DETERMINING THE NUMBER OF FACTORS FROM EMPIRICAL DISTRIBUTION OF EIGENVALUES

Alexei Onatski*

Abstract—We develop a new estimator of the number of factors in the approximate factor models. The estimator works well even when the idiosyncratic terms are substantially correlated. It is based on the fact, established in the paper, that any finite number of the largest “idiosyncratic” eigenvalues of the sample covariance matrix cluster around a single point. In contrast, all the “systematic” eigenvalues, the number of which equals the number of factors, diverge to infinity. The estimator consistently separates the diverging eigenvalues from the cluster and counts the number of the separated eigenvalues. We consider a macroeconomic and a financial application.

I. Introduction

FACTORS models with large cross-sectional and temporal dimensions have become a popular tool in finance and macroeconomics. In finance, such models are at the heart of Chamberlain and Rothschild’s (1983) extension of arbitrage pricing theory. In macroeconomics, they are used in an exploding number of applications (for a survey, see Breitung & Eickmeier, 2005).

Much of the related econometric research, including Connor and Korajczyk (1993), Bai and Ng (2002), Hallin and Liska (2007), and Watson and Amengual (2007), has been exclusively focused on the important question: How many factors are there? A critical assumption made by this research is that factors’ cumulative effect on the n × n cross-sectional units grows proportionally to n. Consequently, if there are r static factors, then r eigenvalues of the data’s covariance matrix grow proportionally to n while the rest of the eigenvalues stay bounded. Therefore, the number of factors can be estimated by separating relatively large eigenvalues from the rest using threshold functions such as in Bai and Ng’s (2002) information criteria.

This paper estimates the number of factors without making any assumptions on the speed of growth of the factors’ cumulative effect. One reason to be interested in factors with slowly increasing cumulative effect is provided by the arbitrage pricing theory. As Chamberlain and Rothschild (1983) showed, if asset returns admit the approximate factor structure, an asset’s risk premium is approximately equal to a linear combination of its factor loadings. Excluding relatively weakly influential factors from consideration makes the cumulative approximation error arbitrarily large as the number of assets increases.

Another reason to be interested in factors that have cumulative effects growing more slowly than n is the empirical observation that the eigenvalues of the sample covariance matrices of large macroeconomic and financial data sets do not obviously separate into groups of large and small eigenvalues. Although this observation does not invalidate the fast-growth assumption, it questions its empirical content and motivates this paper’s search for an alternative approach.

Instead of requiring that the cumulative effect of factors grow as fast as n, this paper imposes a structure on the idiosyncratic components of the data. Precisely, we assume that the n × T matrix e of the idiosyncratic components of the data is such that

\[ e = A \times B, \]

where A and B are two largely unrestricted deterministic matrices and e is an n × T matrix with i.i.d. gaussian entries, so that both the cross-sectional and temporal correlations of the idiosyncratic terms are allowed. We can lift the gaussianity assumption, but then we require that either A or B is a diagonal matrix while the other matrix remains relatively unrestricted. As we explain in the next section, structure (1) is more flexible than most of the alternative structures proposed in similar literature. Exploiting the structure, we show that any finite number of the largest of the bounded eigenvalues of the sample covariance matrix cluster around a single point. Our estimator consistently separates the diverging eigenvalues from the cluster and counts the number of the separated eigenvalues. This number is our estimate of the number of factors.

The main difference between our estimator and the Bai-Ng estimator is that in our case, the implicit threshold separating the bounded and the diverging eigenvalues is sharp, in the sense that it cannot be decreased without hurting the consistency of the estimator, whereas the threshold in the Bai-Ng criterion can be arbitrarily scaled, leaving too much freedom for its choice. The Monte Carlo analysis shows that our estimator outperforms estimators of Bai and Ng (2002) in samples of empirically relevant sizes when there is a non-trivial amount of correlation in the idiosyncratic terms. More specifically, the Bai-Ng criteria tend to severely overestimate the number of factors relative to our criterion.

The rest of the paper is organized as follows. In section II, we review the related literature. In section III, we describe the approximate factor model, formulate our assumptions, and develop a consistent estimator of the number of factors. In section IV, we calibrate a parameter that our estimator depends on. Section V contains a Monte Carlo analysis of the estimator. Section VI uses the new method to estimate the number of factors in the arbitrage pricing theory and in a large macroeconomic panel. Section VII concludes. Technical proofs are in the appendix, which can be downloaded from this REVIEW’s Web site at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00043.

Received for publication July 17, 2006. Revision accepted for publication January 12, 2009.

* Columbia University.

This paper has benefited from the comments and suggestions of a great number of friends and colleagues. I am not listing names in fear of missing someone.

II. Related Literature

This paper’s general idea of exploiting a structure of idiosyncratic terms to find the number of factors is not new. It can be traced back at least to Wachter (1976), who proposes a simple method to decide on the number of principal components in the principal components analysis of large-dimensional data. When the data are i.i.d. and thus have no signals, the empirical distribution of eigenvalues of the sample covariance matrix converges to a known nonrandom distribution as the dimensionality of the data and the number of observations rise (see Marčenko & Pastur, 1967). Wachter suggests comparing a plot of the actual empirical distribution of eigenvalues of the data to the cumulative distribution function of such a limiting law and taking the number of principal components equal to the number of the relatively large eigenvalues responsible for a visible mismatch. Wachter’s analysis was considerably extended and generalized in Silverstein and Combettes (1992).

Kapetanios (2004) is the first to use the idea of exploiting a structure of idiosyncratic terms in the context of approximate factor models. He assumes that the idiosyncratic terms are homoskedastic and unconditionally uncorrelated. He points out that this assumption implies a closed-form expression for a sharp asymptotic upper bound on the idiosyncratic eigenvalues of the sample covariance matrix. Counting the eigenvalues above the bound gives an estimate of the number of factors.

This paper allows for the covariance structure of the idiosyncratic terms that is much less restrictive than Kapetanios’s. This flexibility costs us the existence of a closed-form expression for the upper bound. The upper bound becomes a complicated implicit function of the cross-sectional and temporal correlation structure of the data. We propose a method of consistently estimating the upper bound from the data.

More recently, Kapetanios (2007) has developed another estimator of the number of factors that sequentially tests hypotheses that the number of factors is 0, 1, 2, . . . against the alternative that the number of factors is larger than 0, 1, 2, . . . , respectively. The estimator equals the number of factors specified by the first nonrejected null. To perform the tests, Kapetanios employs a subsampling method to approximate the asymptotic distribution of a test statistic $\lambda_i - \lambda_j$, where $\lambda_i$ is the i-th largest eigenvalue of the sample covariance matrix. He makes high-level assumptions about the existence of a scaling of $\lambda_i - \lambda_j$, which converges in distribution to some unknown limit law, about properties of such a law and about the functional form of the scaling constant.

Unfortunately, the only known sufficient conditions on the correlation structure of the idiosyncratic terms that guarantee Kapetanios’ high-level assumptions exclude either cross-sectional or temporal correlation and require the idiosyncratic terms to be gaussian (see el Karoui, 2005). In this paper we explicitly allow for both cross-sectional and temporal dependence of the gaussian idiosyncratic terms. Alternatively, for nongaussian idiosyncratic terms, we explicitly allow for time-series correlation and cross-sectional heteroskedasticity—general $B$ and diagonal $A$ in equation (1)—or vice versa—general $A$ and diagonal $B$ in equation (1).

III. Assumptions and Basic Results

In this paper, we study approximate factor models of the form

$$X^{(n,T)} = \Lambda^{(n,T)}F^{(n,T)} + e^{(n,T)},$$

(2)

where $X^{(n,T)}$ is an $n \times T$ matrix of data on $n$ cross-sectional units observed over $T$ time periods, $\Lambda^{(n,T)}$ is an $n \times r$ matrix whose $i,j$th element is interpreted as the loading of the $j$th factor on the $i$th cross-sectional unit, $F^{(n,T)}$ is an $r \times T$ matrix whose $j,t$th element is interpreted as the value of the $j$th factor at time $t$, and $e^{(n,T)}$ is an $n \times T$ matrix of the idiosyncratic components of the data. Models (2) corresponding to relatively small $n$ and $T$ are not necessarily nested into the larger models. The only assumptions on $\Lambda^{(n,T)}, F^{(n,T)}$ and $e^{(n,T)}$ made are assumptions 1 and 2, below. Our goal is to consistently estimate the unknown number of factors $r$ from the data as both $n$ and $T$ go to infinity.

To simplify notation, we will omit the superscript $(n,T)$ on matrices $\Lambda, F,$ and $e$ in what follows:

**Assumption 1.** For any constants $B > 0$ and $\delta > 0$, there exist positive integers $n_0$ and $T_0$ such that for any $n > n_0$ and $T > T_0$, the probability that the smallest eigenvalue of $(\Lambda' \Lambda)(F'F/T)$ is below $B$, is smaller than $\delta$.

In an important special case when the columns of $F$ are i.i.d. vectors with an identity covariance matrix, assumption 1 is equivalent to requiring that the smallest eigenvalue of $\Lambda' \Lambda$ diverges to infinity as $n \to \infty$. This is the restriction on factor loadings introduced by Chamberlain and Rothschild (1983). Assumption 1 can be loosely interpreted as saying that the cumulative effect of the “least influential factor” diverges in probability to infinity as the sample size grows.

Our assumption is weaker than that in Connor and Korajczyk (1993), Stock and Watson (2002, 2005), Bai and Ng (2002, 2006), and Bai (2003), who assume that the cumulative effect of the least influential factor rises proportionally to $n$. Such a proportional growth assumption is an essential component of Bai’s (2003) and Bai and Ng’s (2006) proofs of the asymptotic normality of the principal components factor estimators and factor-based forecasts. These asymptotic normality results are likely to break down under our assumption 1.

\footnote{For asymptotic analysis of the principal components estimator of factor models with i.i.d. gaussian idiosyncratic terms under an assumption similar to assumption 1 see Onatski (2006).}
Assumption 1 does not impose any separate restrictions on the convergence or divergence of $A'TA$ and $FF'/T$. Therefore, much room is left for modeling factors. For example, they may be deterministic, or random and stationary, or random and nonstationary. In particular, the factors may follow $I(1)$ processes as in Bai and Ng (2004). In addition, we do not require factors to be independent from the idiosyncratic terms, which are characterized by our next assumption.

To formulate the next assumption, we introduce new notation. Let $\lambda_i(A) \geq \ldots \geq \lambda_n(A)$ be the eigenvalues of a generic $n \times n$ symmetric matrix $A$. We define the eigenvalue distribution function for $A$ as

$$\mathcal{F}_A(x) = 1 - \frac{1}{n} \# \{ i \leq n : \lambda_i(A) > x \},$$

where $\#\{\cdot\}$ denotes the number of elements in the indicated set. Note that $\mathcal{F}_A(x)$ is a valid cumulative probability distribution function (cdf). Further, for a generic probability distribution having a bounded support and cdf $\mathcal{F}_x$, let $u(\mathcal{F})$ be the upper bound of the support, that is, $u(\mathcal{F}) = \min\{x : \mathcal{F}(x) = 1\}$.

**Assumption 2.** For any positive integers $n$ and $T$, there exist an $n \times n$ deterministic matrix $A$, a $T \times T$ deterministic matrix $B$, and an $n \times T$ random matrix $\varepsilon$ with $i, t$ element $\varepsilon_{it}$ such that equation (1) holds, where:

i. $\varepsilon_{it}, 1 \leq i \leq n, 1 \leq t \leq T$ are i.i.d. and satisfy moment conditions $E\varepsilon_{it} = 0, E(\varepsilon_{it})^2 = 1, E(\varepsilon_{it})^4 < \infty$.

ii. $\mathcal{F}_{AA}$ and $\mathcal{F}_{BB}$ weakly converge to probability distribution functions $\mathcal{F}_A$ and $\mathcal{F}_B$, respectively, as $n$ and $T$ go to infinity.

iii. Distributions $\mathcal{F}_A$ and $\mathcal{F}_B$ have bounded support, $u(\mathcal{F}_{AA}) \to u(\mathcal{F}_A) > 0$ and $u(\mathcal{F}_{BB}) \to u(\mathcal{F}_B) > 0$ almost surely as $n$ and $T$ go to infinity.

iv. $\liminf_{\tilde{t} \to 0} f_{\mathcal{F}_{AA}}(\lambda_i) = k_A > 0$ and $\liminf_{\tilde{t} \to 0} f_{\mathcal{F}_{BB}}(\lambda_i) = k_B > 0$.

Assumption 2 allows the idiosyncratic terms to be nontrivially correlated both cross-sectionally and over time. The cross-sectional correlation is caused by matrix $A$ linearly combining different rows of $\varepsilon$, whereas the correlation over time is caused by matrix $B$ linearly combining different columns of $\varepsilon$. Such a way of creating covariances between different elements of matrix $\varepsilon$ is restrictive. Denoting the $nT \times 1$ vector of stacked columns of $\varepsilon$ as vec $\varepsilon$, we have $E(\text{vec } e(\text{vec } e')) = BB' \otimes AA'$. How well $BB' \otimes AA'$ can approximate more interesting covariance structures depends on the details of these structures. (For a general discussion of the quality of approximations with Kronecker products, see Van Loan & Pitsianis, 1993.)

Part ii of assumption 2 is satisfied for rich classes of matrices. In particular, $BB'$ (as well as $AA'$) may be the autocovariance matrix of any covariance-stationary process with a bounded spectral density. Part iii of the assumption implies clustering of the largest eigenvalues of the population covariance matrix of vec $e$. In particular, it does not allow a few linear combinations of idiosyncratic terms to have unusually large variation, which would make these combinations difficult to separate from combinations of the genuine factors in finite samples.

Assumption 2 allows us to use recent results from the large random matrix literature formulated below as lemmas 1 and 2. Let us first introduce some useful concepts and notation. For any probability distribution function $G$, we define its Stieltjes transform as $m_G(z) = \int (\lambda - z)^{-1}dG(\lambda)$, where $z \in \mathbb{C}^+ \equiv \{ z \in \mathbb{C} : \text{Im } z > 0 \}$. The Frobenius-Perron inversion formula,

$$G([a, b]) = \frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b \text{Im } m_G(\xi + i\eta)d\xi,$$

where $a$ and $b$ are points of continuity of $G$, ensures that we can reconstruct a distribution from its Stieltjes transform. We will denote the Stieltjes transforms of $\mathcal{F}_A$ and $\mathcal{F}_B$ as $m_A(z)$ and $m_B(z)$, respectively.

**Lemma 1** (Zhang, 2006). Let $\{T(n), n \in \mathbb{N}\}$ be a sequence of positive integers such that $n/T(n) \to c > 0$ as $n \to \infty$. Then, if assumption 2 holds, the eigenvalue distribution of $ee'/T(n)$ converges to a nonrandom cdf $\mathcal{F}_{c,AB}$ almost surely. For each $z \in \mathbb{C}^+$, the Stieltjes transform of $\mathcal{F}_{c,AB}$, $m(z)$, together with two other analytics on $\mathbb{C}^+$ functions $u(z)$ and $v(z)$, constitute a solution to the system:

$$zm(z) + 1 = u(z)m_A(u(z)) + 1$$
$$zm(z) + 1 = c^{-1}[v(z)m_B(v(z)) + 1]$$
$$zm(z) + 1 = -c^{-1} \frac{z}{u(z)v(z)}$$

which is unique in the set $\{m(z), u(z), v(z)\}$: $\text{Im } m(z) > 0, \text{Im}(u(z)) > 0, \text{Im}(v(z)) > 0$.

This is a slightly reformulated version of theorem 1.2.1 in Zhang (2006).³ Let $\mathcal{F}_{c,A,\varepsilon}$ denote the distribution whose Stieltjes transform is given by equation (5) replacing $c, m_A$, and $m_B$ with $n/T(n)$, the Stieltjes transform of $\mathcal{F}_{AA'}$, and the Stieltjes transform of $\mathcal{F}_{BB'}$, respectively. Paul and Silverstein (2008) prove the following lemma:

**Lemma 2** (Paul & Silverstein, 2008). Let $\{T(n), n \in \mathbb{N}\}$ be a sequence of positive integers such that $n/T(n) \to c > 0$ as $n \to \infty$. Further, let $\varepsilon$ have gaussian entries; alternatively, let either $A$ or $B$ in equation (1) be a diagonal matrix. Then, if assumption 2 holds, $\text{Pr}(\text{no eigenvalue of } ee'/T(n) \text{ appears}}$

³Zhang (2006) formulates her theorem in terms of the Stieltjes transform $s(z)$ of $ee'/T(n)$ (as opposed to that of $ee'/T(n)$). Further, instead of our $u(z)$ and $v(z)$, she uses $p(z) = -1/v(z)$ and $q(z) = -1/u(z)$. Finally, her theorem is formulated under weaker assumptions than those of lemma 1. The technical appendix states the original version of theorem 1.2.1 and derives from it our lemma 1.
in \([a, b]\) for all large \(n\) = 1, where \([a, b]\) is any interval that lies in an open interval outside the support of \(F_{\lambda_0}^{X_n, A, B_n}\).

Lemma 2 corresponds to theorem 1 in Paul and Silverstein (2008). That theorem does not require all parts of assumption 2 to hold. Precisely, part iii of assumption 2 can be substituted by a weaker assumption that functions \(F_\lambda\) and \(F_\mu\) have bounded support and that \(u(F^{MA})\) and \(u(F^{RB})\) are bounded in \(n\). We use the stronger assumption to prove the next lemma:

**Lemma 3.** Under conditions of lemma 2, \(\Pr(\)no eigenvalue of \(ee'/T(n)\) appear in \([a, b]\) for all large \(n\) = 1, where \([a, b]\) is any interval that lies in an open interval to the right of the support of \(F^{C, A, B}\).

The proof of lemma 3 is given in the appendix, which can be downloaded from this REVIEW’s Web page. It consists of showing that under assumption 2, the upper boundary of \(F_{ee'}^{C, A, B_n}\) converges to the upper boundary of \(F^{C, A, B}\). To show such a convergence, we characterize the upper boundary of support of a distribution whose Stieltjes transform solves system (5). Such a characterization is nontrivial and constitutes a result of independent interest. (For similar work in different settings see Marcenko & Pastur, 1967; Silverstein & Choi, 1995; Dozier & Silverstein, 2004.)

For the purpose of this paper, lemma 3 is important because it implies that the eigenvalues of the sample covariance matrix \(XX'/T\) can be separated into \(r\) systematic eigenvalues that diverge to infinity and a group of idiosyncratic eigenvalues, any finite number of the largest of which cluster around \(u(F^{C, A, B})\). Indeed, we have the following theorem:

**Theorem 1.** Let model (2) satisfy assumptions 1 and 2, and let \(\{T(n), n \in \mathbb{N}\}\) be a sequence of positive integers such that \(n/T(n) \to c > 0\) as \(n \to \infty\). Further, let \(e\) have gaussian entries; alternatively, let either \(A\) or \(B\) in equation (1) be a diagonal matrix. Then, as \(n \to \infty\), we have:

i. For any sequence of positive integers \(\{i(n), n \in \mathbb{N}\}\) such that \(i(n)/n \to 0\) as \(n \to \infty\), and \(i(n) > r\) for large enough \(n\) the \(i(n)\)th eigenvalue of \(XX'/T(n)\) converges almost surely to \(u(F^{C, A, B})\).

ii. The \(r\)th eigenvalue of \(XX'/T(n)\) tends to infinity in probability.

**Proof.** Given lemma 3, proof of theorem 1 is simple. We therefore provide it here instead of relegating it to the appendix. According to a singular value analog of Weyl’s eigenvalue inequalities (see theorem 3.3.16 of Horn & Johnson, 1991), for any \(n \times T\) matrices \(P\) and \(Q\), we have

\[
\sigma_{i+j-1}(P + Q) \leq \sigma_i(P) + \sigma_j(Q),
\]

where \(1 \leq i, j \leq \min(n, T)\) and \(\sigma_i(P)\) denotes the \(i\)th largest singular value of matrix \(P\), which is another name for the square root of the \(i\)th largest eigenvalue of matrix \(PP'\). Let \(\lambda_i, \mu_i, \) and \(v_i\) be the \(i\)th largest eigenvalues of \(XX'/T(n)\), \(ee'/T(n)\), and \((\Lambda \Lambda')(FF'/T(n))\), respectively. Note that since the rank of the latter matrix is \(r\), \(v_{r+1}\) equals 0. Inequality (6) implies

\[
\sqrt{\lambda_i} \geq \sqrt{\mu_r} - \sqrt{\mu_1}
\]

(7)

\[
\lambda_i \leq \mu_{i-r}, \text{ for } i = r + 1, \ldots, n
\]

(8)

\[
\lambda_i \geq \mu_{i+r}, \text{ for } i = 1, \ldots, n - r
\]

(9)

where the first inequality follows by taking \(P = X/\sqrt{T(n)}\) and \(Q = -e/\sqrt{T}\), the second inequality follows by taking \(P = e/\sqrt{T}\) and \(Q = \Lambda F'/\sqrt{T}\), and the last inequality follows by taking \(P = X/\sqrt{T}\) and \(Q = -\Lambda F'/\sqrt{T}\).

Assumption 1, lemma 3, and inequality (7) imply that the first \(r\) eigenvalues of the sample covariance matrix diverge to infinity. On the other hand, according to inequalities (8) and (9), the \(i\)th eigenvalue of the sample covariance matrix is sandwiched between \(\mu_{i(n)−r}\) and \(\mu_{i(n)+r}\). The latter two converge to \(u(F^{C, A, B})\) by lemma 3, the definition of \(u(F^{C, A, B})\), and the assumption that \(i(n)/n \to 0\) as \(n \to \infty\). Hence, \(\lambda_{i(n)}\) also converges to \(u(F^{C, A, B})\).

Theorem 1 suggests a method of estimation of \(r\). Indeed, it implies that for any \(j > r\), the difference \(\lambda_j - \lambda_{j+1}\) converges to 0, while the difference \(\lambda_r - \lambda_{r+1}\) diverges to infinity. Let \(r^n_{\max}, n \in \mathbb{N}\) be a slowly increasing sequence of real numbers such that \(r^n_{\max}/n \to 0\) as \(n \to \infty\). \(r^n_{\max}\) will be interpreted as the maximum possible number of factors that a researcher having a sample of size \(n, T(n)\) assumes a priori. We define a family of estimators,

\[
\hat{r}(\delta) = \max \left\{ i \leq r^n_{\max} : \lambda_i - \lambda_{i+1} \geq \delta \right\},
\]

(10)

parameterized by a positive number \(\delta\). If the set \(\{ i \leq r^n_{\max} : \lambda_i - \lambda_{i+1} \geq \delta \}\) is empty, then \(\hat{r}(\delta) = 0\). The consistency of \(\hat{r}(\delta)\) for any fixed \(\delta > 0\) and any slowly increasing sequence \(\{r^n_{\max}, n \in \mathbb{N}\}\) follows from theorem 1 and from the fact that \(r^n_{\max}\) becomes larger than a fixed number \(r\) for large enough \(n\).

**Corollary 1.** Under assumptions of theorem 1, for any fixed \(\delta > 0\) and any slowly increasing sequence \(\{r^n_{\max}, n \in \mathbb{N}\}\) such that \(r^n_{\max}/n \to 0\) as \(n \to \infty\), \(\hat{r}(\delta) \to r\) in probability as \(n \to \infty\).

Although consistent, estimator (10) may perform poorly in finite samples if \(\delta\) is poorly calibrated. In the next section, we describe a calibration procedure for \(\delta\).

**IV. Calibration of \(\delta\)**

Lemma 1 and inequalities (7), (8), and (9) imply that the empirical distribution of \(\lambda_1, \ldots, \lambda_r\) converges to \(F^{C, A, B}\). Hence, a natural approximation of \(F^{C, A, B}(\lambda_i) = \frac{1}{n}\#(j: \lambda_j \leq \lambda_i) \equiv (n - i + 1)/n\), and we may calibrate the difference \(\lambda_i - \lambda_{i+1}\) for \(i > r\) by \([F^{C, A, B}]^{-1}(\frac{n+i}{n}) - [F^{C, A, B}]^{-1}(\frac{n+i+1}{n})\).
It is therefore reasonable to make $\delta$ in equation (10) no smaller than $[\mathcal{F}^{x:y}]^{-1}(y + \frac{1}{n}) - [\mathcal{F}^{x:y}]^{-1}(y)$, where $y$ is close to $1$.

Unfortunately, as is evident from equation (5), $\mathcal{F}^{x:y}$ is a very complicated function of $A$ and $B$, and it does not have an explicit form. However, locally to $u(\mathcal{F}^{x:y})$, the density corresponding to $\mathcal{F}^{x:y}(x)$ is proportional to the square root function $u(\mathcal{F}^{x:y}(x) - 1)^{1/2}/2$ for a very general class of matrices $A$ and $B$. At a heuristic level, the idea is as follows. Clearly the Stieltjes transform of $\mathcal{F}^{x:y}, m(z) \equiv \int (\lambda - z)^{-1}d\mathcal{F}^{x:y}(\lambda)$, is a well-defined analytical real function on $z \in (u, +\infty)$, where $u \equiv u(\mathcal{F}^{x:y})$. It must, however, have a singularity at $u$ because otherwise, although not defined on $z < u$ by integral $\int (\lambda - z)^{-1}d\mathcal{F}^{x:y}(\lambda), m(z)$ can be analytically extended from $z > u$ to those $z < u$ that are close enough to $u$. The extension will be real for $z < u$. On the other hand, according to the inversion formula (4), $m(z)$ must have a positive imaginary part for $z \in (-\infty, u)$ arbitrarily close to $u$. Since $m(z)$ has a singularity at $u$, the analytical extension of its inverse, $z(\cdot)$, should not be invertible at $m(u)$, which happens if its second-order approximation lacks the first-order term: $z(m) \approx u + k(m - m(u))^{2}$. Rearranging, we get $m(z) \approx m(u) + (z - u)^{1/2}/k$. This approximation, together with the inversion formula (4), explains the square root phenomenon.\footnote{Note that the square root behavior of the density implies that the cdf $\mathcal{F}^{x:y}(x)$ can be approximated by $1 - a(1 - x)^{1/2}$ for some positive $a$ in the neighborhood of $u$, where $a$ takes the positive part of the argument. Therefore, $[\mathcal{F}^{x:y}]^{-1}(y)$ can be approximated by $u - (\frac{y}{2})^{2}/3$ for $y$ close to $u$ but smaller than 1, and a reasonable upper bound for $[\mathcal{F}^{x:y}]^{-1}(y + \frac{1}{2}) - [\mathcal{F}^{x:y}]^{-1}(y)$ for such $y$ is given by $(1 - \frac{1}{n})^{2}/3 + \frac{1}{2} \frac{1}{n^{2}}|_{y=1} \equiv (an)^{-2/3} \equiv \beta$. In our Monte Carlo experiments and applications below, we will make $\delta$ equal $2\beta$, where $\beta$ is an estimate of $(an)^{-2/3}$.}

Any estimate $\hat{\beta}$ can be obtained as follows. Define $y_{j} \equiv \lambda_{j}$ and $x_{j} \equiv (j - 1)^{2}/3$. Then an approximate linear relationship $y_{j} \approx u - (an)^{-2/3}x_{j}$ holds for small enough $j > r$. We obtain $\hat{\beta}$ by running an ordinary least squares regression of $y_{j}, y_{j+1}, \ldots, y_{j+4}$ on the constant and $x_{j}, x_{j+1}, \ldots, x_{j+4}$, where $j$ is any integer larger than $r$.\footnote{An estimate can be obtained as follows. Define $y_{j} \equiv \lambda_{j}$ and $x_{j} \equiv (j - 1)^{2}/3$. Then an approximate linear relationship $y_{j} \approx u - (an)^{-2/3}x_{j}$ holds for small enough $j > r$. We obtain $\hat{\beta}$ by running an ordinary least squares regression of $y_{j}, y_{j+1}, \ldots, y_{j+4}$ on the constant and $x_{j}, x_{j+1}, \ldots, x_{j+4}$, where $j$ is any integer larger than $r$.} One possible choice for $j$ is $r_{\text{max}} + 1$. A better choice can be made using an iterative procedure: starting from $j = r_{\text{max}} + 1$, calibrate $\hat{\beta}$, and estimate $r$ by $\hat{\beta}(\hat{\beta}$); next, take $j = \hat{\beta}(\hat{\beta} + 1$, calibrate $\hat{\beta}$ again, and estimate $r$ by $\hat{\beta}(\hat{\beta}$ using the newly calibrated $\hat{\beta}$. This iterative procedure can be repeated until convergence.

Let us conclude this section by providing an algorithm for estimating the number of factors based on equation (10):

1. Compute eigenvalues $\lambda_{1}, \ldots, \lambda_{n}$ of the sample covariance matrix $XX'/T$. Set $j = r_{\text{max}} + 1$.
2. Compute $\hat{\beta}$, the slope coefficient in the OLS regression of $\lambda_{j}, \ldots, \lambda_{j+4}$, on the constant and $(j - 1)^{2}/3, \ldots, (j + 3)^{2}/3$. Set $\delta = 2\hat{\beta}$.
3. Compute $\hat{r}(\hat{\beta}) = \max\{i \leq r_{\text{max}} : \lambda_{i} - \lambda_{i+1} \geq \delta\}$, or if $\lambda_{i} - \lambda_{i+1} < \delta$ for all $i \leq r_{\text{max}}$, set $\hat{r}(\hat{\beta}) = 0$.
4. Set $j = \hat{r}(\hat{\beta}) + 1$. Repeat steps 2 and 3 until convergence.\footnote{In our Monte Carlo experiments, we set the number of iterations to four, but the convergence was often achieved at the second iteration.}

In what follows, we denote the estimator described by the above algorithm as ED (edge distribution) because it exploits the square root shape of the edge of the eigenvalue distribution. In the next section, we use Monte Carlo analysis to study the performance of ED and its sensitivity to different choices of $r_{\text{max}}$.

V. Monte Carlo Analysis

In these Monte Carlo experiments, we compare our estimator ED with estimators $PC_{p1}$, $IC_{p1}$, proposed by Bai and Ng (2002), Bayesian information criterion estimator $BIC_{3}$ studied in the same paper, and estimators ME and MED, proposed by Kapetanios (2004) and Kapetanios (2007), respectively.\footnote{We are grateful to George Kapetanios for sharing his codes with us.}

Estimators $PC_{p1}, IC_{p1}$, and $BIC_{3}$ are based on the minimization of the following criterion functions:

$$PC_{p1}(k) = V(k) + \kappa \delta^{2}\left(\frac{n + T}{nT}\right) \ln\left(\frac{nT}{n + T}\right),$$

$$IC_{p1}(k) = \ln(V(k)) + k \left(\frac{n + T}{nT}\right) \ln\left(\frac{nT}{n + T}\right),$$

$$BIC_{3}(k) = V(k) + \kappa \delta^{2}\left(\frac{n + T - k}{nT}\right) \ln(nT),$$

where $V(k)$ is $(nT)^{-1}$ times the minimal possible sum of squared errors of a rank $k$ approximation of the data matrix, and $\kappa \delta^{2}$ is an estimator of the unconditional variance of the idiosyncratic errors.\footnote{In our Monte Carlo experiments, we set the number of iterations to four, but the convergence was often achieved at the second iteration.} Note that $V(k)$ and $\kappa \delta^{2}$ in equations (11), (12), and (13) are equal to $(nT)^{-1} \sum_{j=k+1}^{n} \lambda_{j}$ and $(nT)^{-1} \sum_{j=r_{\text{max}}+1}^{n} \lambda_{j}$ respectively, where $\lambda_{j}$ is the $j$th largest eigenvalue of the sample covariance matrix. Therefore, similar to ED, estimators $PC_{p1}, IC_{p1}$, and $BIC_{3}$ are based entirely on the empirical distribution of eigenvalues $\lambda_{j}$.

Kapetanios’s (2004) ME estimator compares the largest eigenvalue of the sample covariance matrix of the standardized data with the theoretical large sample limit $1 + \sqrt{2}$ of the largest eigenvalue of the sample covariance matrix of i.i.d. data. If the largest eigenvalue is smaller than $1 + \sqrt{2}$, ME equals 0. Otherwise, the first principal component is
subtracted from the data, and the largest eigenvalue of the standardized residuals is compared with $1 + \sqrt{\varepsilon}$. If the largest eigenvalue is smaller than $1 + \sqrt{\varepsilon}$, 1, ME equals 1. Otherwise, the two principal components are subtracted from the data until the largest eigenvalue is smaller than $1 + \sqrt{\varepsilon}$ is reached. Kapetanios’s (2007) MED estimator sequentially tests hypotheses that the number of factors is larger than 0, 1, 2, ..., respectively. MED equals the number of factors specified by the first nonrejected null, or to $r_{\text{max}}$ if all the nulls are rejected.\(^9\)

Our first experiment consists of generating 1,000 replications of data produced by the data-generating process used in Bai and Ng (2002):

\[
X_{it} = \sum_{k=1}^{r} \Lambda_{ik} F_{kt} + \sqrt{\varepsilon} e_{it},
\]

\[
e_{it} = \varrho e_{i,t-1} + v_{it} + \sum_{1 \leq j, l \leq 8} \beta_{ij} v_{i,j} F_{kt}, \Lambda_{ik}, v_{it} \sim \text{i.i.d. } N(0, 1). \quad (14)
\]

\(^9\) We compute MED based on the subsampling size $b(n) = 0.6n$ and 5% significance level tests.

We consider three pairs of parameters $\varrho$ and $\beta$ : $(\varrho, \beta) = (0, 0), (0.3, 0.1)$, and $(0.85, 0)$. The latter two pairs are tailored to fit our financial and macroeconomic applications described in the next section, respectively.\(^{10}\) The true number of factors equals $r = 1, 3, 5, 15$. For $r = 1, 3, 5$, we set $r_{\text{max}} = 8$. For $r = 15$, we set $r_{\text{max}} = 20$. Further, we set $\theta = (1 - \rho^2)/(1 + 16\beta^2)$. Such a choice implies that the factors explain exactly 50% variation in the data. We consider sample sizes $(n, T) \in \{40, 100, (200, 60), (1,000, 60), (100, 100), (150, 150), (150, 500), (1,000, 250)\}$.

Tables 1, 2, and 3 report the percentage of the replications that result in overestimation and underestimation of the number of factors in the cases $(\rho, \beta) = (0, 0), (\rho, \beta) = (0.3, 0.1)$, and $(\rho, \beta) = (0.85, 0)$, respectively. The table entries have the form $x/y$, which means that $x\%$ of the replications result in the overestimation, $y\%$ of the replications result in the underestimation.

\(^{10}\) First, we removed ten principal components from our financial data to obtain an estimate $e_{i}$ of the matrix of its idiosyncratic components. Then we chose $\varrho$ and $\beta$ so as to minimize $\sum_{i}(e_{i}(ee') - \lambda_{e}(ee')^{2})$, where $e$ was generated according to equation (14) and normalized so that $\text{tr}(ee') = \text{tr}(e_{i}e'_{i})$. The minimizing $\varrho$ and $\beta$ were 0.3 and 0.1, respectively. For our macroeconomic data, a similar calibration procedure gave us $\varrho = 0.85$ and $\beta = 0$. 

---

**Table 1.** Monte Carlo Replications of Data-Generating Process (14)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>PC$_{p1}$</th>
<th>IC$_{p1}$</th>
<th>BIC$_{3}$</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>100</td>
<td>1/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
</tbody>
</table>

**Table 2.** Monte Carlo Replications of Data-Generating Process (14)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>PC$_{p1}$</th>
<th>IC$_{p1}$</th>
<th>BIC$_{3}$</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>90/0</td>
<td>100/0</td>
<td>100/0</td>
<td>69/0</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>3/0</td>
<td>100/0</td>
<td>94/0</td>
<td>100/0</td>
<td>100/0</td>
<td>5/0</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>100/0</td>
<td>1/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>93/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>100/0</td>
<td>7/0</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>70/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>100/0</td>
<td>3/0</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>100/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>70/0</td>
<td>2/0</td>
</tr>
</tbody>
</table>

**Table 3.** Monte Carlo Replications of Data-Generating Process (14)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>PC$_{p1}$</th>
<th>IC$_{p1}$</th>
<th>BIC$_{3}$</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>57/0</td>
<td>100/0</td>
<td>100/0</td>
<td>52/27</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>99/0</td>
<td>4/0</td>
<td>100/0</td>
<td>95/0</td>
<td>100/0</td>
<td>2/0</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>170/0</td>
<td>1/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>92/0</td>
<td>2/0</td>
<td>0/0</td>
<td>100/0</td>
<td>100/0</td>
<td>2/0</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>70/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>99/0</td>
<td>1/0</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>100/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>91/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>64/0</td>
<td>0/0</td>
</tr>
</tbody>
</table>

Correlated idiosyncratic terms: $\rho = 0.3, \beta = 0.1$. 

Uncorrelated idiosyncratic terms: $\rho = 0, \beta = 0$. 

---

**Downloaded from** http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00043 by guest on 16 May 2021
The Review of Economics and Statistics

1010

The tables reveal at least three sources of problems for the considered estimators: correlation in the idiosyncratic terms, a relatively large number of factors, and a relatively small sample size. Correlation in the idiosyncratic terms ($\rho \neq 0$ or $\beta \neq 0$) leads to the situation when the estimators often overestimate the true number of factors. The exception from this rule is our estimator, and, for the case $\rho = 0.3, \beta = 0.1$, BIC3. Our estimator remains very good for the cases $\rho = 0.3, \beta = 0.1, \gamma > 15$, and $\rho = 0.85, \beta = 0, \gamma = 0.1$. For the other cases, it very often underestimates. Overall, our estimator correctly estimates the number of factors far more often than any other of the studied estimators.

A relatively large number of factors ($r = 15$) leads to the situation when all the estimators are almost always wrong about the true number of factors. Depending on the values of $\rho$ and $\beta$ and on the sample size, the estimators may either overestimate or underestimate the true number of factors. When our estimator works poorly, it underestimates.

Note that depending on the application, underestimation of the number of factors may be more problematic than overestimation. For example, the principal components–based estimators of the factor space remain consistent only when the assumed number of factors is greater than or equal to the true number. However, note also that when $\lambda_r$ is only marginally larger than $\lambda_{r+1}$, as is typically the case when $r$ is relatively large, the principal components–based estimators estimate the factor space very poorly even if the true number of factors is known (see Onatski, 2006).

Relatively small sample sizes in general worsen the quality of the estimators. For some of the estimators, even the largest of the considered sample sizes ($n = 1,000, T = 250$) is not large enough to guarantee good performance in the face of the nontrivial correlation in the idiosyncratic terms or the large number of factors. Our estimator benefits the most from the relatively large sample sizes. For the two largest sample sizes ($n = 1,000, T = 250$ and $n = 150, T = 500$), it practically always correctly estimates the true number of factors for ten out of twelve considered combinations of $(\rho, \beta)$ and $r$. The second best is BIC3, with six out of twelve combinations with correctly estimated number of factors.

In our second Monte Carlo experiment, we explore the sensitivity of the estimators to the amount of the noise in the data. Figure 1 shows the performance of PCp1, ICp1, BIC3, ME, MED, and ED when $\rho = \beta = 0, r = 3$, and $\theta/r$, interpreted as the relative variance of the idiosyncratic terms, varies from 1 to 25. We see that the Bai-Ng estimators start to underestimate the true number of factors as $\theta/r$ increases. Note that our estimator remains correct for a considerably larger range of $\theta/r$. Eventually, however, it is also bound to underestimate the true number of factors because the systematic information becomes significantly dissolved in the noise, which makes correct identification of the true number of factors virtually impossible. Kapetanios’s ME estimator outperforms the Bai-Ng estimators but works worse than our estimator. Kapetanios’s MED estimator differs from the other

---

Table 3.—(% of Overestimation)/(% of Underestimation) in 1,000 Monte Carlo Replications of Data-Generating Process (14)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>PCp1</th>
<th>ICp1</th>
<th>BIC3</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>14/0</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>4/0</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>4/0</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>500</td>
<td>100/0</td>
<td>97/0</td>
<td>0/0</td>
<td>100/0</td>
<td>25/0</td>
<td>2/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>8/50</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>3/85</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/86</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>2/18</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>500</td>
<td>100/0</td>
<td>97/0</td>
<td>0/0</td>
<td>100/0</td>
<td>31/0</td>
<td>1/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>500</td>
<td>100/0</td>
<td>97/0</td>
<td>0/0</td>
<td>100/0</td>
<td>36/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>200</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>60</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>150</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>500</td>
<td>100/0</td>
<td>97/0</td>
<td>0/0</td>
<td>100/0</td>
<td>36/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>100/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
</tbody>
</table>

Note: The true number of factors $r = 3$. Horizontal axis: $\theta/r$.

Figure 1.—Average Estimated Number of Factors According to BIC3, PCp1, ICp1, ME, MED, and ED.

Correlated idiosyncratic terms: $\rho = 0.85, \beta = 0.1$.

---

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00043 by guest on 16 May 2021
estimators in that it overestimates the true number of factors for most of the considered range of $\theta/r$. In addition, its mean is not monotone with respect to $\theta/r$.

In our next experiment, we explore in more detail the relationship between the quality of the estimators and the amount of the dependence in the idiosyncratic terms. We set the true number of factors at three and vary $\rho$ on a grid $0:0.05:0.3$. Figure 2 reports the average estimates of the number of factors produced by PC$_p$, IC$_p$, BIC$_3$, ME, MED, and our estimator ED for $\rho$ and $\beta$ on the grid across 1,000 Monte Carlo replications, for sample size $n = T = 100$. Table 4 reports the numerical height of the surfaces shown in Figure 2.

We see that the mean of our estimator remains very close to the true number of factors 3 for the widest area in the $(\rho, \beta)$-plane. Indeed, the percentage of the grid nodes where the mean remains in the $\pm 10\%$-of-the-true-number-of-factors segment equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors 3 for the widest area in the $\rho$-$\beta$ plane. Indeed, the percentage of the grid nodes where the mean of our estimator equals 13% for PC$_p$, 19% for IC$_p$, 33% for BIC$_3$, 6% for ME, 4% for MED, and 51% for our estimator. As $\rho$ and $\beta$ rise, our estimator starts to underestimate the true number of factors.

To explore the sensitivity of different estimators to the choice of $r_{\text{max}}$, we computed the corresponding average estimates in 1,000 MC replications on the grid $r_{\text{max}} = 6:1:20$. Figure 3 reports the $r_{\text{max}}$-sensitivity results for $\rho = 0.3, \beta = 0.1$ and for the sample size $n = T = 80, 120, 160,$ and 200.

Our estimator (solid line) is the least sensitive to the choice of $r_{\text{max}}$. Estimator BIC$_3$ (dash-dotted line) behaves similar to ours for $n = T \geq 120$, but it is more sensitive to $r_{\text{max}}$ for $n = T = 80$. Similar to our estimator, IC$_p$ (dotted line) is not sensitive to the choice of $r_{\text{max}}$. However, it substantially overestimates the number of factors for the two smallest sample sizes studied. Estimators PC$_p$, ME, and MED (dashed line, solid line with dot markers, and solid line with star markers, respectively) are very sensitive to $r_{\text{max}}$.

The setting of our last Monte Carlo experiment is the same as that of experiment B in Kapetanios (2007). The data-generating process is

$$X_{it} = \sum_{k=1}^{r_0} \Lambda_{1,ik} F_{kt} + \sum_{k=1}^{r_0} \Lambda_{2,ik} F_{kt-1} + e_{it}; \quad F_{kt}, \Lambda_{1,ik}, \Lambda_{2,ik} \sim \text{i.i.d. } N(0, 1),$$

$$e_{it} = \Sigma^{1/2} v_{it}; \quad v_{it} = 0.5 v_{i,t-1} + \xi_{it}; \quad \xi_{it} \sim \text{i.i.d. } N(0, 2\theta I_n),$$

where $e_i$ is the $n \times 1$ vector with components $e_{it}$; $v_i$ is an $n$-dimensional AR(1) process with innovations $\xi_{it}$, which are independent from $F_{kt}, \Lambda_{1,ik}, \Lambda_{2,ik}$; and matrix $\Sigma = [\sigma_{ij}]$ is such that $\sigma_{ij} = 1, \sigma_{ij} = \sigma_{ji} \sim U((-0.1, 0.1))$ for $0 < |i-j| \leq 5$, and $\sigma_{ij} = 0$ for $|i-j| > 5$. We consider $r_0 = 0$ and $r_0 = 2$ so that the true number of static factors equals 0 and 4, respectively. Finally, we consider all combinations of sample sizes $n = 50, 100, 200; T = 50, 100, 200; \text{and}$
even in the easiest of the cases when other estimators. They poorly estimate the number of factors that work similar to IC when 

\[ \theta = 1, 3, 5, 10. \] Note that the ratio of \( \sum_{i,j} E(\sum_{k=1}^{r_0} A_{1,k} F_{ik} + \sum_{k=1}^{r_0} A_{2,k} F_{(k-1)} - \sum_{k=1}^{r_0} F_{k-1}^2) \) equals \( r_0/(r_0 + \theta/1) \). Hence, for \( r_0 = 2 \), the component explains 60%, 33%, 23%, and 13% of the variation in the data when \( \theta = 1, 3, 5, 10 \), respectively.

Tables 5 and 6 report, respectively, the average estimated number of factors and the percentage of the overestimation in 1,000 Monte Carlo replications of the data-generating process (15). As can be seen from the tables, estimators \( PC_{pl} \) and ME work considerably worse than the other estimators. They poorly estimate the number of factors even in the easiest of the cases when \( \theta = 1 \). Estimator \( BIC_3 \) works similar to \( IC_{pl} \) and ED when \( r = 0 \) and when \( \theta = 1 \) and \( r = 4 \). However, it more substantially underestimates the number of factors that \( IC_{pl} \) and ED when \( \theta > 1 \) and \( r = 4 \). Our estimator performs similarly to \( IC_{pl} \). It is more difficult to compare the performances of MED and of \( IC_{pl} \) and ED. Estimator MED almost always overestimates, and it substantially overestimates when \( r = 0 \). On the other hand, \( PC_{pl} \) and ED underestimate (except for the largest sample sizes) for relatively large \( \theta \) when \( r = 4 \).

\[ \text{TABLE 4.—NUMERICAL HEIGHT OF THE SURFACES SHOWN IN FIGURE 2} \]

<table>
<thead>
<tr>
<th>( \rho</th>
<th>\beta )</th>
<th>0.05</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
<th>0.3</th>
<th>0.05</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( PC_{pl} )</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>( IC_{pl} )</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ n = 80 \]

\[ n = 120 \]

\[ n = 160 \]

\[ n = 200 \]

\[ \text{Figure 3.—AVERAGE ESTIMATED NUMBER OF FACTORS ACCORDING TO BIC}_1 \] (DASH-DOTTED LINE), \( PC_{pl} \) (DASHED LINE), \( IC_{pl} \) (DOTTED LINE), ME (SOLID LINE WITH DOT MARKERS), MED (SOLID LINE WITH STAR MARKERS), and ED (SOLID LINE).

VI. Applications

We apply the new estimation procedure to determine the number of pervasive factors driving stock returns and the number of pervasive factors influencing the dynamics of a large set of macroeconomic variables. First, we estimate the number of factors in the approximate factor model of the stock returns. Our data consist of monthly returns on 1,148 stocks (\( n = 1148 \)) traded on the NYSE, AMEX, and NASDAQ during the period January 1983 to December 2003 (\( T = 252 \)), obtained from CRSP database. We included a stock in the data set if it was traded during the whole sample period.

We computed our estimator ED on the grid \( r_{max} = 6:1:20 \). For all \( r_{max} \) on the grid, ED equals 2. For comparison, Bai and Ng (2002) detect two pervasive factors in the stock returns data. Connor and Korajczyk (1993) find evidence for between one and six pervasive factors in the stock returns. Trzcinka (1986) finds some support for the existence of five pervasive factors. Five seems also to be a preferred number for Roll and Ross (1980) and Reinganum (1981). A study by Brown and Weinstein (1983) also suggests that the number of factors is unlikely to be greater than five. Huang and Jo (1995) identify only two common factors.

To check the robustness of our estimator to the choice of the sample and as an empirical approach to decide on the number of factors, we repeat our analysis over different subgroups of the variables. Precisely, we estimate the number of factors in

\[ \text{Note: The true number of factors } r = 3 \]. Horizontal axis: \( r_{max} \).
the returns of randomly selected 50% of the original stocks. We repeat the random selection procedure 1,000 times. For $r_{\text{max}} = 20$, 460 selected samples give the estimate of one factor, 447 selected samples give the estimate of two factors, and the other 93 selected samples give the estimate of three or more factors. For $r_{\text{max}} < 20$, the results are similar. It is therefore likely that there are only one or two common factors in the stock returns.

Our second application is about determining the number of pervasive factors influencing the dynamics of a large set of macroeconomic variables. The pervasive factors can be viewed as corresponding to the basic macroeconomic shocks driving the economy. The existence of such shocks is in the spirit of modern dynamic stochastic general equilibrium macroeconomic models.

The data set we use is the same as in Watson (2003). It includes 215 monthly time series for the United States from 1959:1 to 1998:12 ($T = 480$). The series represent fourteen main categories of macroeconomic time series: real output and income; employment and hours; real retail, manufacturing, and trade sales; real inventories and inventory-sales ratios; orders and unfilled orders; stock prices; exchange rates; interest rates; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous. The variables in the data set were transformed, standardized, and screened for outliers as described in Stock and Watson (2002). The determination of the number of factors was based on the data subset of the transformed and screened 148 variables available for the full sample period ($n = 148$).

As in the case of the financial data, for all $r_{\text{max}}$ on the grid 6:1:20, our estimate of the number of pervasive factors equals two. For comparison, Stock and Watson (2005) use PC and IC criteria to estimate seven to nine pervasive factors in the data set, which is almost identical to ours but has a larger time dimension. When we apply PC$_p$ and IC$_p$ to our data set, we find that they estimate the number of factors as equal to $r_{\text{max}}$, for $r_{\text{max}}$ in the range 6:18. BIC$_p$ finds two factors when $r_{\text{max}}$ is 6; three factors when $r_{\text{max}}$ is 7 and 8; four factors when
Table 6.—Percentage of Overestimation for \( r = 0 \), and (% of Overestimation)/(% of Underestimation) for \( r = 4 \) in 1,000 Monte Carlo Replications of the Data-Generating Process (15)

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>PC</th>
<th>IC</th>
<th>BIC</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
<th>PC</th>
<th>IC</th>
<th>BIC</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>100</td>
<td>10</td>
<td>3</td>
<td>51</td>
<td>89</td>
<td>10</td>
<td>100/0</td>
<td>29/0</td>
<td>0/3</td>
<td>74/0</td>
<td>49/0</td>
<td>2/4</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>82</td>
<td>6</td>
<td>67/0</td>
<td>0/0</td>
<td>0/2</td>
<td>25/0</td>
<td>40/0</td>
<td>1/0</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>57</td>
<td>5</td>
<td>0/0</td>
<td>0/0</td>
<td>0/1</td>
<td>1/0</td>
<td>29/0</td>
<td>1/0</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>0</td>
<td>100</td>
<td>93</td>
<td>5</td>
<td>100/0</td>
<td>18/0</td>
<td>0/1</td>
<td>100/0</td>
<td>52/0</td>
<td>1/0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>85</td>
<td>5</td>
<td>23/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>30/0</td>
<td>1/0</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>72</td>
<td>3</td>
<td>0/0</td>
<td>0/0</td>
<td>0/50</td>
<td>18/0</td>
<td>1/0</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>96</td>
<td>2</td>
<td>100/0</td>
<td>9/0</td>
<td>0/1</td>
<td>100/0</td>
<td>34/0</td>
<td>0/0</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>89</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>73</td>
<td>3</td>
<td>5/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>13/0</td>
<td>0/0</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>65</td>
<td>2</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>100/0</td>
<td>11/0</td>
</tr>
</tbody>
</table>

\( \theta = 1, r = 0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>PC</th>
<th>IC</th>
<th>BIC</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
<th>PC</th>
<th>IC</th>
<th>BIC</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>100</td>
<td>11</td>
<td>3</td>
<td>60</td>
<td>94</td>
<td>9</td>
<td>100/0</td>
<td>11/49</td>
<td>0/98</td>
<td>54/13</td>
<td>80/3</td>
<td>1/91</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>75</td>
<td>7</td>
<td>63/0</td>
<td>0/36</td>
<td>0/100</td>
<td>20/5</td>
<td>76/0</td>
<td>1/42</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>62</td>
<td>4</td>
<td>0/0</td>
<td>0/15</td>
<td>0/100</td>
<td>18/0</td>
<td>47/0</td>
<td>1/77</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>0</td>
<td>100</td>
<td>96</td>
<td>5</td>
<td>100/0</td>
<td>9/28</td>
<td>0/98</td>
<td>100/0</td>
<td>88/2</td>
<td>1/77</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>79</td>
<td>5</td>
<td>21/0</td>
<td>0/2</td>
<td>0/96</td>
<td>100/0</td>
<td>81/0</td>
<td>1/4</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>73</td>
<td>3</td>
<td>0/0</td>
<td>0/86</td>
<td>47/0</td>
<td>63/0</td>
<td>1/0</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>100</td>
<td>2</td>
<td>0</td>
<td>100</td>
<td>98</td>
<td>2</td>
<td>100/0</td>
<td>6/12</td>
<td>0/98</td>
<td>100/0</td>
<td>76/1</td>
<td>0/49</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>75</td>
<td>3</td>
<td>4/0</td>
<td>0/0</td>
<td>0/77</td>
<td>100/0</td>
<td>62/0</td>
<td>0/0</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>61</td>
<td>2</td>
<td>0/0</td>
<td>0/8</td>
<td>100/0</td>
<td>59/0</td>
<td>0/0</td>
<td></td>
</tr>
</tbody>
</table>

\( \theta = 3, r = 0 \)

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>PC</th>
<th>IC</th>
<th>BIC</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
<th>PC</th>
<th>IC</th>
<th>BIC</th>
<th>ME</th>
<th>MED</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>100</td>
<td>11</td>
<td>3</td>
<td>47</td>
<td>93</td>
<td>9</td>
<td>100/0</td>
<td>1/97</td>
<td>0/100</td>
<td>16/66</td>
<td>79/9</td>
<td>0/99</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>79</td>
<td>6</td>
<td>463/0</td>
<td>0/98</td>
<td>0/100</td>
<td>8/57</td>
<td>75/5</td>
<td>1/92</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td>4</td>
<td>0/18</td>
<td>0/94</td>
<td>0/100</td>
<td>0/25</td>
<td>76/0</td>
<td>1/41</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>100</td>
<td>6</td>
<td>0</td>
<td>100</td>
<td>97</td>
<td>5</td>
<td>1/0</td>
<td>0/192</td>
<td>0/100</td>
<td>100/0</td>
<td>86/5</td>
<td>0/99</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>82</td>
<td>4</td>
<td>16/2</td>
<td>0/67</td>
<td>0/100</td>
<td>99/0</td>
<td>87/0</td>
<td>1/59</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>71</td>
<td>3</td>
<td>0/1</td>
<td>0/22</td>
<td>0/100</td>
<td>48/0</td>
<td>80/0</td>
<td>1/1</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>95</td>
<td>3</td>
<td>100/0</td>
<td>1/86</td>
<td>0/100</td>
<td>100/0</td>
<td>81/7</td>
<td>0/98</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>79</td>
<td>2</td>
<td>3/0</td>
<td>0/21</td>
<td>0/100</td>
<td>100/0</td>
<td>76/0</td>
<td>1/14</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>68</td>
<td>3</td>
<td>0/0</td>
<td>0/0</td>
<td>0/100</td>
<td>100/0</td>
<td>69/0</td>
<td>0/0</td>
</tr>
</tbody>
</table>

\( \theta = 5, r = 0 \)

\( \theta = 3, r = 4 \)

\( r_{\text{max}}^n \) is 9, 10, and 11; and more factors when \( r_{\text{max}}^n \) increases further.

A growing literature estimates the number of dynamic factors in the generalizations of the approximate factor model proposed in Forni et al. (2000) and Stock and Watson (2002). This literature finds at least two dynamic factors in macroeconomic data sets. If the true data-generating process is indeed a dynamic factor model, then our estimate should in principle capture the number of factors and their lags. Hence, the number of static factors should be larger than two.

As we know from our Monte Carlo experiment calibrated to fit the macroeconomic data (\( \rho = 0.85, \beta = 0, n = 150, T = 500 \)), our estimator could underestimate the true number of factors if this number is really large, say 15. Such a large number of static factors is not unrealistic a priori. A related reason for possible underestimation may be that the dynamic factors are strongly serially correlated so that the effect of their lags can be easily confused with the contemporaneous effect. Therefore, some of the static factors capturing lagged effects may be weak in the sense that the smallest systematic eigenvalues would be close to the largest idiosyncratic eigenvalues. Such a situation will lead to the underestimation by ED. Note that if some static factors are indeed weak, they can be estimated only with huge error by the principal components technique (see Onatski, 2006).

To check the robustness of our estimator to the choice of the sample, we repeat our analysis over different subgroups of the variables. We set \( r_{\text{max}} = 20 \). Our estimate of the number of factors remains two when we remove from the data the 20 variables that constitute the “real output and income” category, or the 14 variables that constitute the “price indexes” category, or the 18 variables that constitute the “interest rates” category. However, when we remove 25 variables that constitute the “employment and hours” category, we get an estimate of only one factor in the data. When we estimate the number of factors in the data that consist...
of a randomly selected 50% of the original series, we get 31 out of 1,000 cases of zero-factors-estimate, 267 cases of one-factor-estimate, 647 cases of two-factors-estimate, and 55 cases of three-or-more-factors-estimate. For $r_{\text{max}} < 20$, the results are similar.

VII. Conclusion

In this paper we have developed a new, consistent estimator for the number of factors in the approximate factor models. The estimator is based on the fact, established in the paper, that any finite number of the largest "idiotsyncratic" eigenvalues of the sample covariance matrix clusters around a single point. In contrast, all the "systematic" eigenvalues, the number of which equals the number of factors, diverge to infinity. Our estimator consistently separates the diverging eigenvalues from the cluster and counts the number of the separated eigenvalues.

The clustering result is obtained under the assumption that the matrix of idiosyncratic terms, $e$, has form $AEB$, where $e$ is a matrix with i.i.d. gaussian entries, and matrices $A$ and $B$ are relatively unconstrained matrices that encode cross-sectional and temporal correlation of the entries of $e$. We can lift the requirement of the gaussianity of $e$ at the expense of making one of the matrices $A$, $B$ diagonal.

To prove the clustering, we have established a novel result in the literature on large random matrices (see lemma A4). Precisely, we have characterized the upper boundary of support of the limiting spectral distribution of $ee^T/B$ in terms of the limiting spectral distributions of matrices $A$ and $B$. The most advanced results of this type up to date considered the special case when either $A$ or $B$ is proportional to the identity matrix (see, for example, Silverstein and Choi, 1995).

The main advantage of our estimator relative to the Bai-Ng estimators is that it works well in realistically small samples when the amount of cross-sectional and temporal correlation in the idiosyncratic terms is relatively large. It also improves on the Bai-Ng criteria when the portion of the observed variance attributed to the factors is small relative to the variance due to the idiosyncratic term.

We applied our methodology to estimate the number of factors driving stock returns and the number of factors influencing a large set of macroeconomic variables. In both cases our estimate of the number of factors equals two. If factors not captured by our estimator exist, their influence on the cross-sectional terms is likely to be weak, so that it is easily confused with the idiosyncratic influences. As shown by Onatski (2006), such weak factors are poorly estimated by the principal components methods.

REFERENCES

Bai, J., “Inferential Theory for Factor Models of Large Dimensions,” 

Bai, J., and S. Ng, “Determining the Number of Factors in Approximate Factor Models,” 
*Econometrica* 70 (2002), 191–221.

---

Bai, J., “A PANIC Attack on Unit Roots and Cointegration,” 

---


Chamberlain, G., and M. Rothschild, “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets,” 

Connor, G., and R. Korajczyk, “A Test for the Number of Factors in an Approximate Factor Model,” 

Silverstein, J., and S. Choi, “Analysis of the Limiting Spectral Distribution of Large Dimensional Information-Plus-Noise Type Matrices,” 

El Karoui, N., “Recent Results about the Largest Eigenvalue of Random Covariance Matrices and Statistical Application,” 


Hallin, M., and R. Liska, “The Generalized Dynamic Factor Model: Determining the Number of Factors,” 


Huang, R., and H. Jo, “Data Frequency and the Number of Factors in Stock Returns,” 

Kapetanios, G., “A New Method for Determining the Number of Factors in Factor Models with Large Datasets,” Department of Economics, 
Queen Mary University of London, working paper no. 525 (2004).

Kapetanios, G., “A Testing Procedure for Determining the Number of Factors in Approximate Factor Models with Large Datasets,” Department of Economics, 
Queen Mary University of London, working paper no. 551 (2007).


Onatski, A., “Asymptotic Distribution of the Principal Components Estimator of Large Factor Models When Factors Are Relatively Weak,” 
manuscript, Columbia University (2006).

Paull, D., and J. Silverstein, “No Eigenvalues outside the Support of Limiting Empirical Spectral Distribution of a separable Covariance Matrix,” 
North Carolina State University, working paper (2008).


Silverstein, J., and S. Choi, “Analysis of the Limiting Spectral Distribution of Large Dimensional Random Matrices,” 


Stock, J., and M. Watson, “Macroeconomic Forecasting Using Diffusion Indexes,” 

Trzcinka, C., “On the Number of Factors in the Arbitrage Pricing Model,” 


