_4}{\gamma_4 - \gamma_5}$, which equals 3.20. Comparing this value to the critical values from Table I, we find that we can not reject the null with the 15% size test.

A particularly interesting hypothesis for the Stock-Watson macroeconomic data is that of 2 versus 3-7 dynamic factors. The value of the R statistics for such a test is 3.44. We cannot reject the null with the 15% size test. At the same time, the null of zero factors vs. the alternative of one or two factors can be rejected with 2% size test.

The hypothesis of zero factors is especially interesting for the Boivin-Giannoni-Mojon data (we are grateful to Harald Uhlig for suggesting to us to analyze these

data). As mentioned above, Uhlig (2008) points out that the high explanatory power of few principal components extracted from these data may be purely an artifact of the high temporal persistence of the data. The data are very persistent because they represent year-to-year growth rates of macroeconomic indicators (in contrast, the Stock-Watson data mostly represent the monthly growth rates).

For the Boivin-Giannoni-Mojon data, we cannot reject the null of zero factors vs. any of the alternatives: $1, 1$ to $2, ..., 1$ to 7 dynamic factors by the 5% size test. The null of zero factors vs. the alternative of one or two factors can be rejected only by the 15% size test. The p-values for the test of zero factors vs. 1 to $3, ..., 1$ to 7 alternatives are even larger.

The large temporal persistence of the idiosyncratic terms in the Boivin-Giannoni-Mojon data takes us somewhat outside the setting of the Monte Carlo experiments in Section 5. To check the finite sample performance of our test in the high-persistence environment, we simulate 8,000 datasets as in Uhlig (2008) (we thank Harald Uhlig for sharing his codes with us). These datasets consist of 243 *independent* times series (hence, no common factors exist) of length 83 with the first-order autocorrelations equal to the first-order autocorrelations of the corresponding Boivin-Giannoni-Mojon series.

The data were transformed exactly as the Boivin-Giannoni-Mojon data and the tests of zero vs. $1, 1$ to $2, ..., 1$ to 5 dynamic factors were performed. The actual sizes of our 5% nominal size tests (based on 8,000 simulations) were 7.9, 7.9, 8.3, 8.4 and 8.4% for the alternatives of 1, 1 to 2,...,1 to 5 factors, respectively. To assess the power of the test to detect the five factors extracted by Boivin et al. (2008), we add the five-factor systematic components estimated as in that paper to the series simulated as above (normalized to match the standard deviations of the estimated idiosyncratic components). We find that the power of the nominal 5, 10 and 15%

tests of $H_0 : k = 0$ vs. $H_1 : 0 < k \leq 5$ equals 51, 75 and 88%, respectively.

We conclude that there is only very mild evidence that there are one or two common dynamic factors in the Boivin-Giannoni-Mojon data. If such factors exist, they might explain less variation in the data than suggested by the high explanatory power of the few principal components extracted from the data.

7 Discussion

Our test can be based on other statistics than $R \equiv \max_{k_0<i\leq k_1} \frac{\gamma_i - \gamma_{i+1}}{\gamma_{i+1} - \gamma_{i+2}}$. Indeed, under the assumptions of Theorem 1, any function $f\left(\gamma_{k_0+1}, ..., \gamma_{k_0+r}\right)$ which is invariant with respect to scaling and normalizing of its arguments converges in distribution to a function of the r-dimensional TW_2 under the null. It is not difficult to see that, except for a zero probability domain where some γ_i coincide, all such functions have form $g(\rho)$, where $\rho \equiv \left\{ \frac{\gamma_{k_0+i} - \gamma_{k_0+i+1}}{\gamma_{k_0+i+1} - \gamma_{k_0+i+2}}, i = 1, ..., r-2 \right\}$.

A theoretically sound choice of $g(\rho)$ would be the ratio of the likelihood under the alternative to the likelihood under the null, $L_1(\rho)/L_0(\rho)$. However, the alternative does not specify *any* information about the first $k - k_0$ elements of ρ except that $\frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}}$ tends to infinity. A feasible choice of $g(\rho)$ would be a pseudo-likelihood ratio when $L_1(\rho)$ is an "improper density" that diverges to infinity as any of its first $k_1 - k_0$ arguments diverges to infinity.

We tried a test based on $L_1(\rho)/L_0(\rho)$ with $L_1(\rho) = R$. We estimated $L_0(\rho)$ using kernel methods and a simulated sample from multivariate TW_2 distribution. The test has comparable size and somewhat better power properties than our benchmark test. However, the necessity of estimating $L_0(\rho)$ makes the test much more complicated than the benchmark, and thus less attractive from a practical perspective.

Our test is one of many possible "simple" tests based on $g(\rho) = R$. We have tried a

variety of choices of $g(\rho)$ in addition to $g(\rho) = R$. For example, we considered $g(\rho) = \tilde{R} \equiv \max_{k_0<i\leq k_1} \frac{w_i \sum_{j=k_0+1}^{i} \gamma_j - (1-w_i) \sum_{j=i+1}^{k_1+1} \gamma_j}{\gamma_{i+1} - \gamma_{k_1+2}}$, where $w_i = \left(1 - \frac{i-k_0}{k_1-k_0+1}\right)$. In contrast to R, which maximizes "local" measures of the curvature of the scree plot at would-be kink points under the alternative, statistic $\tilde{R}$ maximizes wider measures of the curvature. However, none of the alternative statistics which we tried unambiguously outperformed R in terms of the size and power of the corresponding test.

Now we would like to address another important issue: it is unlikely that the factor model with few common factors literally describes the data. For example, macroeconomic data often have a clear block structure, with blocks corresponding to different categories of economic indicators, such as prices or output measures, included in the data set. Such data are perhaps better described by a factor model with few strong factors and many "weaker" factors corresponding to the different data categories.

Although a comprehensive analysis of such a model is beyond the scope of this paper, we would like to point out two facts. First, as shown in Onatski (2009), when the explanatory power of a factor is not much larger than that of the idiosyncratic components, it is impossible to empirically distinguish between the two. More important, a situation with few dominant and *many* "weaker" factors can alternatively be modeled as a situation where there are few factors and the idiosyncratic components are strongly influential.

We therefore expect that, if the distinction between strong and weak factors is not sharp, our test will work as if there were no factors at all. In contrast, if the distinction is sharp, the test will work as if the true number of factors were equal to the number of the strong factors.

To check this intuition, we generate data with two strong factors, each of which explains $s \cdot 100\%$ of the data's variation, and ten weak factors, each of which explains

$w \cdot 100\%$ of the data's variation, where $s > w$. We consider three (s, w) pairs: $(1/6, 1/30)$, $(1/8, 1/24)$, and $(1/12, 1/20)$. Figure 3 plots the power (based on 2,000 MC replications of the data with MA loadings) against the nominal size of the test of $H_0 : k = 0$ vs. $H_1 : 0 < k \leq 2$ (left panel) and of the test of $H_0 : k = 2$ vs. $H_1 : 2 < k \leq 4$ factors (right panel).

As expected, when the explanatory power of the two strong factors is much larger than that of the weak factors (solid line), the test has a lot of power against the null of zero factors but no power against the null of two factors. When the relative strength of the strong factors fades (dash line corresponds to $(s, w) = (1/8, 1/24)$ and dotted line corresponds to $(s, w) = (1/12, 1/20)$), the power against the null of zero factors gradually disappears.

8 Conclusion

In this paper, we have developed a test of the null hypothesis that there are k_0 factors, versus the alternative that there are more than k_0 but less than k_1+1 factors in a generalized dynamic factor model. Our test is based on the statistic $\max_{k_0<k\leq k_1} \frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}}$, 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 have shown that the test statistic is asymptotically pivotal, and has a distribution related to the Tracy-Widom distribution of type 2. We have tabulated percentiles of this distribution which provide us with critical values for our test.

Our Monte Carlo analysis shows that the new test has good finite-sample properties. It outperforms the Connor-Korajczyk (1993) test and is better sized than the Kapetanios (2005) test for a variety of finite sample situations. In addition to its primary use for testing different hypotheses about the number of factors, our test can

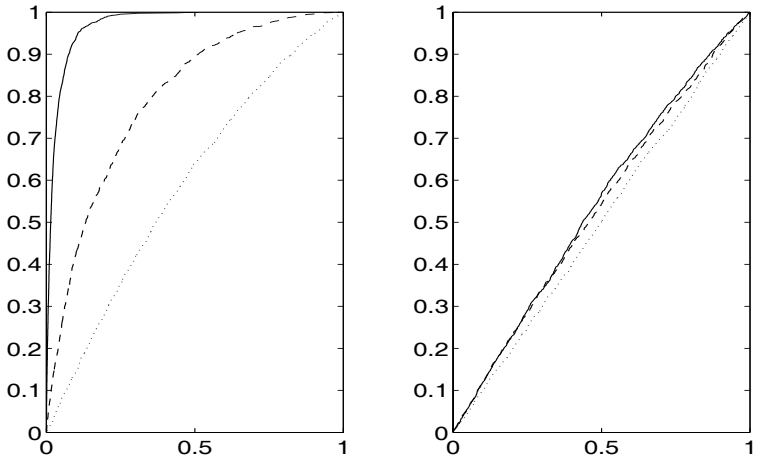

Figure 3: Power plotted against the nominal size. Left panel: the test of $H_0 : k = 0$ vs. $H_1 : 0 < k \leq 2$. Right panel: the test of $H_0 : k = 2$ vs. $H_1 : 2 < k \leq 4$. Dotted, dash and solid lines in that order correspond to the increasing explanatory power of the two strong factors.

be used to determine the number of factors in a dataset. We find that an estimator based on our test performs better than the Bai-Ng (2007) estimator and on par with the Hallin-Liska (2007) estimator.

We have applied our test to test various hypotheses about the number of factors in the macroeconomic datasets of Stock and Watson (2002) and Boivin et al. (2008) and in the data on excess stock returns provided by the Center for Research in Security Prices. For the Stock-Watson data, we have not been able to reject the null of two dynamic factors, in favor of the alternative that there are more than two but no more than seven dynamic factors. For the Boivin-Giannoni-Mojon data, we have found only very mild evidence that there exists one or two common dynamic factors. For the stock return data, we have not been able to reject the hypothesis of one dynamic factor in favor of the hypothesis of two or three dynamic factors.

9 Appendix

Denote the i-th largest singular value of a matrix A, defined as the square root of the i-th largest eigenvalue of AA', as $\sigma_i(A)$. Further, denote the square root of $\mathrm{tr}\,(AA')$ as $\|A\|_2$. Note that $\|A\| = \sigma_1(A)$ and $\|A\|_2 = \left(\sum \sigma_i^2(A)\right)^{1/2}$.

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 and let $A^{(n,m)}$ and $B^{(n,m)}$ be 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.*

Proof: To simplify notation, we omit index (n,m) over A and B. Weyl's inequalities (see p.423 of Horn and Johnson (1985)) for singular values of any $n \times m$ matrices F and G are: $\sigma_{i+j-1}(F+G) \leq \sigma_i(F) + \sigma_j(G)$, where $1 \leq i,j \leq \min(n,m)$ and $i+j \leq \min(n,m)+1$. Taking $i=1$, $j=k$, $F = A-B$ and considering $G=-A$ and $G=B$, we obtain: $|\sigma_k(A) - \sigma_k(B)| \leq \sigma_1(A-B) = o_p(n^{-1/6})$. But $|\sigma_k^2(A) - \sigma_k^2(B)| = |\sigma_k(A) - \sigma_k(B)|\,(\sigma_k(A) + \sigma_k(B)) \leq \sigma_1(A-B)\left(2\sigma_1(B) + o_p(n^{-1/6})\right) = o_p\left(n^{1/3}\right)$ uniformly over k. QED

Lemma 4 Let n, m *and* T *go to infinity so that* $n \sim m = o\left(T^{3/7}\right)$. *Then, under Assumptions 1 and 2i,* $\sigma_1^2\left(\hat{\chi} - \hat{\Lambda}_0\hat{F}\right) = o_p\left(n^{-1/3}\right)$, *where* $\hat{\chi} \equiv [\hat{\chi}_1(n), ..., \hat{\chi}_m(n)]$, $\hat{F} \equiv \left[\hat{F}_1, ..., \hat{F}_m\right]$ *and* $\hat{\Lambda}_s \equiv \sum_{u=0}^{\infty} \Lambda^{(u)}(n) e^{-iu\omega_s}$.

Proof: Write $\hat{\chi} - \hat{\Lambda}_0\hat{F}$ as $P+Q$, where P and Q are $n \times m$ matrices with s-th columns $P_s \equiv \hat{\chi}_s(n) - \hat{\Lambda}_s\hat{F}_s$ and $Q_s \equiv \left(\hat{\Lambda}_s - \hat{\Lambda}_0\right)\hat{F}_s$, respectively. First, consider matrix P. Interchanging the order of the sums in the definition $\hat{\chi}_s(n) \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{u=0}^{\infty}\Lambda^{(u)}(n)F_{t-u}e^{-it\omega_s}$ and changing summation index t to $\tau = t-u$, we obtain the following representation: $P_s \equiv \hat{\chi}_s(n) - \hat{\Lambda}_s\hat{F}_s = \frac{1}{\sqrt{T}}\sum_{u=0}^{\infty}\Lambda^{(u)}(n)e^{-iu\omega_s}r_u$, where $r_u = \sum_{\tau=-u+1}^{\min(T-u,0)} F_\tau e^{-i\tau\omega_s} - \sum_{\tau=\max(0,T-u)+1}^{T} F_\tau e^{-i\tau\omega_s}$.

Using this representation, we obtain: $E\left\|P_s\right\|^2 \leq \frac{1}{T}\sum_{u,v=0}^{\infty} E\left|r_u'\Lambda^{(u)}(n)'\Lambda^{(v)}(n)r_v\right| \leq \frac{1}{T}\sum_{u,v=0}^{\infty}\left\|\Lambda^{(u)}(n)\right\|\left\|\Lambda^{(v)}(n)\right\| E\left(\left\|r_u\right\|\left\|r_v\right\|\right) \leq \frac{1}{T}\sum_{u,v=0}^{\infty}\left\|\Lambda^{(u)}(n)\right\|\left\|\Lambda^{(v)}(n)\right\|\left(E\left\|r_u\right\|^2 * E\left\|r_v\right\|^2\right)^{1/2} = \frac{1}{T}\left(\sum_{u=0}^{\infty}\left\|\Lambda^{(u)}(n)\right\|\left(E\left\|r_u\right\|^2\right)^{1/2}\right)^2$. But $E\left\|r_u\right\|^2 = \sum_{\tau=-u+1}^{\min(T-u,0)}\left\|F_\tau\right\|^2 + \sum_{\tau=\max(0,T-u)+1}^{T}\left\|F_\tau\right\|^2 = 2k\min(u,T)$ because F_t is a k-dimensional white noise. Therefore, by Lemma 2: $E\sigma_1^2(P) \leq \sum_{s=1}^{m} E\left\|P_s\right\|^2 \leq \frac{m}{T}\left(\sum_{u=0}^{\infty}\left\|\Lambda^{(u)}(n)\right\|(2ku)^{1/2}\right)^2$ which is $o\left(n^{-1/3}\right)$ by Assumption 2i and by the assumption that $n \sim m = o\left(T^{3/7}\right)$. Hence, by Markov's inequality, $\sigma_1(P) = o_p\left(n^{-1/6}\right)$.

Next, consider matrix Q. Note that $\left\|\hat{\Lambda}_s - \hat{\Lambda}_0\right\| \leq \sum_{u=0}^{\infty}\left\|\Lambda^{(u)}(n)\right\|\left|e^{-iu\omega_s} - e^{-iu\omega_0}\right| \leq \sum_{u=0}^{\infty}\left\|\Lambda^{(u)}(n)\right\| u\frac{2\pi(m+1)}{T}$ which is $o\left(n^{-5/6}\right)$ uniformly in s by Assumption 2i and by the assumption that $n \sim m = o\left(T^{3/7}\right)$. Further, since F_t is a k-dimensional white noise, $\sum_{s=1}^{m}\left\|\hat{F}_s\right\|^2 = O_p(m)$. Finally, by Lemma 2, $\sigma_1^2(Q) \leq \sum_{s=1}^{m} E\left\|Q_s\right\|^2 \leq \max_s\left\|\hat{\Lambda}_s - \hat{\Lambda}_0\right\|^2\sum_{s=1}^{m}\left\|\hat{F}_s\right\|^2 = o\left(n^{-5/3}\right)\cdot O_p(m) = o_p\left(n^{-2/3}\right)$ so that $\sigma_1(Q) = o_p\left(n^{-1/3}\right)$. Now, the statement of the lemma follows from the fact that $\sigma_1^2\left(\hat{\chi} - \hat{\Lambda}_0\hat{F}\right) \leq \left(\sigma_1(P) + \sigma_1(Q)\right)^2 = o_p\left(n^{-1/3}\right)$.QED

Lemma 5 *Let $\hat{e} \equiv \left[\hat{e}_1(n), ..., \hat{e}_m(n)\right]$. Then, under the assumptions of Theorem 1, there exists an $n \times m$ matrix $\tilde{e}$ with independent $N_n^{\mathbb{C}}\left(0, 2\pi S_n^e(\omega_0)\right)$ columns, independent from $\hat{F}$, and such that $\sigma_1^2\left(\hat{e} - \tilde{e}\right) = o_p\left(n^{-1/3}\right)$.*

Proof: First, suppose that Assumption 2ii holds and $n \sim m = o\left(T^{3/8}\right)$. Define $\eta \equiv \left(\left(\operatorname{Re}\hat{e}_1(n)\right)', \left(\operatorname{Im}\hat{e}_1(n)\right)', ..., \left(\operatorname{Re}\hat{e}_m(n)\right)', \left(\operatorname{Im}\hat{e}_m(n)\right)'\right)'$. Using Theorem 4.3.2 of Brillinger (1981), which characterizes mixed cumulants of discrete Fourier transforms, it is not difficult to show[7] that $E\eta\eta' = V + R$ with a block-diagonal $V = \pi I_m \otimes \begin{pmatrix} \operatorname{Re} S_n^e(\omega_0) & -\operatorname{Im} S_n^e(\omega_0) \\ \operatorname{Im} S_n^e(\omega_0) & \operatorname{Re} S_n^e(\omega_0) \end{pmatrix}$ and $R_{ij} = \delta_{[i/2n],[j/2n]}O(m/T) + O\left(T^{-1}\right)$, where δ_{st} is the Kronecker's delta, and $O(m/T)$ and $O\left(T^{-1}\right)$ are uniform in i and j running from 1 to $2nm$.

[7]For details of the derivation see the Supplementary Appendix.

Construct $\tilde{\eta} = V^{1/2}(V+R)^{-1/2}\eta$ and define an $n \times m$ matrix $\tilde{e}$ with the s-th columns $\tilde{e}_s$ so that $\left((\mathrm{Re}\,\tilde{e}_1)', (\mathrm{Im}\,\tilde{e}_1)', ..., (\mathrm{Re}\,\tilde{e}_m)', (\mathrm{Im}\,\tilde{e}_m)'\right)' = \tilde{\eta}$. Note that $\tilde{e}$ has independent $N_n^{\mathbb{C}}(0, 2\pi S_n^e(\omega_0))$ columns by construction.

Using inequalities $\|BA\|_2 \leq \|B\|\,\|A\|_2$ and $\|AB\|_2 \leq \|A\|_2\,\|B\|$ (see, for example, Horn and Johnson's (1985) problem 20, p.313), we obtain: $E\|\eta - \tilde{\eta}\|^2 = \left\|(V+R)^{1/2} - V^{1/2}\right\|_2^2 \leq \left\|V^{1/4}\right\|^4 \left\|\left(I + V^{-1/2}RV^{-1/2}\right)^{1/2} - I\right\|_2^2$. Denote the i-th largest eigenvalue of $V^{-1/2}RV^{-1/2}$ as μ_i and note that $|\mu_i| \leq 1$ for large enough T. Since $\left|(1+\mu_i)^{1/2} - 1\right| \leq |\mu_i|$ for any $|\mu_i| \leq 1$, the i-th eigenvalue of $\left(I + V^{-1/2}RV^{-1/2}\right)^{1/2} - I$ is no larger by absolute value than the i-th eigenvalue of $V^{-1/2}RV^{-1/2}$ for large enough T. Therefore, $E\|\eta - \tilde{\eta}\|^2 \leq \left\|V^{1/4}\right\|^4 \left\|V^{-1/2}RV^{-1/2}\right\|_2^2 \leq \left\|V^{1/4}\right\|^4 \left\|V^{-1/2}\right\|^4 \|R\|_2^2$. But $\left\|V^{1/4}\right\| = (\pi l_{1n})^{1/4}$ and $\left\|V^{-1/2}\right\| = (\pi l_{nn})^{-1/2}$ by construction, and $\|R\|_2^2 = \sum_{i,j=1}^{2nm}\left(\delta_{[i/2n],[j/2n]}O(m/T) + O(T^{-1})\right)^2 = m(2n)^2 O(m^2/T^2) + (2mn)^2 O(T^{-2}) = o\left(n^{-1/3}\right)$ because $n \sim m = o\left(T^{3/8}\right)$. Hence, $E\|\eta - \tilde{\eta}\|^2 \leq (\pi l_{1n})(\pi l_{nn})^{-2} o\left(n^{-1/3}\right) = o\left(n^{-1/3}\right)$, where the last equality holds because l_{1n} and l_{nn}^{-1} remain bounded as $n, m \to \infty$ by Assumption 3. Finally, Lemma 2 and Markov's inequality imply that $\sigma_1^2(\hat{e} - \tilde{e}) = o_p\left(n^{-1/3}\right)$.

Now, suppose that Assumption 2iia holds and $m = o\left(T^{1/2-1/p}\log^{-1}T\right)^{6/13}$. In this case, $\hat{e}_{is} = \sum_{j=1}^{\infty} A_{ij}\hat{u}_{js}$, where $\hat{u}_{js}$ is the dft of u_{jt} at frequency ω_s. For fixed j and $\omega_0 = 0$, Phillips (2007) shows that there exist iid complex normal variables ξ_{js}, $s = 1, ..., m$, such that $\hat{u}_{js} - \xi_{js} = o_p\left(\frac{m}{T^{1/2-1/p}}\right)$ uniformly over $s \leq m$. Lemmas 1S, 2S and 3S in the Supplementary Appendix extend Phillips' proof to the case $\omega_0 \neq 0$ and show that there exist Gaussian processes u_{jt}^G with the same auto-covariance structure as u_{jt} and independent over $j \in \mathbb{N}$ such that the differences between the dft's $\hat{u}_{js} - \hat{u}_{js}^G \equiv r_{js}$ satisfy: $\sup_{j>0} E\left(\max_{s\leq m}|r_{js}|\right)^2 \leq Km^2T^{2/p-1}\log^2 T$ for large enough T, where $K > 0$ depends only on $p, \mu_p, \sup_{j\geq 1}\left(\sum_{k=0}^{\infty} k\,|c_{jk}|\right)^p$ and $\sup_{j\geq 1}|C_j(e^{-i\omega_0})|$.

Assumption 2iia implies that the process $e_{it}^G = \sum_{j=1}^{\infty} A_{ij}u_{jt}^G$ satisfies Assumption

2ii. Let $\hat{e}^G$ be the $n \times m$ matrix with i,s-th entry equal to the dft of e_{it}^G at frequency ω_s. Since $\left(T^{1/2-1/p}\log^{-1}T\right)^{6/13} = o\left(T^{3/8}\right)$ for positive p, the above analysis of the Gaussian case implies the existence of an $n \times m$ matrix $\tilde{e}$ described in Lemma 5 and such that $\sigma_1^2\left(\hat{e}^G - \tilde{e}\right) = o_p\left(n^{-1/3}\right)$. On the other hand, $\sigma_1\left(\hat{e} - \tilde{e}\right) \leq \sigma_1\left(\hat{e} - \hat{e}^G\right) + \sigma_1\left(\hat{e}^G - \tilde{e}\right)$. Hence, to complete the proof, we only need to show that $\sigma_1^2\left(\hat{e} - \hat{e}^G\right) = o_p\left(n^{-1/3}\right)$. But we have: $\sum_{i=1}^n \sum_{s=1}^m E\left|\left(\hat{e} - \hat{e}^G\right)_{is}\right|^2 = \sum_{i=1}^n \sum_{s=1}^m \sum_{j=1}^\infty A_{ij}^2 E\left|r_{js}\right|^2 \leq m\sum_{i=1}^n \sum_{j=1}^\infty A_{ij}^2 E\left(\max_{s\leq m}\left|r_{js}\right|\right)^2$ which is $o\left(n^{-1/3}\right)$ because $\sup_{i>0}\sum_{j=1}^\infty A_{ij}^2 < \infty$ by Assumption 2iia and $m\sum_{i=1}^n E\left(\max_{s\leq m}\left|r_{js}\right|\right)^2 \leq mnKm^2T^{2/p-1}\log^2 T = o\left(n^{-1/3}\right)$ if $n \sim m = o\left(T^{1/2-1/p}\log^{-1}T\right)^{6/13}$ as have been assumed. Therefore, Lemma 2 and Markov's inequality imply that $\sigma_1^2\left(\hat{e} - \hat{e}^G\right) = o_p\left(n^{-1/3}\right)$.QED

Lemma 6 *Let* $A(\varkappa) = A + \varkappa A^{(1)}$, *where* $A \equiv \text{diag}\left(a_1, ..., a_k, 0, ..., 0\right)$ *is an* $n \times n$ *matrix with* $a_1 \geq a_2 \geq ... \geq a_k > 0$ *and* $A^{(1)}$ *is a symmetric non-negative definite* $n \times n$ *matrix with lower right* $(n-k) \times (n-k)$ *block* $A_{22}^{(1)}$. *Let* $r_0 = a_k/2$. *Then, for any real* $\varkappa$ *such that* $0 < \varkappa < r_0/\left\|A^{(1)}\right\|$ *and for any* $i = 1, 2, ..., n-k$, *we have*: $\left|\lambda_{k+i}\left(A(\varkappa)\right) - \varkappa\lambda_i\left(A_{22}^{(1)}\right)\right| \leq \frac{\varkappa^2\left\|A^{(1)}\right\|^2}{r_0 - \varkappa\left\|A^{(1)}\right\|}$.

Proof: Let $P(\varkappa)$ and P be the orthogonal projections in $\mathbb{R}^n$ on the invariant subspaces of $A(\varkappa)$ and A, respectively, corresponding to their smallest $n-k$ eigenvalues. Then $A_{22}^{(1)} = PA^{(1)}P$ and the $k+i$-th eigenvalue of $A(\varkappa)$ is the i-th eigenvalue of $\tilde{A}^{(1)}(\varkappa) \equiv P(\varkappa)A(\varkappa)P(\varkappa)$ so that by Weyl's inequalities (see Theorem 4.3.1 of Horn and Johnson (1985)) $\left|\lambda_{k+i}\left(A(\varkappa)\right) - \varkappa\lambda_i\left(A_{22}^{(1)}\right)\right| \leq \left\|\tilde{A}^{(1)}(\varkappa) - \varkappa PA^{(1)}P\right\|$. Kato (1980) shows (see formulas 2.37,2.38 and 2.17 on pages 81 and 87 with Kato's $\tilde{T}^{(1)}(\varkappa)$ equivalent to our $\frac{1}{\varkappa}\tilde{A}^{(1)}(\varkappa)$) that for $|\varkappa| < r_0/\left\|A^{(1)}\right\|$, $\tilde{A}^{(1)}(\varkappa) - \varkappa PA^{(1)}P = \frac{\varkappa}{2\pi i}\sum_{k=1}^\infty(-\varkappa)^k\int_\Gamma R(z)\left(A^{(1)}R(z)\right)^{k+1}zdz$, where $R(z) \equiv (A - zI_n)^{-1}$ and Γ is a positively oriented circle in the complex plane with center 0 and radius r_0. Note that $\max_{z\in\Gamma}\|R(z)\| = \frac{1}{r_0}$ so that $\left\|\int_\Gamma R(z)\left(A^{(1)}R(z)\right)^{k+1}zdz\right\| \leq 2\pi\frac{\left\|A^{(1)}\right\|^{k+1}}{r_0^k}$, and we have

the desired estimate: $\left\|\tilde{A}^{(1)}(\varkappa) - \varkappa P A^{(1)} P\right\| \leq \sum_{k=1}^{\infty} \frac{|\varkappa|^{k+1}\left\|A^{(1)}\right\|^{k+1}}{r_0^k} = \frac{|\varkappa|^2\left\|A^{(1)}\right\|^2}{r_0 - |\varkappa|\left\|A^{(1)}\right\|}$. QED

Proof of Theorem 1: Lemmas 3, 4 and 5 imply that there exists a matrix $\tilde{e}$ with $iidN_n^{\mathbb{C}}(0, 2\pi S_n^e(\omega_0))$ columns such that $\left|\gamma_j - \sigma_j^2\left(\frac{\hat{\Lambda}_0\hat{F}+\tilde{e}}{\sqrt{2\pi m}}\right)\right| = o_p\left(m^{-2/3}\right)$ uniformly over j. Then, since $\sigma_{m,n} \sim m^{2/3}$, the asymptotic joint distribution of interest in Theorem 1 is the same as that of $\left\{\sigma_{m,n}^{-1}\left(\lambda_{k+i}\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right) - \mu_{m,n}\right), i = 1, ..., r\right\}$, where $\tilde{X} \equiv \hat{\Lambda}_0\hat{F} + \tilde{e}$ and $\lambda_j(A)$ denotes the j-th largest eigenvalue of matrix A. We will prove that this distribution is the multivariate TW_2.

Let U be a unitary matrix such that $\hat{F}U$ has all but the first k columns zero. Partition $\hat{F}U$ as $\left[\hat{F}_1, 0\right]$ and $\tilde{e}U$ as $[\tilde{e}_1, \tilde{e}_2]$, where $\tilde{e}_1$ is $n \times k$ and $\hat{F}_1$ is $k \times k$. Then, we have a decomposition: $\frac{\tilde{X}\tilde{X}'}{2\pi m} = \frac{\left(\hat{\Lambda}_0\hat{F}_1+\tilde{e}_1\right)\left(\hat{\Lambda}_0\hat{F}_1+\tilde{e}_1\right)'}{2\pi m} + \frac{\tilde{e}_2\tilde{e}_2'}{2\pi m}$, where matrix $\tilde{e}_2$ has $iidN_n^{\mathbb{C}}(0, 2\pi S_n^e(\omega_0))$ columns. Let $V_n' A V_n$ be a spectral decomposition of $\frac{\left(\hat{\Lambda}_0\hat{F}_1+\tilde{e}_1\right)\left(\hat{\Lambda}_0\hat{F}_1+\tilde{e}_1\right)'}{2\pi mn}$ so that A is a rank-k diagonal matrix with the first k diagonal elements $a_1 \geq a_2 \geq ... \geq a_k > 0$. Then, we have: $\frac{V_n\tilde{X}\tilde{X}'V_n'}{2\pi mn} = A + \frac{1}{n}A^{(1)}$, where $A^{(1)} \equiv \frac{V_n\tilde{e}_2\tilde{e}_2'V_n'}{2\pi m}$. Denote the matrix of the last $n-k$ rows of $V_n\tilde{e}_2$ as $\tilde{e}_{22}$ so that $A_{22}^{(1)} \equiv \frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}$ is the lower right $(n-k)\times(n-k)$ block of $A^{(1)}$. Since $\frac{1}{n}\lambda_i\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right) = \lambda_i\left(\frac{V_n\tilde{X}\tilde{X}'V_n'}{2\pi mn}\right)$, we have, by Lemma 6: $\left|\frac{1}{n}\lambda_{k+i}\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right) - \frac{1}{n}\lambda_i\left(A_{22}^{(1)}\right)\right| \leq \frac{\left\|A^{(1)}\right\|^2/n^2}{r_0 - \left\|A^{(1)}\right\|/n}$ whenever $\frac{1}{n} < \frac{r_0}{\left\|A^{(1)}\right\|}$, where $r_0 = a_k/2$.

Note that $\left\|A^{(1)}\right\| \equiv \lambda_1\left(\frac{\tilde{e}_2\tilde{e}_2'}{2\pi m}\right) \leq \lambda_1\left(\frac{\tilde{e}\tilde{e}'}{2\pi m}\right) = O_p(1)$ by Lemma 1. On the other hand, r_0 decreases slower than $n^{-1/3}$. Indeed, by Weyl's inequalities for singular values (see Lemma 3): $a_k^{1/2} \geq \lambda_k^{1/2}\left(\frac{\hat{\Lambda}_0\hat{F}\hat{F}'\hat{\Lambda}_0'}{2\pi mn}\right) - \lambda_1^{1/2}\left(\frac{\tilde{e}_1\tilde{e}_1'}{2\pi mn}\right)$, where $\lambda_1\left(\frac{\tilde{e}_1\tilde{e}_1'}{2\pi mn}\right) \leq \frac{1}{n}\lambda_1\left(\frac{\tilde{e}\tilde{e}'}{2\pi m}\right) = O_p\left(\frac{1}{n}\right)$ and $\lambda_k\left(\frac{\hat{\Lambda}_0\hat{F}\hat{F}'\hat{\Lambda}_0'}{2\pi mn}\right) \geq \frac{1}{n}\lambda_k\left(\frac{\hat{F}\hat{F}'}{2\pi m}\right)\lambda_k\left(\hat{\Lambda}_0'\hat{\Lambda}_0\right)$ decreases slower than $o_p\left(n^{-1/3}\right)$, by Assumptions 1 and 4. Therefore, for large enough n, the inequality $\frac{1}{n} < \frac{r_0}{\left\|A^{(1)}\right\|}$ is satisfied and $\left|\lambda_{k+i}\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right) - \lambda_i\left(A_{22}^{(1)}\right)\right| \leq \frac{\left\|A^{(1)}\right\|^2/n}{r_0 - \left\|A^{(1)}\right\|/n} = o_p\left(n^{-2/3}\right)$. This implies that, since $\sigma_{m,n} \sim m^{-2/3} \sim n^{-2/3}$, the random variables $\sigma_{m,n}^{-1}\left(\lambda_{k+i}\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right) - \mu_{m,n}\right)$,

$i = 1, ..., r$ have the same asymptotic joint distribution as $\sigma_{m,n}^{-1}\left(\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right) - \mu_{m,n}\right)$, $i = 1, ..., r$.

Now, note that the distribution of $\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}$ conditional on V_n is $W_{n-k}^{\mathbb{C}}\left(m-k, \frac{\bar{S}_n^e(\omega_0)}{m}\right)$, where $\bar{S}_n^e(\omega_0)$ is obtained from $V_n S_n^e(\omega_0) V_n'$ by eliminating its first k rows and columns. Matrix $\bar{S}_n^e(\omega_0)$ satisfies an assumption analogous to Assumption 3 for $S_n^e(\omega_0)$. Precisely, let $\bar{l}_{1n} \geq ... \geq \bar{l}_{n-k,n}$ be the eigenvalues of $\bar{S}_n^e(\omega_0)$, $\bar{H}_n$ be the spectral distribution of $\bar{S}_n^e(\omega_0)$ and $\bar{c}_{m,n}$ be the unique root in $\left[0, \bar{l}_{1n}^{-1}\right)$ of the equation $\int\left(\frac{\lambda\bar{c}_{m,n}}{1-\lambda\bar{c}_{m,n}}\right)^2 d\bar{H}_n(\lambda) = \frac{m-k}{n-k}$. Then, as n and m tend to infinity so that m/n remains in a compact subset of $(0,\infty)$, $\limsup \bar{l}_{1n} < \infty$, $\liminf \bar{l}_{n-k,n} > 0$, and $\limsup \bar{l}_{1n}\bar{c}_{m,n} < 1$. The first two of the latter inequalities follow from Assumption 3 and the fact that, by Theorem 4.3.15 of Horn and Johnson (1985), $l_{k+i,n} \leq \bar{l}_{in} \leq l_{in}$ for $n-k \leq i \leq 1$. The third inequality follows from the fact that $\bar{c}_{m,n} - c_{m,n} = o(1)$. This fact, and even a stronger result that $\bar{c}_{m,n} - c_{m,n} = O(1/n)$, can be established by finding bounds on function $\bar{f}(c) \equiv \int\left(\frac{\lambda c}{1-\lambda c}\right)^2 d\bar{H}_n(\lambda)$ in terms of function $f(c) \equiv \int\left(\frac{\lambda c}{1-\lambda c}\right)^2 dH_n(\lambda)$ (see the Supplementary Appendix for details).

Since $\bar{S}_n^e(\omega_0)$ satisfies an assumption analogous to Assumption 3, Lemma 1 implies that the joint distribution of $\left\{\bar{\sigma}_{m,n}^{-1}\left(\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right) - \bar{\mu}_{m,n}\right); i = 1, ..., r\right\}$ conditional on V_n converges to TW_2, where $\bar{\mu}_{m,n} = \frac{1}{\bar{c}_{m,n}}\left(1 + \frac{n-k}{m-k}\int\frac{\lambda\bar{c}_{m,n}}{1-\lambda\bar{c}_{m,n}} d\bar{H}_n(\lambda)\right)$ and $\bar{\sigma}_{m,n} = \frac{1}{(m-k)^{2/3}\bar{c}_{m,n}}\left(1 + \frac{n-k}{m-k}\int\left(\frac{\lambda\bar{c}_{m,n}}{1-\lambda\bar{c}_{m,n}}\right)^3 d\bar{H}_n(\lambda)\right)^{1/3}$. We however, would like to show that the convergence to TW_2 still takes place if we replace $\bar{\sigma}_{m,n}$ by $\sigma_{m,n}$ and $\bar{\mu}_{m,n}$ by $\mu_{m,n}$. It is enough to show that $\bar{\sigma}_{m,n}^{-1}\left(\bar{\mu}_{m,n} - \mu_{m,n}\right)$ and $\left(\bar{\sigma}_{m,n}^{-1} - \sigma_{m,n}^{-1}\right)\left(\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right) - \mu_{m,n}\right)$ are $o_p(1)$. But the inequality $l_{k+i,n} \leq \bar{l}_{in} \leq l_{in}$ for $n-k \leq i \leq 1$ and the fact that $\bar{c}_{m,n} - c_{m,n} = O(1/n)$ imply that $\bar{\mu}_{m,n} - \mu_{m,n} = O(n^{-1})$ and $\bar{\sigma}_{m,n}^{-1} - \sigma_{m,n}^{-1} = O\left(n^{-1/3}\right)$. Then, since $\bar{\sigma}_{m,n} \sim n^{-2/3}$, we indeed have: $\bar{\sigma}_{m,n}^{-1}\left(\bar{\mu}_{m,n} - \mu_{m,n}\right) = o(1)$. Further, since $\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right) = O_p(1)$ and $\mu_{m,n} = O(1)$, we indeed have $\left(\bar{\sigma}_{m,n}^{-1} - \sigma_{m,n}^{-1}\right)\left(\lambda_i\left(\bar{e}_{22}\bar{e}_{22}'/m\right) - \mu_{m,n}\right) =$

$o_p(1)$.

Therefore, we have: $\left\{\sigma_{m,n}^{-1}\left(\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right)-\mu_{m,n}\right); i=1,...,r\right\}$ conditional on V_n converges to TW_2. In particular, the conditional probabilities $\Pr\left(\sigma_{m,n}^{-1}\left(\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right)-\mu_{m,n}\right)\leq x_i; i=1,...,r|V_n\right)$ converge to the cumulative distribution function $TW_2(x_1,...,x_r)$ with probability 1. Then, by the dominated convergence theorem, the unconditional probabilities $\Pr\left(\sigma_{m,n}^{-1}\left(\lambda_i\left(\frac{\tilde{e}_{22}\tilde{e}_{22}'}{2\pi m}\right)-\mu_{m,n}\right)\leq x_i; i=1,...,r\right)$, which are just the expected values of the conditional probabilities, also converge to $TW_2(x_1,...,x_r)$. QED

Proof of Theorem 2: The convergence of R to $\max_{0<i\leq k_1-k_0}\frac{\lambda_i-\lambda_{i+1}}{\lambda_{i+1}-\lambda_{i+2}}$ when $k=k_0$ follows from Theorem 1. When $k_0<k\leq k_1$, $R\geq\frac{\gamma_k-\gamma_{k+1}}{\gamma_{k+1}-\gamma_{k+2}}$. Therefore, we only need to show that $\frac{\gamma_k-\gamma_{k+1}}{\gamma_{k+1}-\gamma_{k+2}}\xrightarrow{p}\infty$. As was shown in the proof of Theorem 1, $\left|\gamma_i-\lambda_i\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right)\right|=o\left(n^{-2/3}\right)$ uniformly in i. Using Weyl's inequalities for singular values (see Lemma 3), we obtain: $\left|\lambda_i^{1/2}\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right)-\lambda_i^{1/2}\left(\frac{\hat{\Lambda}_0\hat{F}\hat{F}'\hat{\Lambda}_0'}{2\pi m}\right)\right|\leq\lambda_1^{1/2}\left(\frac{\tilde{e}\tilde{e}'}{2\pi m}\right)$ for $i=1,...,n$, where $\lambda_1\left(\frac{\tilde{e}\tilde{e}'}{2\pi m}\right)=O_p(1)$ by Lemma 1. Take $i=k$. By Assumption 4, $\lambda_k\left(\frac{\hat{\Lambda}_0\hat{F}\hat{F}'\hat{\Lambda}_0'}{2\pi m}\right)\xrightarrow{p}\infty$. Therefore, $\lambda_k\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right)\xrightarrow{p}\infty$, and hence, $\gamma_k\xrightarrow{p}\infty$. Now, take $i>k$. Then $\lambda_i^{1/2}\left(\frac{\hat{\Lambda}_0\hat{F}\hat{F}'\hat{\Lambda}_0'}{2\pi m}\right)=0$. Therefore, $\lambda_i\left(\frac{\tilde{X}\tilde{X}'}{2\pi m}\right)=O_p(1)$, and hence, $\gamma_i=O_p(1)$. Summing up, $\gamma_k-\gamma_{k+1}\xrightarrow{p}\infty$, while $\gamma_{k+1}-\gamma_{k+2}=O_p(1)$. Hence, $\frac{\gamma_k-\gamma_{k+1}}{\gamma_{k+1}-\gamma_{k+2}}\xrightarrow{p}\infty$.QED

References

Bai, J., and S. Ng (2007) "Determining the number of primitive shocks in factor models", *Journal of Business and Economic Statistics* 25, 52-60.

Boivin, J., Giannoni, M.P. and Mojon, B. (2008) "How has the Euro Changed the monetary Transmission?", NBER Macroeconomics Annual 23.

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

Brillinger, D. R. (1981) Time series. Data analysis and theory. Holden Day, Inc., San Francisco.

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

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

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

El Karoui, N. (2006) "On the largest eigenvalue of Wishart matrices with identity covariance when n, p and n/p tend to infinity", manuscript, Berkeley University.

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

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.

Giannone, D., L. Reichlin, and L. Sala (2005) "Monetary Policy in real Time", in M. Gertler and K. Rogoff, Eds., NBER Macroeconomic Annual 2004, Cambridge: MIT Press, 161-200

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.

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

Jacobs, J.P.A.M. and P.W. Otter (2008) "Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach", Econometric Reviews 27, pp. 385-397.

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.

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

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

Onatski, A. (2009) "Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors", manuscript, Columbia University.

Onatski, A. (2008) "The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices", Annals of Applied Probability 18, pp.470-490

Péché, S. (2009) "Universality results for the largest eigenvalues of some sample covariance matrix ensembles", Probability Theory and Related Fields 143, pp. 481-516.

Phillips, P.C.B. (2007) "Unit root log periodogram regression", Journal of Econometrics 138, pp. 104-124.

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

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

Stock, J., and M. Watson (2005) "Implications of Dynamic Factor Models for VAR Analysis", manuscript, Princeton University.

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

Uhlig, H. (2008) "Macroeconomic Dynamics in the Euro Area. Discussion by Harald Uhlig", NBER Macroeconomics Annual 23.