Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise.

Alexei Onatski*

Economics Department, Columbia University

First draft: November, 2005

This draft: August, 2007

Abstract

We consider large factor models where factors' explanatory power does not strongly dominate the explanatory power of the idiosyncratic terms asymptotically. We find the first and second order asymptotics of the principal components estimator of such a weak factors as the dimensionality of the data and the number of observations tend to infinity proportionally. The principal components estimator is inconsistent but asymptotically normal.

JEL code: C13, C33. Key words: large factor models, principal components, phase transition, weak factors, inconsistency, asymptotic distribution, Marčenko-Pastur law.

1 Introduction

High-dimensional factor models have recently attracted an increasing amount of attention from researchers in macroeconomics and finance. Factors extracted from hundreds of macroeconomic and financial variables observed for a period of several decades have been used for macroeconomic forecasting, monetary policy and business

*420 West 118th St., International Affairs Building, Economics Department, Room 1011, New York, NY, 10027. Tel: (212) 854-3685. Fax: (212) 854 8059. E-mail: ao2027@columbia.edu

cycle analysis, arbitrage pricing theory tests, and portfolio performance evaluation (see, for example, Stock and Watson (2005), Bernanke, Boivin, and Eliasz (2005), Forni and Reichlin (1998), and Connor and Korajczyk (1988)). A popular technique for factor extraction is the principal components method, which estimates the factors by the principal eigenvectors of a sample-covariance-type matrix. In this paper we study the asymptotic distribution of the principal components estimator when the dimensionality of the data, n, and the number of observations, T, go to infinity proportionally.

The consistency and asymptotic normality of the principal components estimator when both n and T go to infinity have been recently shown by Bai (2003). To prove his results, Bai makes a strong assumption equivalent to requiring that the ratio between the k-th largest and the $k+1$-th largest eigenvalues of the population covariance matrix of the data, where k is the number of factors, is rising *proportionally* to n so that the cumulative effects of the normalized factors on the cross-sectional units strongly dominate the idiosyncratic influences asymptotically. In practice, the ratio of the adjacent eigenvalues of the finite sample analog of the population covariance matrix turns out to be rather small. For example, for the set of the 148 macroeconomic indicators used in Stock and Watson (2002), the ratio of the i-th to the $i+1$-th eigenvalues of the sample covariance matrix is smaller than 1.75 for any positive integer $i \leq 20$, where 20 is a generous *a priori* upper bound on the number of factors. Hence, for the macroeconomic data, the cumulative effect of the "least influential factor" on the cross-sectional units is comparable to the strongest idiosyncratic influence so that, even if the ratio of the k-th to the $k+1$-th eigenvalues does increase proportionally to n, the coefficient of proportionality must be very small and the usefulness of the "strong-factor asymptotics" is questionable.

In this paper, we, therefore, focus on the principal components estimation of models with factors having bounded, instead of increasing with n, cumulative effects on the cross-sectional units. We call such factors weak. More precisely, we consider a sequence of factor models indexed by n :

$$X_{it}^{(n)} = L_i^{(n)\prime} F_t^{(n)} + \varepsilon_{it}^{(n)}, \text{ with } i \in \mathbb{N} \text{ and } t \in \mathbb{N}, \tag{1}$$

where $F_t^{(n)}$ and $L_i^{(n)}$ are $k \times 1$ vectors of factors at time t and factor loadings on the cross-sectional unit i, respectively, and $\varepsilon_{it}^{(n)}$ is an idiosyncratic or noise component of $X_{it}^{(n)}$. Suppose that the data consist of the observations of $X_{it}^{(n)}$ with $i = 1, ..., n$ and

$t = 1, ..., T^{(n)}$, and let the factors be normalized so that $E\left(\sum_{t=1}^{T^{(n)}} F_t^{(n)} F_t^{(n)\prime} / T^{(n)}\right) = I_k$. Then, our key assumption capturing the notion of weak factors is that

$$\sum_{i=1}^{n} L_i^{(n)} L_i^{(n)\prime} - D \to 0,$$

where $D = \text{diag}\,(d_1, ..., d_k)$ with $d_1 > ... > d_k > 0$. This assumption is in contrast to the "strong factor" assumption made by Bai (2003) which, when specialized to our notation and setup, takes the form:

$$\sum_{i=1}^{n} \frac{L_i^{(n)} L_i^{(n)\prime}}{n} - D \to 0.$$

Informally, the "strong factor" assumption requires factors to load non-trivially on an infinite number of the cross-sectional units whereas the weak factor assumption requires factors to load non-trivially on a possibly large but finite number of the cross-sectional units.

This paper answers the question: what is the first and the second order asymptotics of the principal components estimators of the factors and factor loadings when the factors are weak. We find that, in contrast to the "strong factor" case, the estimators are inconsistent. We give explicit formulae for the amount of this inconsistency. Further, we show that, when centered around their respective probability limits, the principal components estimators of the factors and factor loadings are asymptotically normal, and we establish explicit formulae for the asymptotic variance. A Monte Carlo analysis shows that our asymptotic formulae work very well even in samples as small as $n = 40$ and $T^{(n)} = 20$.

We derive all our results under a strong assumption that the noise $\varepsilon_{it}^{(n)}$ is Gaussian and i.i.d. both cross-sectionally (for $i = 1, ..., n$) and over time (for $t = 1, ..., T^{(n)}$). Hence, all non-trivial cross-sectional and time series dependence in the data is due to the presence of factors. Making such an assumption substantially reduces the generality of model (1) but allows us to overcome substantial technical difficulties of analyzing the weak factor case. We leave an important issue of the generalization of our results to less restrictive noise structures to future research.

Our main findings can be summarized in more detail as follows. In what follows, we will omit the superscript (n) from our notations to make them easier to read.

However, occasionally we will use the superscript to emphasize dependence on n. Let X be an $n \times T$ matrix of the data, let F and L be $T \times k$ and $n \times k$ matrices of factors and factor loadings, respectively (that is, the t-th row of F is F_t' and the i-th row of L is L_i'), and let ε be an $n \times T$ matrix of the noise. According to (1), we have: $X = LF + \varepsilon$. The principal components estimator of F, $\hat{F}$, is defined as $\sqrt{T}$ times the matrix of the principal k eigenvectors of a sample-covariance-type matrix $X'X/T$, and the principal components estimator of L, $\hat{L}$, is defined as $X\hat{F}/T$.

In Theorem 1, we establish the following representation of the principal components estimator of the factors:

$$\hat{F} = F \cdot Q + F^{\perp}, \tag{2}$$

where Q is a random $k \times k$ matrix which tends in probability to a diagonal matrix with positive diagonal elements *strictly smaller than unity*, and $F^{\perp}$ is a random $T \times k$ matrix which has columns orthogonal to the columns of F and is such that the joint distribution of the entries of $F^{\perp}$ conditional on F is invariant with respect to the multiplication of $F^{\perp}$ from the left by any orthogonal matrix having $\operatorname{span}(F)$ as an invariant subspace. Matrix Q centered by its probability limit and scaled by $\sqrt{T}$ has asymptotically jointly normal entries, and we find explicit formulae for the probability limit and for the covariance matrix of the asymptotic distribution of Q.

The above representation is illustrated in Figure 1. The principal components estimates $\hat{F}$ "randomly circle" around the true F so in the limit of large n there remains a non-zero angle between $\hat{F}$ and F. When the cumulative effects of the factors on the cross-sectional units, measured by the the diagonal elements of D (the limit of $L'L$), are large, $\operatorname{plim} Q$ is close to an identity matrix and $\hat{F}$ is close to F. When the cumulative effects are small, $\operatorname{plim} Q$ is close to zero and $\hat{F}$ is nearly orthogonal to F. In the extreme case, when the cumulative effect of one of the factors goes below a certain threshold, representation (2) breaks down and the corresponding factor estimate starts to point in a completely random direction. The width of the darker band on the sphere of radius $\sqrt{T}$ represents the size of the asymptotic variance of Q. The more narrow the band, the smaller the asymptotic variance of Q.

A formula completely analogous to (2) holds for the normalized principal components estimator of factor loadings $\hat{\mathcal{L}} \equiv \hat{L}\left(\hat{L}'\hat{L}\right)^{-1/2}$. Precisely, our Theorem 2 shows that $\hat{\mathcal{L}} = \mathcal{L} \cdot R + \mathcal{L}^{\perp}$, where $\mathcal{L}$ is a matrix of normalized factor loadings $L\left(L'L\right)^{-1/2}$

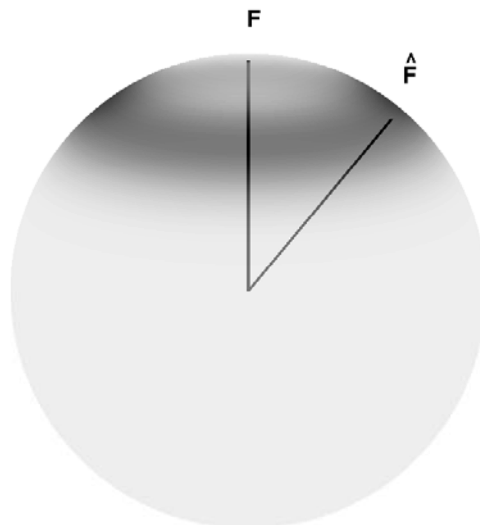

Figure 1: Distribution of $\hat{F}^{(n)}$. The darker areas on the sphere represent the regions of relatively higher probability for $\hat{F}^{(n)}$.

and a random matrix R has properties parallel to those of Q in (2).

Representations of type (2) can be used to obtain the asymptotic distributions of the principal components estimator of factors at particular time periods or factor loadings corresponding to specific cross-sectional units. We find such distributions in Theorems 3 and 4. The distributions are centered at the true values of the factors and of the factor loadings, *shrunk towards zero*. As the cumulative effects of the factors on the cross-sectional units tend to infinity, the bias disappears and our asymptotic formulae converge to formulae found by Bai (2003) for the case of strong factors. The Monte Carlo analysis shows that our asymptotic distribution provides a better approximation for the finite sample distribution than the asymptotic distribution found by Bai (2003) even for relatively "strong" factors.

In the special case when factors are i.i.d. Gaussian random variables, the principal components estimator of the normalized factor loadings is the maximum likelihood estimator. Its asymptotic distribution in the case of fixed n and large T is well known (see Anderson (1984), Chapter 13). In this special case, our asymptotic distribution converges to the classical analog when the limit of the n/T ratio converges to zero. The Monte Carlo analysis shows that for n comparable to T our asymptotic approximation works much better than the classical one.

In this paper we also find the asymptotic distribution of the principal eigenvalues of the sample covariance matrix XX'/T. It is easy to show that the i-th eigenvalue measures the square of the Euclidean length of the i-th column of $\hat{L}$. Hence, the i-th eigenvalue can be interpreted as the principal components estimator of the cumulative effect of the i-th factor on the cross-sectional units. We find that the first k eigenvalues of the sample covariance matrix of the data converge in probability to values strictly larger than the first k eigenvalues of the population covariance matrix. When the "population eigenvalues" are large enough, the "sample eigenvalues" centered by their probability limits and multiplied by $\sqrt{T}$ are asymptotically jointly normal, and we find explicit formulae for the probability limits and the covariance matrix of the asymptotic distribution. If a "population eigenvalue" is below a certain threshold, the corresponding "sample eigenvalue" converges to a positive constant that does not depend on the population eigenvalue.

The rest of the paper is organized as follows. Section 2 explains how our paper is related to the previous statistical literature on large random matrices. In Section 3 we introduce the model, state our assumptions, and formulate our main results. Section 4 provides an intuition for the inconsistency of the principal components estimator. Monte Carlo analysis is given in Section 5. The main steps of our proofs are given in Section 6. Section 7 concludes. All auxiliary results are proven in the Appendix.

2 Connection to the literature on large random matrices

Our paper is related to the mathematical, physical and statistical literature studying eigenvalues and eigenvectors of the sample covariance matrix of high-dimensional data. Most of this literature is concerned with i.i.d. data, which may have factor structure only in the special case when factors are i.i.d. The basic result in the literature is due to Marčenko and Pastur (1967). They show that if an n by T matrix X has i.i.d. entries (hence, there are no factors in the data) with zero mean and unit variance and such that $E\left|X_{it}\right|^{2+\delta} < \infty$ for some $\delta > 0$, then for any real x, as n and T tend to infinity so that $n/T \to c > 0$:

$$\frac{1}{n}\,(\text{number of eigenvalues of } XX'/T \leq x) \to F(x) \tag{3}$$

almost surely. Here $F(x)$ is such that $F'(x) = \frac{1}{2\pi xc}\sqrt{(b-x)(x-a)}$ for $a < x < b$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. In cases when $c > 1$, the limiting measure has an additional mass $1 - \frac{1}{c}$ at zero. This result has been extended and generalized by many authors (see Bai (1999) for a review). Although Marčenko-Pastur result describes the limiting behavior of the whole distribution of the eigenvalues of XX'/T, it says little about the behavior of the largest eigenvalue, which, in the absence of factors, estimates the "strongest" idiosyncratic influence on the cross-sectional units.

Geman (1980) shows that under the i.i.d. assumption and additional growth restrictions on the moments of X_{it}, which are satisfied , for example, if X_{it} is Gaussian, the largest eigenvalue of XX'/T converges almost surely to the upper boundary of the support of $F(x) : \lambda_1 \to (1+\sqrt{c})^2$. Bai et al. (1988) generalize Geman's result by establishing the above convergence under the existence of the fourth moment of X_{it}. Bai (1999) cites other generalizations and extensions of Geman's result and provides details. In the language of factor models, Geman's result says that the principal components estimator of the "strongest" idiosyncratic influence on the cross-sectional units may substantially overestimate the influence. The larger the parameter c, that is the larger the ratio of the cross-sectional dimension to the time series dimension, the larger the amount of the overestimation.

Johnstone (2001) proposes the "spiked covariance" model for X. According to this model, columns of X are i.i.d. observations of n-dimensional Gaussian vectors with covariance matrix having a few relatively large eigenvalues and the rest of the eigenvalues being equal to unity. The spiked covariance model corresponds to model (1) with i.i.d. Gaussian factors. Then, the number of the relatively large eigenvalues corresponds to the number of factors. Johnstone (2001) shows that in the null case when all the eigenvalues of the covariance matrix are equal to one (no factors), the largest eigenvalue of the sample covariance matrix, λ_1, centered by $(1+\sqrt{c})^2$ and multiplied by $n^{2/3}/\left(\sqrt{c}\,(1+\sqrt{c})^{4/3}\right)$ converges in distribution to the so called Tracy-Widom law (see Tracy and Widom (1994)).

Baik et al. (2005) study the limiting behavior of λ_1 in non-null cases (which are consistent with the data having a non-trivial i.i.d. factor component). However, they only consider the case when X has complex as opposed to real Gaussian entries, and the complexity of the entries of X plays a central role in their proofs. Non-null cases of the spiked covariance model for X with real entries are studied in Baik and Silverstein (2006) and Paul (2006). Baik and Silverstein (2006) find the almost sure limit of the largest eigenvalue of XX'/T for X's that have i.i.d. columns with correlated and not necessarily Gaussian elements. Paul (2006) studies the asymptotic distribution of the largest eigenvalue of XX'/T for Gaussian X's.

Our Corollary 2 states essentially the same result as Theorem 3 in Paul (2006). However, we derive Corollary 2 from a more general result (Theorem 5), which allows the factors to be non-Gaussian and non-i.i.d. over time. Whether Paul's proofs can be extended to handle the case of non-i.i.d. and non-Gaussian factors is an open question. Paul starts his proofs from the fact that if X has i.i.d. Gaussian columns with spiked covariance matrix $\Sigma = diag\left(\lambda_1, ..., \lambda_k, 1, 1, ..., 1\right)$, then it can be represented in the form $X' = \left[Z_1'\Lambda^{1/2}, Z_2'\right]$, where $\Lambda = diag\left(\lambda_1, ..., \lambda_k\right)$ and $[Z_1', Z_2']$ is a $T \times n$ matrix with i.i.d. Gaussian elements. Then he uses the above representation to obtain a non-linear equation that the first k largest eigenvalues of XX'/T must satisfy, simplifies these equations and derives his asymptotic results based on these simplifications. In the case of non-i.i.d. and non-Gaussian factors that we analyze in this paper, X cannot be represented in Paul's form. Instead, an orthogonal transformation of X, $\tilde{X}$, can be represented in the form $\tilde{X}' = [Z_1' + Q, Z_2']$, where Q is a diagonal matrix that depends on factors and factor loadings. Based on this representation one may, potentially, write an equation analogous to that used by Paul and try to work from this equation. This is, however, not the way we followed in this paper. We learned about Paul's work only after the first draft of this paper was written so that none of our proofs mimic or extend his proofs.

There has been much less previous work on the eigenvectors of the sample covariance matrices of large dimension. It is well known (see Anderson, 1984) that the matrix of eigenvectors of XX'/T, where X has i.i.d. standard Gaussian entries, is Haar distributed on the orthogonal group O_n. For X with non-Gaussian i.i.d. elements, Silverstein (1990) shows weak convergence of random functions defined by the eigenvectors of XX'/T. Reimann et al. (1996) consider i.i.d. observations of a high-dimensional Gaussian vector that has a systematic and idiosyncratic components and study the process of learning the direction of the systematic component. They describe the phenomenon of "retarded classification" when, if the ratio of the dimensionality of the vector to the number of the i.i.d. observations is above a certain threshold, nothing at all can be learned about the systematic direction. If the ratio is above the threshold, then as the dimensionality and the number of observations grow proportionally, the cosine of the angle between the best hypothesized direction and the true direction converges to a number from the interval $(0,1)$. The same phenomenon is observed when there are several systematic directions to learn about (see Hoyle and Rattray (2007)). The "retarded classification" phenomenon is directly related to our finding, mentioned in the introduction, that when the cumulative effects of the factors are less than a threshold (which depends on the ratio n/T), the principal components estimates of the factors point in completely random directions, and when the cumulative effects are larger than the threshold, the estimates asymptotically "live" on a cone around the true direction (see Figure 1).

For a 1-factor model with i.i.d. Gaussian factor, Johnstone and Lu (2004) show that the cosine of the angle between principal eigenvector of the sample covariance matrix and the principal eigenvector of the population covariance matrix remains separated from zero as n and T go to infinity proportionately. Paul (2006) quantifies the amount of the inconsistency pointed out by Johnstone and Lu (2004) for the case of i.i.d. Gaussian data such that all but k distinct eigenvalues of the population covariance matrix are the same. For the same model, Paul (2006) finds the asymptotic distribution of the eigenvectors corresponding to the k largest eigenvalues. Our paper generalizes Paul's (2006) result to factor models with non-i.i.d. and non-Gaussian factors. As was mentioned above, a non-linear equation which is central for Paul's proofs breaks down for the case of general factors. Our proofs, therefore, use different machinery than that used in Paul (2006).

3 Model, assumptions, and main results

We assume that (1) satisfies Assumptions 1 (or 1'), 2, and 3, formulated below.

In what follows, $A_{i\cdot}$ ($A_{\cdot i}$) denotes the i-th row (column) of matrix A, and I_i denotes an i-dimensional identity matrix. Our first assumption comes in two varieties. Assumption 1 treats factors as random variables. It allows us to identify factor loadings. Assumption 1' deals with deterministic factors. It allows us to identify both factor loadings and factors. Both assumptions are standard (see Anderson (1984), pp. 552-553).

Assumption 1: *For each $n \geq 1$, factors $\left\{F_{t\cdot}^{(n)\prime}; t=1,...,T^{(n)}\right\}$ form a sample of length $T^{(n)}$ from a stationary zero-mean $k \times 1$ vector process, normalized so that $E\left(F_{t\cdot}^{(n)\prime}F_{t\cdot}^{(n)}\right) = I_k$. The loadings are normalized so that the first non-zero elements* *of the columns of* $L^{(n)}$ *are positive and* $L^{(n)\prime}L^{(n)}$ *is a* $k \times k$ *diagonal matrix with non-increasing positive elements along the diagonal.*

In the special case when the rows of $F^{(n)}$ represent i.i.d. observations of normally distributed factors, model (1) becomes the so-called spherical Gaussian case of the standard factor model (see Anderson (1984)).

Assumption 1': *For each* $n \geq 1$, *factors form a deterministic sequence of* k*-dimensional vectors. The factors are normalized so that* $F^{(n)\prime}F^{(n)}/T^{(n)} = I_k$ *and the loadings are normalized so that the first non-zero elements of the columns of* $L^{(n)}$ *are positive and* $L^{(n)\prime}L^{(n)}$ *is a* $k \times k$ *diagonal matrix with non-increasing positive elements along the diagonal.*

The next assumption allows us to make orthogonal transformations of the data without changing the joint distribution of the noise components. A particularly important property of the Gaussian noise that we use in this paper is that the orthogonal matrix of eigenvectors of the sample covariance matrix of such noise has conditional Haar invariant distribution (see Anderson (1984), p.536).

Assumption 2: *For each* $n \geq 1$, *entries of* $\varepsilon^{(n)}$ *are i.i.d.* $N(0, \sigma^2)$ *random variables independent of the factors.*

Our last assumption describes the conditions that need to be satisfied for the asymptotic analysis below to be correct as n goes to infinity.

Assumption 3: *There exist a scalar* $c > 0$ *and a* $k \times k$ *diagonal matrix* $D \equiv \operatorname{diag}(d_1, ..., d_k)$, $d_1 > ... > d_k > 0$[1], *such that, as* $n \to \infty$,
i) $n/T^{(n)} - c = o\left(n^{-1/2}\right)$,
ii) $L^{(n)\prime}L^{(n)} - D = o\left(n^{-1/2}\right)$, *where the equality should be understood in the element by element sense,*
iii) $\sqrt{T^{(n)}}\left(\frac{1}{T^{(n)}}F^{(n)\prime}F^{(n)} - I_k\right) \xrightarrow{d} \Phi$, *where entries of* Φ *have a joint normal distribution (degenerate in the case of deterministic factors) with covariance function* $\operatorname{cov}(\Phi_{st}, \Phi_{s_1t_1}) \equiv \phi_{sts_1t_1}$.

Part i) of the assumption requires that n and $T^{(n)}$ be comparable even asymptotically. The requirement that the convergence is faster than $n^{-1/2}$ eliminates any possible effects of this convergence on our asymptotic results. In our opinion, the

[1] We generalized Theorem 5 to the case of some or all of the diagonal elements of D being the same. To save space, we do not report these results below.

behavior of $n/T^{(n)}$ is likely to be application-specific and any consequential assumption about the rate of convergence of $n/T^{(n)}$ will be arbitrary. The assumption about the rate of convergence of $L^{(n)\prime}L^{(n)}$ is made for the same reason. A generalization of part ii) would be to assume that $L^{(n)\prime}L^{(n)}/q^{(n)} - D = o\left(n^{-1/2}\right)$, where $q^{(n)}$ is some deterministic sequence. Although very interesting, such a generalization creates some difficult technical problems in our proofs, so we do not consider it here.

The high-level assumption about the convergence of $\sqrt{T^{(n)}}\left(\frac{1}{T^{(n)}}F^{(n)\prime}F^{(n)} - I_k\right)$ is important because parameters $\phi_{sts_1t_1}$ enter our asymptotic formulae established below. A primitive assumption that implies the convergence is that the individual factors can be represented as infinite linear combinations, with absolutely summable coefficients, of i.i.d. random variables with a finite fourth moment (see Anderson (1971), Theorem 8.4.2). In the special case when $F_{t\cdot}^{(n)}$ are i.i.d. standard multivariate normal, the covariance function of the asymptotic distribution of $\sqrt{T^{(n)}}\left(\frac{1}{T^{(n)}}F^{(n)\prime}F^{(n)} - I_k\right)$ has a particularly simple form: $\phi_{iji_1j_1} = 2$ if $(i,j) = (i_1, j_1)$ and $i = j$, $\phi_{iji_1j_1} = 1$ if $(i,j) = (i_1, j_1)$ or $(i,j) = (j_1, i_1)$ and $i \neq j$, and $\phi_{iji_1j_1} = 0$ otherwise.

In this paper, we study the principal components estimators $\hat{F}^{(n)}$ and $\hat{L}^{(n)}$ of factors and factor loadings, respectively. To define the estimators we introduce the following notation. Denote the largest k eigenvalues of matrices $\frac{1}{T^{(n)}}X^{(n)}X^{(n)\prime}$ and $\frac{1}{T^{(n)}}X^{(n)\prime}X^{(n)}$ as $\mu_1^{(n)} \geq ... \geq \mu_k^{(n)}$. Note that the matrices have the identical sets of largest $\min\left(n, T^{(n)}\right)$ eigenvalues, and we assume that $k < \min\left(n, T^{(n)}\right)$. Further, denote the corresponding eigenvectors for $\frac{1}{T^{(n)}}X^{(n)}X^{(n)\prime}$ and $\frac{1}{T^{(n)}}X^{(n)\prime}X^{(n)}$ as $u_1^{(n)}, ..., u_k^{(n)}$, and $v_1^{(n)}, ..., v_k^{(n)}$, respectively. Then the principal components estimator $\hat{F}^{(n)}$ is defined as a matrix with columns $\sqrt{T^{(n)}}v_1^{(n)}, ..., \sqrt{T^{(n)}}v_k^{(n)}$, and the principal components estimator $\hat{L}^{(n)}$ is defined as $\frac{1}{T^{(n)}}X^{(n)}\hat{F}^{(n)}$. It is easy to verify that the i-th column of $\hat{L}^{(n)}$ is equal to $\sqrt{\mu_i^{(n)}}u_i^{(n)}$. Therefore, the square of the Euclidean length of $\hat{L}_{\cdot i}^{(n)}$, which estimates the cumulative effect of the i-th factor on the cross-sectional units, is equal to $\mu_i^{(n)}$, and the normalized principal components estimator of factor loadings $\hat{\mathcal{L}}^{(n)} \equiv \hat{L}^{(n)}\left(\hat{L}^{(n)\prime}\hat{L}^{(n)}\right)^{-1/2}$ is equal to a matrix with columns $u_1^{(n)}, ..., u_k^{(n)}$.

Of course, without further restrictions the eigenvectors $u_i^{(n)}$ and $v_i^{(n)}$, and, therefore, the principal components estimators $\hat{F}^{(n)}$ and $\hat{L}^{(n)}$, are defined only up to a change in the sign. To eliminate this indeterminacy, we require that the direction of the eigenvectors is chosen so that $u_i^{(n)\prime}L_{\cdot i}^{(n)} > 0$ and $v_i^{(n)\prime}F_{\cdot i}^{(n)} > 0$.

We now formulate and discuss our main results, postponing all proofs until Section

6. As in the Introduction, we will omit the superscript (n) from our notations to make them easier to read. For any $q \leq k$, denote the matrix of the first q columns of $\hat{F}$ as $\hat{F}_{1:q}$, and let $F_q^{\perp}$ be a $T \times q$ matrix with columns orthogonal to the columns of F such that the joint distribution of its entries conditional on F is invariant with respect to multiplication from the left by any orthogonal matrix having span (F) as its invariant subspace. We establish the following

Theorem 1: *Let q be such that $d_i > \sqrt{c}\sigma^2$ for $i \leq q$, and $d_i \leq \sqrt{c}\sigma^2$ for $i > q$. Let Assumptions 1 (or 1'), 2, and 3 hold and let, in addition, $\phi_{ijst} = 0$ when $(i,j) \neq (s,t)$ and $(i,j) \neq (t,s)$. Then, we have:*
i)

$$\begin{aligned} \hat{F}_{1:q} &= F \cdot Q + F_q^{\perp}, \\ Q &= Q^{(1)} + \frac{1}{\sqrt{T}} Q^{(2)}, \end{aligned}$$

where $Q^{(1)}$ is diagonal with $Q_{ii}^{(1)} = \sqrt{\frac{d_i^2 - \sigma^4 c}{d_i(d_i+\sigma^2)}}$, and $\operatorname{vec} Q^{(2)}$ is an asymptotically zero mean Gaussian vector with $\operatorname{Acov}\left(Q_{ij}^{(2)}, Q_{st}^{(2)}\right)$ given by the following formulae:

a) $\frac{\left(d_j^2+\sigma^2 d_i\right)}{(d_j-d_i)^2} + (\phi_{ijij} - 1)\frac{d_j\left(d_j^2 - c\sigma^4\right)}{(d_j+\sigma^2)(d_j-d_i)^2}$ *if* $(i,j) = (s,t)$ *and* $i \neq j$

b) $\frac{\sqrt{d_i d_j}\sqrt{(d_i+\sigma^2)(d_j+\sigma^2)\left(d_i^2-c\sigma^4\right)\left(d_j^2-c\sigma^4\right)}}{(d_j-d_i)^2(c\sigma^4-d_i d_j)} - \left(\phi_{ijij} - 1\right)\frac{\sqrt{d_i d_j}\sqrt{\left(d_i^2-c\sigma^4\right)\left(d_j^2-c\sigma^4\right)}}{(d_j-d_i)^2\sqrt{(d_j+\sigma^2)(d_i+\sigma^2)}}$ *if* $(i,j) = (t,s)$ *and* $i \neq j$

c) $\frac{\left(c^2\sigma^8+d_i^4\right)\left(d_i+\sigma^2\right)}{2d_i\left(d_i^2-c\sigma^4\right)^2} + \frac{d_i\sigma^4(c-1)}{2\left(d_i^2-c\sigma^4\right)(d_i+\sigma^2)} + (\phi_{iiii} - 2)\frac{\left(\left(d_i+\sigma^2\right)^2 - \sigma^4(1-c)\right)^2 d_i}{4\left(d_i^2-c\sigma^4\right)(d_i+\sigma^2)^3}$ if $(i,j) = (t,s)$ *and* $i = j$

d) 0 *if* $(i,j) \neq (s,t)$ *and* $(i,j) \neq (t,s)$

ii) $\hat{F}_{q+1:k} = F \cdot \tilde{Q} + F_{k-q}^{\perp}$, *where* $\tilde{Q} \xrightarrow{p} 0$ *as* $n \to \infty$.

A graphic interpretation of the above representation of $\hat{F}_{1:q}$ for the case of deterministic factors was given in the Introduction. In the case of random factors, the interpretation is complicated by the fact that the columns of F have random length, not necessarily equal to $\sqrt{T}$. Hence, "vector" F in Figure 1 does not "live" on the sphere and a potential graphic interpretation would not be so clean as in the case of deterministic factors. The theorem's requirement that $\phi_{ijst} = 0$ when $(i,j) \neq (s,t)$ and $(i,j) \neq (t,s)$ holds, for example, if the different factors are mutually independent. It trivially holds if the factors are treated as non-random. This requirement was introduced solely to simplify formulae for Acov $\left(Q_{ij}^{(2)}, Q_{st}^{(2)}\right)$, which would otherwise become non-trivial even in the case $(i,j) \neq (s,t)$ and $(i,j) \neq (t,s)$.

The formula $Q_{ii}^{(1)} = \sqrt{(d_i^2 - \sigma^4 c)/d_i(d_i + \sigma^2)}$ established in Theorem 1 reveals a trade-off associated with using more cross-sectional data for factor estimation. On one hand, using more cross-sectional data may call for using higher d_i in the approximating asymptotics, which would increase $Q_{ii}^{(1)}$ and, hence, decrease the bias in the estimate of the factor. On the other hand, using more data would increase the ratio n/T, which would be associated with higher c. The rise in c will lead to a decrease in $Q_{ii}^{(1)}$ and, hence, to an increase in the bias. That more data are not always better for the estimation was empirically demonstrated by Boivin and Ng (2006). They explain that an estimator that uses more data may be less efficient depending on the information content of the new data. Our formula provides an additional theoretical justification for the Boivin-Ng observation.

Our next result is an analog of Theorem 1 for factor loadings. Denote the matrix of normalized factor loadings $L(L'L)^{-1/2}$ as $\mathcal{L}$ and let $\mathcal{L}_q^{\perp}$ be an $n \times q$ random matrix with columns orthogonal to the columns of $\mathcal{L}$ and such that the joint distribution of its entries is invariant with respect to multiplication from the left by any orthogonal matrix having $\operatorname{span}(\mathcal{L})$ as its invariant subspace. We have the following

Theorem 2: *Let q be such that $d_i > \sqrt{c}\sigma^2$ for $i \leq q$, and $d_i \leq \sqrt{c}\sigma^2$ for $i > q$. Let Assumptions 1 (or 1'), 2, and 3 hold and let, in addition, $\phi_{ijst} = 0$ when $(i,j) \neq (s,t)$ and $(i,j) \neq (t,s)$. Then, we have:*
i)

$$\begin{aligned} \hat{\mathcal{L}}_{1:q} &= \mathcal{L} \cdot R + \mathcal{L}_q^{\perp}, \\ R &= R^{(1)} + \frac{1}{\sqrt{T}} R^{(2)}, \end{aligned}$$

where $R^{(1)}$ is diagonal with $R_{ii}^{(1)} = \sqrt{\frac{d_i^2 - \sigma^4 c}{d_i(d_i + \sigma^2 c)}}$, and $\operatorname{vec} R^{(2)}$ is an asymptotically zero mean Gaussian vector with Acov $\left(R_{ij}^{(2)}, R_{st}^{(2)}\right)$ *given by the following formulae:*

a) $\frac{d_j(d_j+\sigma^2)(d_i+\sigma^2) + d_i(\phi_{ijij}-1)(d_j^2 - \sigma^4 c)}{(d_j + \sigma^2 c)(d_j - d_i)^2}$ *if* $(i,j) = (s,t)$ *and* $i \neq j$

b) $-\frac{\sqrt{d_i d_j}\sqrt{(d_i^2 - \sigma^4 c)(d_j^2 - \sigma^4 c)}}{(d_j - d_i)^2 \sqrt{(d_i + \sigma^2 c)(d_j + \sigma^2 c)}} \left(\phi_{ijij} - 1 + \frac{(d_j + \sigma^2)(d_i + \sigma^2)}{(d_i d_j - c\sigma^4)}\right)$ *if* $(i,j) = (t,s)$ *and* $i \neq j$

c) $\frac{c\sigma^4 d_i\left(d_i+\sigma^2\right)^2}{2(d_i+c\sigma^2)\left(d_i^2-c\sigma^4\right)^2}\left(1+c\left(\frac{d_i+\sigma^2}{d_i+c\sigma^2}\right)^2\right)+(\phi_{iiii}-2)\frac{\left(\left(d_i+\sigma^2\right)^2-\sigma^4(1-c)\right)^2 c^2\sigma^4}{4d_i\left(d_i^2-\sigma^4 c\right)(d_i+c\sigma^2)^3}$ if $(i,j)=(t,s)$ *and* $i=j$

d) 0 *if* $(i,j)\neq(s,t)$ *and* $(i,j)\neq(t,s)$

ii) $\hat{\mathcal{L}}_{q+1:k}=\mathcal{L}\cdot\tilde{R}+\mathcal{L}_{k-q}^{\perp}$, where $\tilde{R}\xrightarrow{p}0$ *as* $n\to\infty$.

Theorems 1 and 2 can be used to obtain the asymptotic distributions of the principal components estimator of factors at particular time periods or factor loadings corresponding to specific cross-sectional units. We find such distributions in Theorems 3 and 4 below. Let δ_{ij} denote the Kronecker delta. Then we have:

Theorem 3: *Suppose the assumptions of Theorem 1 hold. Let* $\tau_1,...,\tau_r$ *be such that the probability limits of the* τ_1*-th,...,*τ_r*-th rows of matrix* $F/\sqrt{T}$ *as* n *and* T *approach infinity exist and equal* $\bar{F}_{\tau_1\cdot},...,\bar{F}_{\tau_r\cdot}$*. Then,*
i) Random variables $\left\{\hat{F}_{\tau_g i}-Q_{ii}^{(1)}F_{\tau_g i}:\ g=1,...,r;\ i=1,...,q\right\}$ *are asymptotically jointly mean-zero Gaussian. The asymptotic covariance between* $\hat{F}_{\tau_s i}-Q_{ii}^{(1)}F_{\tau_s i}$ *and* $\hat{F}_{\tau_f p}-Q_{pp}^{(1)}F_{\tau_f p}$ *is equal to* $\sum_{s=1}^{k}\bar{F}_{\tau_g s}\bar{F}_{\tau_f s}\operatorname{Avar}\left(Q_{si}^{(2)}\right)+\left(\delta_{gf}-\sum_{s=1}^{k}\bar{F}_{\tau_g s}\bar{F}_{\tau_f s}\right)\left(1-\left(Q_{ii}^{(1)}\right)^2\right)$ *when* $i=p$ *and to* $-\bar{F}_{\tau_g p}\bar{F}_{\tau_f i}\operatorname{Acov}\left(Q_{pi}^{(2)},Q_{ip}^{(2)}\right)$ *when* $i\neq p$.
ii) For any $i>q$*, and any* $\tau\leq T$*,* $\hat{F}_{\tau i}/\sqrt{T}\xrightarrow{p}0$.

When factors are deterministic, allowing for non-zero limits $\bar{F}_{\tau_1\cdot},...,\bar{F}_{\tau_r\cdot}$ takes into account a possibility that special time periods exist for which the values of some or all factors are "unusually" large. Alternatively, non-zero limits $\bar{F}_{\tau_1\cdot},...,\bar{F}_{\tau_r\cdot}$ can be viewed as a device to improve asymptotic approximation for relatively small T when the rows of $F/\sqrt{T}$ are not expected to be small. When the factors are random and satisfy Assumption 1, then, obviously, the probability limits $\bar{F}_{\tau_1\cdot},...,\bar{F}_{\tau_r\cdot}$ exist and equal zero. In such a case, the above formula for the asymptotic covariance between $\hat{F}_{\tau_s i}-Q_{ii}^{(1)}F_{\tau_s i}$ and $\hat{F}_{\tau_f p}-Q_{pp}^{(1)}F_{\tau_f p}$ simplifies to $\delta_{gf}\frac{\sigma^2\left(d_i+\sigma^2 c\right)}{d_i(d_i+\sigma^2)}$ if $i=p$ and to zero if $i\neq p$.

Theorem 3 can be compared to Theorem 1 of Bai (2003). He finds that, under his "strong-factor" requirement, $\sqrt{n}\left(\hat{F}_{t\cdot}-H'F_{t\cdot}\right)\xrightarrow{d}N\left(0,\Omega\right),$ where H and Ω are matrices that depend on the parameters describing factors, loadings, and noise. For our normalization of factors and factor loadings, it can be shown that H equals the identity matrix and Ω must be well approximated by $n\sigma^2 D^{-1}$ in large samples. Hence, Bai's asymptotic approximation of the finite sample distribution of $\hat{F}_{ti}-F_{ti}$

can be represented as $N\left(0,\frac{\sigma^2}{d_i}\right)$. The variance of the latter distribution is close to our asymptotic variance $\frac{\sigma^2\left(d_i+\sigma^2 c\right)}{d_i(d_i+\sigma^2)}$ when d_i is very large, as it should be under the "strong-factor" assumption, or if c is close to 1. Note that the multiplier $Q_{ii}^{(1)}$, causing the inconsistency of $\hat{F}_{ti}$ in our case, becomes very close to 1 as d_i increases. Hence, Bai's asymptotic formula is consistent with ours in the case of factors with very large cumulative effects on the cross-sectional units.

For factor loadings, we have the following:

Theorem 4: *Suppose the assumptions of Theorem 2 hold. Let $j_1, ..., j_r$ be such that the limits of the j_1-th, ..., j_r-th rows of matrix $\mathcal{L}$ as n and T go to infinity exist and equal $\bar{\mathcal{L}}_{j_1\cdot}, ..., \bar{\mathcal{L}}_{j_r\cdot}$. Then,*
i) Random variables $\left\{\sqrt{T}\left(\hat{\mathcal{L}}_{j_g i} - R_{ii}^{(1)}\mathcal{L}_{j_g i}\right),\ g = 1, ..., r;\ i = 1, ..., q\right\}$ are asymptotically jointly mean-zero Gaussian. The asymptotic covariance between $\sqrt{T}\left(\hat{\mathcal{L}}_{j_g i} - R_{ii}^{(1)}\mathcal{L}_{j_g i}\right)$ and $\sqrt{T}\left(\hat{\mathcal{L}}_{j_f p} - R_{pp}^{(1)}\mathcal{L}_{j_f p}\right)$ equals
$\sum_{s=1}^{k}\bar{\mathcal{L}}_{j_g s}\bar{\mathcal{L}}_{j_f s}\,\mathrm{Avar}\left(R_{si}^{(2)}\right) + \left(\delta_{gf} - \sum_{s=1}^{k}\bar{\mathcal{L}}_{j_g s}\bar{\mathcal{L}}_{j_f s}\right)\left(1 - \left(R_{ii}^{(1)}\right)^2\right)$ *when $i = p$ and equals $-\bar{\mathcal{L}}_{j_g p}\bar{\mathcal{L}}_{j_f i}\,\mathrm{Acov}\left(R_{pi}^{(2)}, R_{ip}^{(2)}\right)$ when $i \neq p$.*
ii) For any $i > q$, and any $j \leq n$, $\hat{\mathcal{L}}_{ji} \xrightarrow{p} 0$

For the special case when the factors are i.i.d. k-dimensional standard normal variables, the formula for the asymptotic covariance of the components of $\hat{\mathcal{L}}$ simplifies. We have:

Corollary 1: *Suppose that, in addition to the assumptions of Theorem 4, the factors $F_{t\cdot}$ are i.i.d. standard multivariate random variables. Then, for any $i \leq q$*

$$\sqrt{T}\left(\left(\hat{\mathcal{L}}_{j_1 i} - R_{ii}^{(1)}\mathcal{L}_{j_1 i}\right), ..., \left(\hat{\mathcal{L}}_{j_r i} - R_{ii}^{(1)}\mathcal{L}_{j_r i}\right)\right) \xrightarrow{d} N\left(0, \Gamma\right),$$

where

$$\begin{aligned}\Gamma_{gf} &= \sum_{\substack{s=1\\ s\neq i}}^{k}\bar{\mathcal{L}}_{j_g s}\bar{\mathcal{L}}_{j_f s}\frac{d_i\left(d_i+\sigma^2\right)\left(d_s+\sigma^2\right)}{\left(d_i+c\sigma^2\right)\left(d_i-d_s\right)^2} + \left(\delta_{gf} - \sum_{s=1}^{k}\bar{\mathcal{L}}_{j_g s}\bar{\mathcal{L}}_{j_f s}\right)\frac{\sigma^2\left(d_i+\sigma^2\right)}{d_i\left(d_i+c\sigma^2\right)} \\ &\quad + \bar{\mathcal{L}}_{j_g i}\bar{\mathcal{L}}_{j_f i}\frac{c\sigma^4 d_i\left(d_i+\sigma^2\right)^2}{2\left(d_i+c\sigma^2\right)\left(d_i^2-c\sigma^4\right)^2}\left(1 + c\left(\frac{d_i+\sigma^2}{d_i+c\sigma^2}\right)^2\right)\end{aligned}$$

Note that when factors are i.i.d. Gaussian random variables, the principal components estimator of the normalized factor loadings is equal to the matrix of the principal eigenvectors of the sample covariance matrix of i.i.d. Gaussian data. The asymptotic distribution of such principal eigenvectors in the case when only T approaches infinity is well known. According to Theorem 13.5.1 of Anderson (1984),

$$\sqrt{T}\left(\hat{\mathcal{L}}_{\cdot i}-\mathcal{L}_{\cdot i}\right)\rightarrow N\left(0,\Pi\right), \tag{4}$$

where

$$\Pi_{gf}=\sum_{\substack{s=1\\ s\neq i}}^{n}\mathcal{L}_{gs}\mathcal{L}_{fs}\frac{\left(d_i+\sigma^2\right)\left(d_s+\sigma^2\right)}{\left(d_i-d_s\right)^2} \tag{5}$$

and it is understood that $\mathcal{L}_{\cdot s}$ is defined as the eigenvector of the population covariance matrix corresponding to the s-th largest eigenvalue, and $d_s=0$ for $s>k$. Note that $\sum_{s=k+1}^{n}\mathcal{L}_{gs}\mathcal{L}_{fs}=\delta_{gf}-\sum_{s=1}^{k}\mathcal{L}_{gs}\mathcal{L}_{fs}$ because the matrix of "population eigenvectors" is orthogonal. Therefore, we can rewrite (5) as

$$\Pi_{gf}=\sum_{\substack{s=1\\ s\neq i}}^{k}\mathcal{L}_{gs}\mathcal{L}_{fs}\frac{\left(d_i+\sigma^2\right)\left(d_s+\sigma^2\right)}{\left(d_i-d_s\right)^2}+\left(\delta_{gf}-\sum_{s=1}^{k}\mathcal{L}_{gs}\mathcal{L}_{fs}\right)\frac{\sigma^2\left(d_i+\sigma^2\right)}{d_i^2}. \tag{6}$$

Since in the classical case n is fixed, the requirement that rows of $\mathcal{L}$ have limits as T approaches infinity is trivially satisfied. For the same reason, there is no need to focus attention on a subset of components $j_1,...,j_r$ of the "population eigenvectors", so that formula (4) describes the asymptotic behavior of all components of $\mathcal{L}_{\cdot i}$. More substantially, the large dimensionality of the data introduces inconsistency (towards zero) to the components of $\hat{\mathcal{L}}_{\cdot i}$ viewed as estimates of the corresponding components of $\mathcal{L}_{\cdot i}$. Indeed, from Corollary 1, we see that the probability limit of $\hat{\mathcal{L}}_{j_s i}$ equals $\mathcal{L}_{j_s i}$ multiplied by $0\leq R_{ii}^{(1)}<1$. Comparing Π and Γ, we see that the high dimensionality of data introduces a new component to the asymptotic covariance matrix, which depends solely on the limits of the components of the i-th "population eigenvector". At the same time, it reduces the "classical component" of the asymptotic covariance by multiplying it by $\frac{d_i}{d_i+c\sigma^2}$. As c becomes very small, our formula for Γ_{gf} converges to the classical formula for Π_{gf}, as should be the case, intuitively.

The asymptotic result for high-dimensional data differs strikingly from the classical result when d_i is below the threshold $\sqrt{c}\sigma^2$. In such a case, $\hat{\mathcal{L}}_{\cdot i}$ has nothing to do with $\mathcal{L}_{\cdot i}$. It just points out the direction of maximal spurious "explanatory power" of the idiosyncratic terms. It is only when the cumulative effect of the i-th factor on the cross-sectional units passes the threshold that $\hat{\mathcal{L}}_{\cdot i}$ becomes related to $\mathcal{L}_{\cdot i}$. As d_i becomes larger and larger, components of $\hat{\mathcal{L}}_{\cdot i}$ approximate those of $\mathcal{L}_{\cdot i}$ better and better, eventually matching them.

The rest of our results concern the asymptotic behavior of eigenvalues $\mu_1, ..., \mu_k$ which, as explained above, can be interpreted as the principal components estimators of the cumulative effects of the 1st, 2nd,..., k-th factors, respectively, on the cross-sectional units. In fact, a better estimator of the cumulative effect of the i-th factor would be $\mu_i - \hat{\sigma}^2$, where $\hat{\sigma}^2$ is any consistent estimator of σ^2. This can be understood by noting that the i-th eigenvalue of the population covariance matrix of data $EX_{\cdot t}X'_{\cdot t}$ equals $d_i + \sigma^2$, where d_i is the true cumulative effect. According to our next theorem, even such a corrected estimator would be inconsistent.

Theorem 5: *Let q be such that $d_i > \sqrt{c}\sigma^2$ for $i \leq q$, and $d_i \leq \sqrt{c}\sigma^2$ for $i > q$. For $i = 1, ..., q$, define constants $m_i = \frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2 c\right)}{d_i}$. Under Assumptions 1 or (1'), 2, and 3, we have:*
i) $\sqrt{T}\left(\mu_1 - m_1, ..., \mu_q - m_q\right)' \xrightarrow{d} N\left(0, \Sigma\right),$ *where*

$$\Sigma_{ij} = \phi_{iijj}\frac{\left(d_i^2 - \sigma^4 c\right)\left(d_j^2 - \sigma^4 c\right)}{d_i d_j} + 2\delta_{ij}\sigma^2\left(2d_j + c\sigma^2 + \sigma^2\right)\frac{d_i^2 - \sigma^4 c}{d_i^2}$$

ii) For any $i > q$, $\mu_i \xrightarrow{p} \left(1 + \sqrt{c}\right)^2 \sigma^2$

Note that according to Theorem 5, $\mu_i - \sigma^2$ converges to $m_i - \sigma^2 = d_i + c\sigma^2\left(1 + \frac{\sigma^2}{d_i}\right) > d_i$. Hence, if we estimate the cumulative effect of the i-th factor by subtracting a true known σ^2 from μ_i, we are making a systematic positive mistake, which may be very large if c and σ^2 are large.

In the case of deterministic factors, the formula for the asymptotic covariance matrix significantly simplifies because $\phi_{iijj} \equiv 0$. The formula also simplifies in the case when the factors are i.i.d. standard multivariate normal random variables. In such a case, we have

Corollary 2: *If, in addition to the assumptions of Theorem 5, factors $F_{t\cdot}$ are i.i.d. standard multivariate normal random variables, then $\sqrt{T}\left(\mu_1 - m_1, ..., \mu_q - m_q\right)' \xrightarrow{d} N\left(0, \Sigma\right)$, where Σ is a diagonal matrix such that $\Sigma_{ii} = 2\left(d_i + \sigma^2\right)^2\left(1 - \frac{\sigma^4 c}{d_i^2}\right)$.*

If we keep the framework of the above corollary, but consider the classical case, when only T goes to infinity, then according to Theorem 13.5.1 of Anderson (1984), μ_i consistently estimates $d_i + \sigma^2$, and the asymptotic variance of μ_i is equal to $2\left(d_i + \sigma^2\right)^2$. This result can be recovered by setting $c = 0$ in Corollary 2. We see that the large dimensionality of the data introduces inconsistency but reduces the asymptotic variance of μ_i, viewed as an estimate of $d_i + \sigma^2$. Indeed, under our assumptions, the probability limit of μ_i is $d_i + \sigma^2$, multiplied by $\frac{d_i+\sigma^2 c}{d_i} > 1$, and the asymptotic variance is $2\left(d_i + \sigma^2\right)^2$ multiplied by $1 - \frac{\sigma^4 c}{d_i^2}$, which is positive if $i \leq q$, but less than 1.

A striking difference from the classical case occurs when the cumulative effect of the i-th factor on the cross-sectional units, measured by d_i, is below the threshold $\sqrt{c}\sigma^2$. In such a case, the i-th largest eigenvalue of $\frac{1}{T}XX'$ converges to a constant $\left(1+\sqrt{c}\right)^2\sigma^2$ which does not depend on d_i. Hence, if the cumulative effect of the i-th factors on the cross-sectional units is weak relative to the variance of idiosyncratic noise and/or if the number of the cross-sectional units in the sample is much larger than the number of the observations, the size of the i-th largest "sample eigenvalue" does not reflect the strength of the cumulative effect, but measures the maximal amount of variation in the data that can be spuriously "explained" by a linear combination of the idiosyncratic terms. The i-th largest "sample eigenvalue" starts to be related to the cumulative effect of the i-th factor only after the cumulative effect passes the threshold.

4 An intuition for the inconsistency result

To see intuitively why the principal components estimator is inconsistent, consider a special situation when $F = \left(\sqrt{T}, 0, ..., 0\right)'$ and $L = \left(\sqrt{d}, 0, ..., 0\right)'$, where d measures the cumulative effect of F on the cross-sectional units. In such a case, matrix $X'X/T$ can be decomposed into a sum of two matrices: $X'X/T = d\left(F+\varepsilon_1\right)\left(F+\varepsilon_1\right)'/T + \varepsilon_{-1}'\varepsilon_{-1}/T$, where ε_1' is the first row of the matrix of idiosyncratic terms ε, and ε_{-1} is obtained from ε by deleting its first row. By definition, the principal components estimator $\hat{F}$ is a vector of length $\sqrt{T}$ which maximizes $\hat{F}'\left(X'X/T\right)\hat{F} = d\hat{F}'\left[\left(F+\varepsilon_1\right)\left(F+\varepsilon_1\right)'/T\right]\hat{F} + \hat{F}'\left[\varepsilon_{-1}'\varepsilon_{-1}/T\right]\hat{F}$. Had $\hat{F}$ been maximizing just the first term in the sum, it would have been close to F. However, the maximization is achieved by balancing the marginal gains from increasing the first and the second terms. Hence, for $\hat{F}$ to be close to F one of the following two scenarios must hold. Either d is very large so that the high weight is put on the maximization of the first term, or matrix $\varepsilon'_{-1}\varepsilon_{-1}/T$ is close to the identity matrix so that the second term is insensitive to the choice of $\hat{F}$. However, the first scenario is ruled out because we do not want to assume the overwhelming domination of factors over the idiosyncratic influences, and the second scenario does not hold because, although the elements of ε_{-1} are i.i.d., the dimensionality of each row of ε_{-1} is so large that there always exists a spurious direction which seems to agree with the directions of a significant proportion of the rows. Therefore, the principal components estimator ends up mixing the direction of the true factor with a spurious direction along which the variation of the idiosyncratic terms seems to be maximized.

Although this intuition explains inconsistency of the principal components estimator, it does not explain why there exists a separation between the directions of F and $\hat{F}$ which makes the darker region in Figure 1 look like a ring rather than a cap. Such a separation is closely related to an observation, which Milman (1988) credits to Poincaré, that in spaces of large dimensions, a randomly chosen direction is nearly orthogonal to any fixed direction with high probability. Since the spurious direction of high idiosyncratic variation is completely random, it turns out to be nearly orthogonal to the factor direction. Therefore, the principal components estimator mixes the factor direction not just with some other direction, but with a nearly orthogonal direction, which leads to a separation between F and $\hat{F}$ with high probability.

5 A Monte Carlo study

In this section we will perform a Monte Carlo analysis to check whether our asymptotic results approximate finite sample situations well. We perform three different experiments. The setting of our first experiment is as follows. We simulate 1000 replications of data having 1-factor structure with $n = 40$, $T = 20$, where F_{t1} is an AR(1) process with AR coefficient 0.5 and variance 1, $\sigma^2 = 1$, $L_{i1} = \sqrt{d/n}$, and d is on a grid 0.1:0.1:20. We repeat the experiment for $n = 200$, $T = 100$. Figure 2 shows the Monte Carlo and theoretical means and 5% and 95% quantiles of the regression coefficient in the regression of $\hat{F}$ on F as functions of d. Smooth solid lines correspond to the theoretical lines obtained using formulae of Theorem 1. According to that theorem, the regression coefficient should be equal to $Q^{(1)} + \frac{1}{\sqrt{T}}Q^{(2)}$. Note that

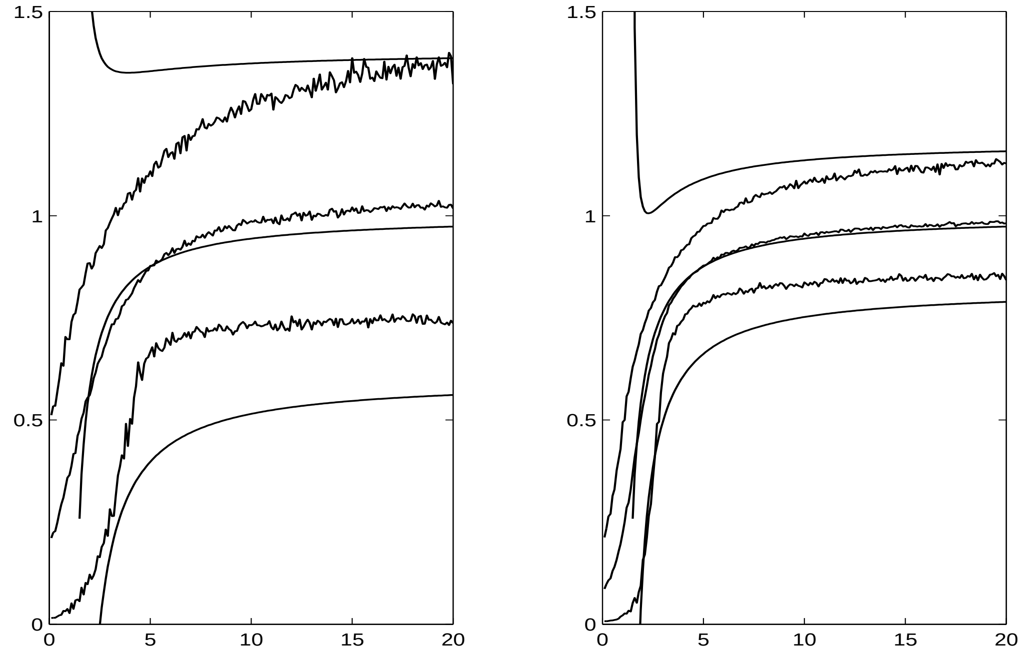

Figure 2: Monte Carlo and theoretical means and 5% and 95% quantiles of the regression coefficient in the regression of $\hat{F}$ on F as functions of d. Horizontal axis: d. Left panel: $n = 40, T = 20$; right panel: $n = 200, T = 100$.

the theoretical lines do not start from $d = 0.1$. It is because our formulae are valid for d larger than the threshold, which is equal to $\sqrt{2}$ in all Monte Carlo experiments below. Rough solid lines correspond to the Monte Carlo sample data. The left panel is for $n = 40$, $T = 20$. The right panel is for $n = 200$, $T = 100$.

The theoretical mean of the regression coefficient, $Q^{(1)}$, approximates the Monte Carlo mean reasonably well for $n = 40$, $T = 20$ and very well for $n = 200$, $T = 100$. For relatively small cumulative effects of the factor, the asymptotic quantiles tend to overestimate the amount of finite sample variation in the coefficient. When the cumulative effect approaches the threshold $\sqrt{2}$, the amount of overestimation explodes.

In our next experiment, we simulate 1000 replications of data having 2-factor structure with $n = 40$, $T = 20$, where F_{t1} and F_{t2} are i.i.d. $N(0, 1)$, $\sigma^2 = 1$, and the factor loadings are defined as follows. We set $L'_{\cdot 1}L_{\cdot 1} = 10\sqrt{2}$ and $L'_{\cdot 2}L_{\cdot 2} = 2\sqrt{2}$, so that the cumulative effect of the first factor on the cross-sectional units is 10 times the threshold, and the cumulative effect of the second factor is only 2 times the threshold. The vectors of loadings are designed so that their first two components are "unusually" large and the other components are equal by absolute value. Precisely, $L_{11} = L_{21} = \left(10\sqrt{2}/3\right)^{1/2}$, $L_{i1} = \left(10\sqrt{2}/3\,(n-2)\right)^{1/2}$ for $i > 2$, and $L_{12} = -L_{22} = -\left(2\sqrt{2}/3\right)^{1/2}$, $L_{i1} = (-1)^i \left(2\sqrt{2}/3\,(n-2)\right)^{1/2}$ for $i > 2$.

Figure 3 shows the results of the second experiment. The upper three graphs correspond to the joint distributions of (from left to right) the (1st, 2nd), (2nd, 3rd), and (3rd, 4th) components of the normalized (to have unit length) vector of factor loadings corresponding to the first factor. The bottom three graphs correspond to the joint distributions of the same components of the normalized vector of factor loadings corresponding to the second factor. The dots on the graphs correspond to the Monte Carlo draws, the solid lines correspond to 95% confidence ellipses of our theoretical asymptotic distribution (see Corollary 1), the dashed lines correspond to the 95% confidence ellipses of the classical asymptotic distribution (see equation 6), and the dotted lines correspond to the 95% confidence ellipses of the asymptotic distribution under the "strong factor" requirement.

Starting from the upper left graph and going in a clockwise direction, the percentage of the Monte Carlo draws falling inside our ellipse, a classical ellipse, and a "strong factor ellipse" are, respectively, $(90, 63, 64)$, $(92, 91, 76)$, $(92, 94, 93)$, $(93, 98, 94)$, $(87, 64, 66)$, and $(84, 23, 47)$. Of course, ideally the percentage should be equal to 95. We see that our asymptotic distribution provides a much better approximation to the Monte Carlo distribution than the classical and the "strong factor" asymptotic distributions. The advantage of our distribution is particularly strong for relatively weak factors and unusually large factor loadings (loadings on the first and second cross-sectional units in our experiment).

In our third experiment, we simulate 1000 replications of data having 1-factor structure with $n = 40$, $T = 20$, where F_{t1} are i.i.d. $N(0, 1)$, $\sigma^2 = 1$, $L_{i1} = \sqrt{d/n}$, and d is on a grid 0.1:0.1:20. Figure 4 shows the Monte Carlo and theoretical means and 5% and 95% quantiles of the first eigenvalue of XX'/T as functions of d. Smooth solid lines correspond to the theoretical lines obtained using formulae in Corollary 2. Rough solid lines correspond to the Monte Carlo sample data. Dotted lines are classical theoretical lines (fixed n large T asymptotics). Remarkably, our asymptotic formula for the mean traces the actual finite sample mean very well for all d on the grid. The 5% and 95% asymptotic quantiles also work well. Clearly, our asymptotic distribution provides a much better approximation to the finite sample distribution

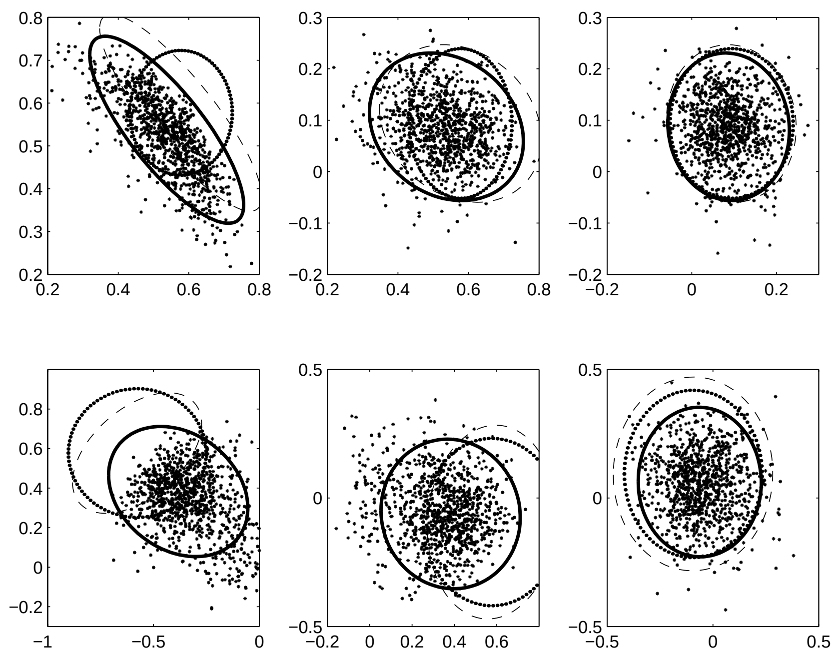

Figure 3: Monte Carlo draws and 95% asymptotic confidence ellipsoids for (from left to right) (1st, 2nd), (2nd, 3rd), (3rd, 4th) components of the normalized vectors of factor loadings. Upper panel: loadings correspond to the first factor. Lower panel: loadings correspond to the second factor. Solid line: our asymptotics. Dashed line: classical asymptotics. Dotted line: "strong factor" asymptotics.

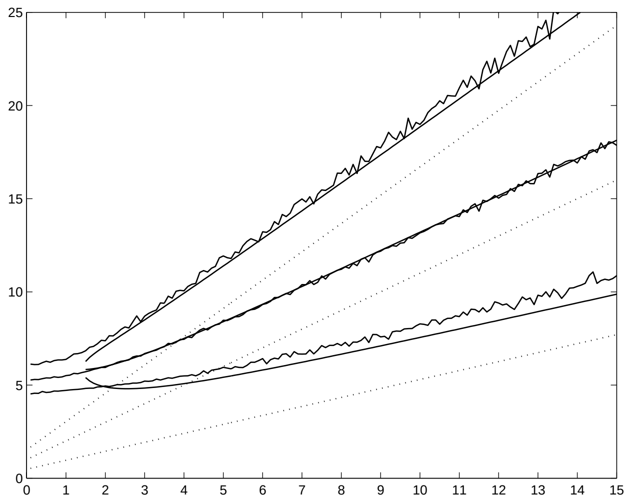

Figure 4: Monte Carlo and asymptotic means and 5% and 95% quantiles of the eigenvalue distribution. Smooth solid lines: our asymptotics. Dotted lines: classical asymptotics. Horizontal axis: the cumulative effect d of the factor. $n = 40$, $T = 20$.

than the classical distribution.

6 The proofs

In this section, we first prove Theorem 5 and then prove Theorem 2, Theorem 4, and Proposition 1, in that order. (The proofs of Theorems 1 and 3 are completely analogous to those of Theorems 2 and 4 and we omit them to save space.)

6.1 Proof of Theorem 5

Let O_L and O_F be $n \times n$ and $T \times T$ orthogonal matrices such that the first k columns of O_L are equal to the columns of $L\,(L'L)^{-1/2}$ and the first k columns of O_F are equal to the columns of $F(F'F)^{-1/2}$. Define $\tilde{\varepsilon} = O_L'\varepsilon O_F$ and let $\frac{1}{T}\tilde{\varepsilon}_{k+1:T}\tilde{\varepsilon}_{k+1:T}' = O'\Lambda O$ be the spectral decomposition of $\frac{1}{T}\tilde{\varepsilon}_{k+1:T}\tilde{\varepsilon}_{k+1:T}'$, where $\tilde{\varepsilon}_{k+1:T}$ denotes a matrix that consists of the last $T-k$ columns of $\tilde{\varepsilon}$. Note that, since $\tilde{\varepsilon}_{k+1:T}\tilde{\varepsilon}_{k+1:T}'$ is distributed according to Wishart $W\left(\sigma^2 I_n, T-k\right)$, its spectral decomposition can be chosen so that O has the Haar invariant distribution (see Anderson (1984)).[2] Define $\hat{X} = OO_L'XO_F$

[2] The decomposition is not unique because each of the columns of O can be multiplied by -1 and the last $\max(0, n-T+k)$ columns can be arbitrarily rotated.

and $\Psi = O_{1:k}(L'L)^{1/2}\left(\frac{F'F}{T}\right)^{1/2} + \frac{1}{\sqrt{T}}O\tilde{\varepsilon}_{1:k}$. Then, matrix $\frac{1}{T}\hat{X}\hat{X}'$ has a convenient representation $\frac{1}{T}\hat{X}\hat{X}' = \Psi\Psi' + \Lambda$ and the same eigenvalues as matrix $\frac{1}{T}XX'$.

Let $\mu_i(A)$ denote the i-th largest eigenvalue of a symmetric matrix A, y_{ij} denote the i-th component of an eigenvector of $\frac{1}{T}\hat{X}\hat{X}'$, corresponding to eigenvalue $\mu_j\left(\frac{1}{T}XX'\right)$, and λ_i denote the i-th largest diagonal element of Λ. Then, if $\mu_j\left(\frac{1}{T}XX'\right) \neq \lambda_i$ for any $i = 1, ..., n$, we have $y_{ij} = \frac{1}{\mu_j - \lambda_i}\Psi_{i\cdot}\Psi' y_{\cdot j}$. Multiplying this equality by $\Psi'_{i\cdot}$ and summing over all i, we get $\Psi' y_{\cdot j} = M_n^{(1)}\left(\mu_j\right)\Psi' y_{\cdot j}$, where $M_n^{(1)}(x) \equiv \sum_{i=1}^{n} \frac{\Psi'_{i\cdot}\Psi_{i\cdot}}{x-\lambda_i}$. Note that $M_n^{(1)}\left(\mu_j\right)$ must have an eigenvalue equal to 1. In fact, we can prove a stronger result:

Lemma 1: *Let $\mu \neq \lambda_i$, $i = 1, ..., n$. Then, μ is an eigenvalue of $\frac{1}{T}XX'$ if and only if there exists $m \leq k$ such that $x = \mu$ satisfies equation*

$$\mu_m\left(M_n^{(1)}(x)\right) = 1. \tag{7}$$

A proof of this lemma as well as all other auxiliary propositions stated in this section can be found in the Appendix. We plan to study the asymptotic behavior of $M_n^{(1)}(x)$ and its eigenvalues considered as random functions of x and to deduce from it the asymptotic properties of solutions to (7), which by Lemma 1 are the eigenvalues of $\frac{1}{T}XX'$.

The key fact for the analysis below was established by Marčenko and Pastur (1967). They showed that the empirical distribution of the elements along the diagonal of Λ defined as $\mathcal{F}^{\Lambda} \equiv \frac{\#\{\lambda_i \leq \lambda\}}{n}$ almost surely converges to a non-random cumulative distribution function $\mathcal{F}_c$, which has density

$$\begin{aligned} f_c(\lambda) &= \begin{cases} \frac{1}{2\pi\lambda c\sigma^2}\sqrt{(b-\lambda)(\lambda-a)} & \text{if } a \leq \lambda \leq b \\ 0 & \text{otherwise} \end{cases} \\ a &= \left(1-\sqrt{c}\right)^2\sigma^2, \quad b = \left(1+\sqrt{c}\right)^2\sigma^2, \end{aligned} \tag{8}$$

and a point mass $1 - 1/c$ at $\lambda = 0$ if $c > 1$.

To see the significance of the Marčenko-Pastur result for our analysis, assume for a moment that $k = 1$ and note that $M_n^{(1)}(x)$ is a weighted linear combination of terms Ψ_i^2 with weights $(x - \lambda_i)^{-1}$. Now, by definition, $\Psi_i = O_{i,1}(L'L)^{1/2}\left(\frac{F'F}{T}\right)^{1/2} + \frac{1}{\sqrt{T}}O_{i\cdot}\tilde{\varepsilon}_1$. The second element in this sum is independent of the first and, by Assumption 2, is $N(0, \sigma^2/T)$. The first term is asymptotically $N(0, d_1/n)$. Indeed, since O has the

Haar invariant distribution, the joint distribution of the entries of its first column is the same as that of the entries of $\xi/\|\xi\|$, where $\xi \sim N(0, I_n)$ and $\|\xi\| = \sqrt{\xi'\xi}$. Hence, $M_n^{(1)}(x)$ asymptotically behaves as a weighted sum of $\chi^2(1)$ independent random variables with weights $\frac{1}{n}(d_1 + c\sigma^2)(x-\lambda_i)^{-1}$. Intuitively, such a sum should converge to $(d_1 + c\sigma^2)\int (x-\lambda)^{-1} d\mathcal{F}_c(\lambda)$, which we confirm below. The properties of $M_n^{(1)}(x)$ centered by its probability limit and scaled by $\sqrt{n}$ can be analyzed using similar ideas.

Now, let us formally establish the asymptotic behavior of $M_n^{(1)}(x)$. As was shown by Bai, Silverstein and Yin (1988), for any fixed k, $\lambda_1, ..., \lambda_k$ almost surely converge to b. This result implies that, with high probability, $M_n^{(1)}(x)$ belongs to the space $C[\theta_1, \theta_2]^{k^2}$ of continuous $k \times k$-matrix-valued functions on $x \in [\theta_1, \theta_2]$, where $\theta_2 > \theta_1 > b$. Since the weak convergence in $C[\theta_1, \theta_2]$ is well-studied, it will be convenient to modify $M_n^{(1)}(x)$ on a small probability set so that the modification is a random element of $C[\theta_1, \theta_2]^{k^2}$ equipped with the max sup norm. To construct such a modification, define $h(x, \lambda_i) = \max\left(x - \lambda_i, \frac{\theta_1 - b}{2}\right)$ and let $\hat{M}_n^{(1)}(x) \equiv \sum_{i=1}^n \frac{\Psi_{i\cdot}'\Psi_{i\cdot}}{h(x,\lambda_i)}$. We will study the asymptotic properties of $\hat{M}_n^{(1)}(x)$ keeping in mind that they are equivalent to the asymptotic properties of $M_n^{(1)}(x)$ because $P\left(M_n^{(1)}(x) = \hat{M}_n^{(1)}(x), \forall x \in [\theta_1, \theta_2]\right) = P\left(\lambda_1 < \frac{\theta_1 + b}{2}\right) \to 1$ as $n \to \infty$.

To prove Theorems 2 and 4 we will also need to analyze the asymptotic properties of $M_n^{(2)}(x) \equiv \sum_{i=1}^n \frac{\Psi_{i\cdot}'\Psi_{i\cdot}}{(x-\lambda_i)^2}$ and $M_n^{(3)}(x) \equiv \sum_{i=1}^n \frac{O_{i,1:k}'\Psi_{i\cdot}}{x-\lambda_i}$. Define $\hat{M}_n^{(2)}(x) = \sum_{i=1}^n \frac{\Psi_{i\cdot}'\Psi_{i\cdot}}{h^2(x,\lambda_i)}$, $\hat{M}_n^{(3)}(x) = \sum_{i=1}^n \frac{O_{i,1:k}'\Psi_{i\cdot}}{h(x,\lambda_i)}$, $M_0^{(1)}(x) = (D + \sigma^2 c I_k)\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}$, $M_0^{(2)}(x) = (D + \sigma^2 c I_k)\int \frac{d\mathcal{F}_c(\lambda)}{(x-\lambda)^2}$, and $M_0^{(3)}(x) = D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}$. Appendix proves the following

Lemma 2: *Let Assumptions 1 (or 1'), 2, and 3 hold. Then, for the random elements of $C^{k^2}[\theta_1, \theta_2]$ defined as $N_n^{(p)}(x) = \sqrt{n}\left(\hat{M}_n^{(p)}(x) - M_0^{(p)}(x)\right), p = 1, 2, 3,$ we have:*

$$\left\{N_n^{(p)}(x), p = 1, 2, 3\right\} \xrightarrow{d} \left\{N^{(p)}(x), p = 1, 2, 3\right\}, \tag{9}$$

where, for any $\{x_1, ..., x_J\} \in [\theta_1, \theta_2]$, the joint distribution of entries of $\left\{N^{(p)}(x_j); p = 1, 2, 3, j = 1, ..., J\right\}$ is a $3Jk^2$-dimensional normal distribution with covariance between entry in row s and column t of $N^{(p)}(x_j)$ and entry in row s_1 and column t_1 of $N^{(r)}(x_{j_1})$ equal to $\Omega^{(p,r)}(\tau, \tau_1)$, where $\tau = (s, t, j)$ and $\tau_1 = (s_1, t_1, j_1)$, and $\Omega^{(p,r)}(\tau, \tau_1)$ is defined in the Appendix.

Using Lemma 2, it is easy to establish the probability limits of the first k eigenvalues of XX'/T. Recall that by Lemma 1, we should look at the probability limits of the solutions to $\mu_j\left(M_n^{(1)}(x)\right)=1$. Consider, first, solutions to a related equation $\mu_j\left(M_0^{(1)}(x)\right)=1$. Function $\mu_j\left(M_0^{(1)}(x)\right)=(d_j+\sigma^2 c)\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}$ is continuous and strictly decreasing on $(b,+\infty)$, and tends to zero as $x\to+\infty$. In addition, since, as is straightforward to check, $\lim_{x\downarrow b}\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}=\frac{1}{c\sigma^2}\frac{\sqrt{c}}{1+\sqrt{c}}$, we have: $\lim_{x\downarrow b}\mu_m(M_0^{(1)}(x))>1$ if and only if $d_j>\sqrt{c}\sigma^2$. Therefore, there exist unique solutions $x_{0j}\in(b,+\infty)$ to equations $\mu_j\left(M_0^{(1)}(x)\right)=1$ for $j\leq q$, and there are no solutions to the equations on $(b,+\infty)$ for $q<j\leq k$.

Now, fix θ_1 and θ_2 so that $\theta_2>\theta_1>b$; $\{x_{0j}:j\leq q\}\in(\theta_1,\theta_2)$, and $(d_k+\sigma^2 c)\int\frac{d\mathcal{F}_c(\lambda)}{\theta_2-\lambda}<\frac{1}{2}$. The continuous mapping theorem and Lemma 2 imply that $\mu_j\left(\hat{M}_n^{(1)}(x)\right)\xrightarrow{d}\mu_j\left(M_0^{(1)}(x)\right)$, in the sense of the weak convergence of the random elements of $C[\theta_1,\theta_2]$. Using this convergence and the monotonicity of $\mu_j\left(\hat{M}_n^{(1)}(x)\right)$ it is easy to show that with probability arbitrarily close to 1, there are no solutions to $\mu_j\left(\hat{M}_n^{(1)}(x)\right)=1$ larger than θ_1 for $q<j\leq k$ and large enough n. Therefore, $P\left\{\mu_j(\frac{1}{T}XX')<\theta_1,q<j\leq k\right\}\to 1$ as $n\to\infty$. But, since $\frac{1}{T}\hat{X}\hat{X}'-\Lambda$ is a positive semi-definite matrix, $\mu_j(\frac{1}{T}XX')$, $q<j\leq k$ cannot be smaller than λ_k which tends almost surely to b. Since θ_1 can be chosen arbitrarily close to b, we have $\mu_j(\frac{1}{T}XX')\xrightarrow{p}b$ for $q<j\leq k$ which proves statement ii of Theorem 5.

In contrast, with high probability there exist unique solutions $x_{nj}\in[\theta_1,\theta_2]$ to $\mu_j\left(\hat{M}_n^{(1)}(x)\right)=1$ for $j\leq q$, and $x_{nj}\xrightarrow{p}x_{0j}$.[3] Therefore, $\mu_j(\frac{1}{T}XX')\xrightarrow{p}x_{0j}$ for $j\leq q$. A short technical derivation relegated to the Appendix shows that $x_{0j}=\frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2 c\right)}{d_i}$ which is denoted as m_j in the condition of Theorem 5.

Next, we show that, for any $j\leq q$,

$$\mu_j(\hat{M}_n^{(1)}(x))=\mu_j\left(M_0^{(1)}(x)\right)+\frac{1}{\sqrt{n}}N_{n,jj}^{(1)}(x)+o_p\left(\frac{1}{\sqrt{n}}\right), \tag{10}$$

where $o_p\left(\frac{1}{\sqrt{n}}\right)$ is understood as a random element of $C[\theta_1,\theta_2]$, which, when multiplied by $\sqrt{n}$, tends in probability to zero as $n\to\infty$. Formula (10) is an easy consequence of Lemma 2 and part i of the following lemma.[4]

Lemma 3: *Let $A(\varkappa)=A+\varkappa A^{(1)}$, where $A^{(1)}$ is a symmetric $k\times k$ ma-*

[3] When there is no solution to $\mu_j\left(\hat{M}_n^{(1)}(x)\right)=1$ on $[\theta_1,\theta_2]$, we can define $x_{nj}\in[\theta_1,\theta_2]$ arbitrarily.

[4] We will need part ii of the lemma to prove Theorem 2.

trix and $A = \operatorname{diag}(a_1, a_2, ..., a_k)$, $a_1 > a_2 > ... > a_k > 0$. *Further, let* $r_0 = \frac{1}{2}\min_{j=1,...,k}|a_j - a_{j+1}|$, *where we define* a_{k+1} *as zero. Then, for any real* $\varkappa$ *such that* $|\varkappa| < r_0/\left\|A^{(1)}\right\|$, *the following two statements hold:*
i) Exactly one eigenvalue of $A(\varkappa)$ *belongs to the segment* $(a_j - r_0, a_j + r_0)$. *Denoting this eigenvalue as* $a_j(\varkappa)$, *we have:*[5] $\left|\frac{1}{\varkappa}(a_j(\varkappa) - a_j) - A_{jj}^{(1)}\right| \leq |\varkappa| \left\|A^{(1)}\right\| \left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^{-1}$.
ii) Let $P_j(\varkappa)$ *be the orthogonal projection on the invariant subspace of* $A(\varkappa)$ *corresponding to eigenvalue* $a_j(\varkappa)$ *and let*
$S_j = \operatorname{diag}\left((a_1 - a_j)^{-1}, ..., (a_{j-1} - a_j)^{-1}, 0, (a_{j+1} - a_j)^{-1}, ..., (a_k - a_j)^{-1}\right)$. *Then* $e_j(\varkappa) \equiv P_j(\varkappa)e_j/\|P_j(\varkappa)e_j\|$ *is an eigenvector of* $A(\varkappa)$ *corresponding to eigenvalue* $a_j(\varkappa)$, *and* $\left\|\frac{1}{\varkappa}(e_j(\varkappa) - e_j) + S_j A^{(1)} e_j\right\| \leq 2|\varkappa| \left\|A^{(1)}\right\|^2 \left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^{-2}$.

Define function $\nu_j(y)$ for $y > 0$ so that it is equal to b if $y > \lim_{x \downarrow b} \mu_j(M_0^{(1)}(x))$ and to the inverse function to function $\mu_j(M_0^{(1)}(x))$ otherwise. Since $\frac{d}{dx}\mu_j(M_0^{(1)}(x)) = -(d_j + \sigma^2 c)\int \frac{d\mathcal{F}_c(\lambda)}{(x-\lambda)^2}$, it is easy to see that $\lim_{x\downarrow b} \frac{d}{dx}\mu_j(M_0(x)) = +\infty$, and, hence, $\nu_j(y)$ is differentiable for $y > 0$. Applying ν_j to both sides of (10) and using the first order Taylor expansion of the right hand side, we have for $x \in [\theta_1, \theta_2]$: $\nu_j\left(\mu_j(\hat{M}_n^{(1)}(x))\right) = x + \nu_j'(\tau_n(x)) \frac{1}{\sqrt{n}} N_{n,jj}^{(1)}(x) + o_p\left(\frac{1}{\sqrt{n}}\right)$, where $\tau_n(x)$ is a random element of $C[\theta_1, \theta_2]$ such that $\tau_n(x) \xrightarrow{p} \mu_j(M_0^{(1)}(x))$ as $n \to \infty$.

Note that by definition of x_{nj}, definition of $\nu_j(\cdot)$, and Lemma 1, $\mu_j\left(\hat{M}_n^{(1)}(x_{nj})\right) = 1$, $\nu_j\left(\mu_j\left(\hat{M}_n^{(1)}(x_{nj})\right)\right) = x_{0j}$, and $x_{nj} = \mu_j\left(\frac{1}{T}XX'\right)$ with probability arbitrarily close to 1 for large enough n. Substituting x by x_{nj} in the above expansion of $\nu_j\left(\mu_j(\hat{M}_n^{(1)}(x))\right)$ and using these facts, we obtain: $\sqrt{n}\left(\mu_j\left(\frac{1}{T}XX'\right) - x_{0j}\right) = -\nu_j'(\tau_n(x_{nj})) N_{n,jj}^{(1)}(x_{nm}) + o_p(1)$.

Further, since $x_{nj} \xrightarrow{p} x_{0j} = m_j$, we have $\nu_j'(\tau_n(x_{nj})) \xrightarrow{p} \nu_j'(1)$. Finally, $N_{n,jj}^{(1)}(x_{nj}) - N_{n,jj}^{(1)}(m_j) \xrightarrow{p} 0$, which follows from Lemma 2 and the following additional

Lemma 4: *Let* $f_n(x)$ *and* $f_0(x)$ *be random elements of* $C[\theta_1, \theta_2]$ *such that* $f_n(x) \xrightarrow{d} f_0(x)$ *as* $n \to \infty$. *And let* x_n *be random variables with values form* $[\theta_1, \theta_2]$ *and such that* $x_n \xrightarrow{p} x_0$, *where* $x_0 \in [\theta_1, \theta_2]$ *Then* $f_n(x_n) - f_n(x_0) \xrightarrow{p} 0$.

Therefore, $\sqrt{n}\left(\mu_j\left(\frac{1}{T}XX'\right) - x_{0j}\right)$ has the following form

$$\sqrt{n}\left(\mu_j\left(\frac{1}{T}XX'\right) - m_j\right) = -\nu_j'(1) N_{n,jj}^{(1)}(m_j) + o_p(1). \tag{11}$$

[5] For any matrix (or vector) B, $\|B\| = (\max eig(B^*B))^{1/2}$, where $*$ denotes the operation of transposition and complex conjugation.

the Appendix shows that $-\nu_j'(1) = \left(d_j^2 - \sigma^4 c\right)\left(d_j + \sigma^2 c\right) d_j^{-2}$. The latter equality, formula (11), and Lemma 2 imply statement i of Theorem 5.□

6.2 Proof of Theorem 2

First, note that representation $\hat{\mathcal{L}}_{1:q}=\mathcal{L}\cdot R+\mathcal{L}_q^{\perp}$, where $\mathcal{L}_q^{\perp}$ is a matrix with q columns orthogonal to span $(\mathcal{L})$ is a trivial coordinate decomposition statement. The value of Theorem 2 is, therefore, contained in describing properties of $\mathcal{L}_q^{\perp}$ and R. Recall that the columns of $\hat{\mathcal{L}}_{1:q}$ are equal to the q principal eigenvectors of $\frac{1}{T}XX'$. By Assumption 2, the joint distribution of elements of X is invariant with respect to multiplication of X from the left by any orthogonal matrix leaving columns of L unchanged. This immediately implies that the joint distribution of entries of $\mathcal{L}_q^{\perp}$ is invariant with respect to the multiplication of $\mathcal{L}_q^{\perp}$ from the left by any orthogonal matrix that has span $(\mathcal{L})$ = span (L) as its invariant subspace. In the rest of the proof we, therefore, focus on the properties of R.

Since $\hat{\mathcal{L}}_{\cdot j}$ is an eigenvector of $\frac{1}{T}XX'$ corresponding to $\mu_j\left(\frac{1}{T}XX'\right)$, we have $\hat{\mathcal{L}}_{\cdot j} = O_L O' y_{\cdot j}$, where $y_{\cdot j}$ is an eigenvector of $\frac{1}{T}\hat{X}\hat{X}'$ corresponding to the same eigenvalue. This implies that the vector of coordinates of $\hat{\mathcal{L}}_{\cdot j}$ in the basis formed by columns of O_L is equal to $O' y_{\cdot j}$. Further, since the first k columns of O_L form matrix $\mathcal{L}$, $R_{\cdot j}$ must be equal to the vector of the first k coordinates $O_{1:k}' y_{\cdot j}$. Using this fact, the fact established in the proof of Theorem 5, that $y_{ij} = \left(\mu_j\left(\frac{1}{T}XX'\right) - \lambda_i\right)^{-1}\Psi_{i\cdot}\Psi' y_{\cdot j}$, and the definitions of Ψ, $\hat{M}_n^{(3)}(x)$, and x_{nj}, it is straightforward to check that

$$R_{\cdot j} = \hat{M}_n^{(3)}(x_{ni})\,\Psi' y_{\cdot j} \tag{12}$$

with probability arbitrarily close to one for large enough n. The analysis below will be based on this representation of $R_{\cdot j}$.

We, first, find the probability limit $R^{(1)}$ of R. Lemma 2, Lemma 4, and the fact that $x_{nj} \xrightarrow{p} m_j$ imply that $\hat{M}_n^{(3)}(x_{nj}) \xrightarrow{p} M_0^{(3)}(m_j) \equiv D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{m_j - \lambda}$. Further, since $\Psi' y_{\cdot j} = M_n^{(1)}\left(\mu_j\left(\frac{1}{T}XX'\right)\right)\Psi' y_{\cdot j}$, $w_{nj} \equiv \Psi' y_{\cdot j}/\left\|\Psi' y_{\cdot j}\right\|$ is a unit-length eigenvector of $\hat{M}_n^{(1)}(x_{nj})$ with high probability for large enough n. By part ii of Lemma 3, $w_{nj} \xrightarrow{p} e_j$. Finally, since $y_{ij} = \left(\mu_j\left(\frac{1}{T}XX'\right) - \lambda_i\right)^{-1}\Psi_{i\cdot}\Psi' y_{\cdot j}$ and $\|y_{\cdot j}\| = 1$, $\|\Psi' y_{\cdot j}\| = \left(w_{nj}'\hat{M}_n^{(2)}(x_{nj})\, w_{nj}\right)^{-1/2}$ with high probability for large enough n. But by Lemma 2, and Lemma 4, $\hat{M}_n^{(2)}(x_{nj}) \xrightarrow{p} (D + \sigma^2 c I_k)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j - \lambda)^2}$. Therefore,

$\|\Psi' y_{\cdot i}\| \xrightarrow{p} \left((d_j+\sigma^2 c)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2}\right)^{-1/2}$. Using representation (12) and these convergence results, we get: $R_{\cdot j} \xrightarrow{p} d_j^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{m_j-\lambda}\left((d_j+\sigma^2 c)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2}\right)^{-1/2} e_j$. Formulae (24) and (25) from the Appendix imply that this limit simplifies so that we get: $R_{\cdot j} \xrightarrow{p} \left(\frac{d_j^2-\sigma^4 c}{d_j(d_j+\sigma^2 c)}\right)^{1/2} e_j$ which establishes the form of $R^{(1)}$.

Now, we will study the asymptotic behavior of R around its probability limit $R^{(1)}$. Starting from representation (12) the Appendix shows that the asymptotic joint distribution of the components of q $k\times 1$ vectors $\sqrt{n}\left(R_{\cdot j}-R_{\cdot j}^{(1)}\right)$, $j=1,...,q$ is the same as that of the components of q $k\times 1$ vectors $\sum_{s=1}^{4}\varkappa_j\tilde{A}_j^{(s)}$, $j=1,...,q$, where $\tilde{A}_j^{(1)} = N_n^{(3)}(m_j)e_j$, $\tilde{A}_j^{(2)} = -0.5\left(d_j^2-c\sigma^4\right)d_j^{-3/2}N_{n,jj}^{(2)}(m_j)e_j$, $\tilde{A}_j^{(3)} = \sigma^4 c\left(d_i^2-c\sigma^4\right)^{-1}d_i^{-1/2}N_{n,jj}^{(1)}(m_j)\,e_j$, $\tilde{A}^{(4)} = -D^{1/2}S_jN_n^{(1)}(m_j)\,e_j$, and $\varkappa_j = \left(d_j^2-\sigma^4 c\right)^{1/2}\left(d_j+\sigma^2 c\right)^{1/2}d_j^{-1}$. Using Lemma 2, we conclude that the joint asymptotic distribution of the elements of $\sqrt{n}\left(R-R^{(1)}\right)$ is Gaussian. The elements of the covariance matrix of the asymptotic distribution of $\sqrt{n}\left(R-R^{(1)}\right)$ can be found[6] using the above definitions of $\tilde{A}_j^{(s)}$, $s=1,...,4$, the expressions for the covariance of $N_n^{(1)}(m_j)$, $N_n^{(2)}(m_j)$, and $N_n^{(3)}(m_j)$, $j=1,...,q$ summarized in the definition of $\Omega^{(\cdot,\cdot)}$ given in the Appendix, and formulae (24),(25),(27), and (28).

Let us now complete the proof by considering the case when $d_j \le \sqrt{c}\sigma^2$. Consider a number $\gamma > b$. For large enough n, with high probability $\mu_j \equiv \mu_j\left(\frac{1}{T}XX'\right) < \gamma$ because, by Theorem 5, $\mu_j \xrightarrow{p} b$. Therefore, for large enough n with high probability $\min\operatorname{eval}\left(\sum_{i=k+1}^n \Psi_{i\cdot}'\left(\mu_j-\lambda_i\right)^{-2}\Psi_{i\cdot}\right) \ge \min\operatorname{eval}\left(\sum_{i=k+1}^n \Psi_{i\cdot}'\left(\gamma-\lambda_i\right)^{-2}\Psi_{i\cdot}\right)$. Since k is fixed, the right hand side of the latter inequality is asymptotically equivalent to $\min\operatorname{eval} M_n^{(2)}(\gamma)$ which, by Lemma 2, converges to $\min\operatorname{eval}\left(D+\sigma^2 cI_k\right)\int \frac{\mathcal{F}_c(d\lambda)}{(\gamma-\lambda)^2}$. Note that the latter expression can be made arbitrarily large by choosing γ close enough to b. Therefore, $\min\operatorname{eval}\left(\sum_{i=k+1}^n \Psi_{i\cdot}'\left(\mu_j-\lambda_i\right)^{-2}\Psi_{i\cdot}\right) \xrightarrow{p} \infty$. But, since $y_{ij} = \left(\mu_j-\lambda_i\right)^{-1}\Psi_{i\cdot}\Psi' y_{\cdot j}$ and $\|y_{\cdot j}\| = 1$, $(\Psi' y_{\cdot j})'\left(\sum_{i=k+1}^n \Psi_{i\cdot}'\left(\mu_j-\lambda_i\right)^{-2}\Psi_{i\cdot}\right)(\Psi' y_{\cdot j}) = \sum_{i=k+1}^n y_{ij}^2 \le 1$, and hence, $\Psi' y_{\cdot j} \xrightarrow{p} 0$.

Now, let τ be a number $0 < \tau < 1$. We have: $\sum_{i=[\tau n]+1}^n y_{ij}^2 = (\Psi' y_{\cdot j})'\left(\sum_{i=[\tau n]+1}^n \Psi_{i\cdot}'\left(\mu_j-\lambda_i\right)^{-2}\Psi_{i\cdot}\right)(\Psi' y_{\cdot j})$. Note that, for $i \ge [\tau n]+1$, $\left(\mu_j-\lambda_i\right)^{-2} \ge \left(\mu_j-\lambda_{[\tau n]+1}\right)^{-2}$. By Marčenko and Pastur (1967) result and Theorem 5 the right

[6] To obtain these formulas we used symbolic manipulation software of the Scientific Workplace, version 5.

hand side of the latter inequality converges to $\left(b-\mathcal{F}_c^{-1}(1-\tau)\right)^{-2}<\infty$. Therefore, $\left\|\sum_{i=[\tau n]+1}^{n}\Psi_{i\cdot}'\left(\mu_j-\lambda_i\right)^{-2}\Psi_{i\cdot}\right\|$ is bounded in probability and, since $\Psi' y_{\cdot j}\xrightarrow{p}0$, $\sum_{i=[\tau n]+1}^{n}y_{ij}^2\xrightarrow{p}0$. Loosely speaking, for any $0<\tau<1$, with high probability for large enough n, almost all "mass" in vector $y_{\cdot j}$ is concentrated in the first $\tau 100\%$ of its components.

Finally, the i-th coordinate of $\hat{\mathcal{L}}_{\cdot j}$ in the basis O_L are equal to $\left(O_{\cdot i}\right)' y_{\cdot j}$. We have $\left|\left(O_{\cdot i}\right)' y_{\cdot j}\right| = \sum_{s=1}^{[\tau n]}O_{si}y_{sj}+\sum_{s=[\tau n]+1}^{n}O_{si}y_{sj}\le\left(\sum_{s=1}^{[\tau n]}O_{si}^2\right)^{1/2}+\left(\sum_{s=[\tau n]+1}^{n}y_{sj}^2\right)^{1/2}$. The last term in the right hand side of the above inequality converges in probability to zero. As to the first term, since O is Haar distributed, $\sum_{s=1}^{[\tau n]}O_{si}^2$ has the same distribution as $\frac{1}{\|\varsigma\|^2}\sum_{j=1}^{[\tau n]}\varsigma_j^2$, where ς is an $n\times 1$ standard normal vector. Clearly, $\frac{1}{\|\varsigma\|^2}\sum_{j=1}^{[\tau n]}\varsigma_j^2\xrightarrow{p}\tau$. Therefore, $\Pr\left(\left|\left(O_{\cdot i}\right)' y_{\cdot j}\right|>2\tau\right)\to 0$ as $n\to\infty$ for any $0<\tau<1$. In other words, all coordinates of $\hat{\mathcal{L}}_{\cdot j}$ in basis O_L converge in probability to zero.□

6.3 Proof of Theorem 4

First, note that since the distribution of the data X does not depend on the multiplication of X from the left by any orthogonal matrix having $\operatorname{span}(L)$ as its invariant subspace, the joint distribution of the coordinates of the columns of $\hat{\mathcal{L}}$ in the basis formed by the columns of O_L does not depend on how the $k+1$-th, $k+2$-th, ..., n-th columns of O_L are chosen.

Denote an $n\times 1$ unit-length vector with all entries but the j-th equal to zero as e_j. Let the $k+1$-th column of O_L be chosen as $M(L)e_{j_1}/\left\|M(L)e_{j_1}\right\|$, where $M(L)$ denotes the operator of taking the residual from the orthogonal projection on $\operatorname{span}(L)$, the $k+2$-th column be chosen as $M([L,e_{j_1}])e_{j_2}/\left\|M([L,e_{j_1}])e_{j_2}\right\|$, ..., and the $k+r$-th column be chosen as $M(\left[L,e_{j_1},...,e_{j_{r-1}}\right])e_{j_r}/\left\|M(\left[L,e_{j_1},...,e_{j_{r-1}}\right])e_{j_r}\right\|$. For example, if $r=2$ and $j_1=1$ and $j_2=2$, then matrix O_L has the following structure

$$O_L=\left[\begin{array}{c|ccccc} & x & 0 & 0 & \cdots & 0\\ LD^{-\frac{1}{2}} & y & z & 0 & \cdots & 0\\ \cline{2-6} & \multicolumn{5}{c}{*}\end{array}\right],\tag{13}$$

where $x = \|M(L)e_1\|$, $y = e_2'M(L)e_1/\|M(L)e_1\|$, and $z = \|M([L, e_1])e_2\|$. Note that:

$$x^2 = e_{j_1}'M(L)e_{j_1} = 1 - e_{j_1}'L(L'L)^{-1}L'e_{j_1} = 1 - \sum_{i=1}^{k}\mathcal{L}_{j_1 i}^2 \tag{14}$$

$$y = \frac{1}{x}e_{j_2}'M(L)e_{j_1} = -\frac{1}{x}\sum_{i=1}^{k}\mathcal{L}_{j_1 i}\mathcal{L}_{j_2 i}. \tag{15}$$

Let us denote the $n-k$ coordinates of the columns of $\hat{\mathcal{L}}_{1:q}$ in the basis formed by the columns of O_L as $R^{\perp}$. That is, $R_{ij}^{\perp}$ is the scalar product of $\hat{\mathcal{L}}_{\cdot j}$ and the $k+i$-th column of O_L. Then, $\hat{\mathcal{L}}_{j_s i} = \mathcal{L}_{j_s\cdot}\cdot R_{\cdot i} + \sum_{t=1}^{r} O_{L,j_s t}\cdot R_{ti}^{\perp}$. Hence, we can obtain the asymptotic joint distribution of $\left\{\hat{\mathcal{L}}_{j_s i}; s = 1, ..., r; i = 1, ..., q\right\}$ from the asymptotic joint distribution of the entries of R and the first r columns of $R^{\perp}$.

It is easy to see that matrix $\tilde{R}^{\perp} \equiv R^{\perp}(I_q - R'R)^{-1/2}$, where R is as defined in Theorem 2, has orthonormal columns. Moreover, as a consequence of the invariance of the distribution of X with respect to the orthogonal transformations leaving L unchanged, the joint distribution of the entries of $\tilde{R}^{\perp}$ conditional on R is invariant with respect to multiplication of $\tilde{R}^{\perp}$ from the left by any orthogonal matrix. This implies that the joint distribution of the entries of $\tilde{R}^{\perp}\alpha$ conditional on R, where α is any $q\times 1$ unit-length vector, is the same as the joint distribution of the entries of $\xi/\|\xi\|$, where ξ is an $(n-k)\times 1$ vector with i.i.d. Gaussian entries.

As a consequence of the above result, the entries of $\tilde{R}^{\perp}\alpha$ are independent from the entries of $R\alpha$, and their unconditional joint distribution is the same as that of the entries of $\xi/\|\xi\|$. This fact, together with Theorem 2 and Cramer-Wold theorem (see White (1999), p.114), implies that the entries of $\sqrt{n}\left(R - R^{(1)}\right)$ and of the first r rows of $\sqrt{n}R^{\perp}$, where r is any fixed positive number, are asymptotically independent and have asymptotic joint zero-mean Gaussian distribution. The covariance matrix of the asymptotic distribution of the first r raws of $\sqrt{n}R^{\perp}$ is diagonal and Avar $\left(\sqrt{n}R_{ji}^{\perp}\right) = 1 - \left(R_{ii}^{(1)}\right)^2$.

The asymptotic joint Gaussianity of the entries of $\sqrt{n}\left(R - R^{(1)}\right)$ and $\sqrt{n}R^{\perp}$ implies that $\left\{\sqrt{n}\left(\hat{\mathcal{L}}_{j_g i} - R_{ii}^{(1)}\mathcal{L}_{j_g i}\right); g = 1, ..., r; i = 1, ..., q\right\}$ are asymptotically jointly mean-zero Gaussian. We will now find the variances and covariances of the asymptotic distribution. Consider the random variables $\sqrt{n}\left(\hat{\mathcal{L}}_{j_g i} - R_{ii}^{(1)}\mathcal{L}_{j_g i}\right)$ and $\sqrt{n}\left(\hat{\mathcal{L}}_{j_f p} - R_{pp}^{(1)}\mathcal{L}_{j_f p}\right)$. Without loss of generality assume that $g = 1$, $f = 2$. If $g \neq 1$ and/or $f \neq 2$, construct O_L so that its $k+1$-th column is $M(L)e_{j_g}/\left\|M(L)e_{j_g}\right\|$ and its $k+2$-th column is $M([L, e_{j_g}])e_{j_f} / \left\| M([L, e_{j_g}])e_{j_f} \right\|$. From (13), we have: $\sqrt{n}\left(\hat{\mathcal{L}}_{j_g i} - R^{(1)}_{ii}\mathcal{L}_{j_g i}\right) = \sum_{1 \le s \le k} \mathcal{L}_{j_g s}\sqrt{n}\left(R_{si} - R^{(1)}_{si}\right) + x\sqrt{n}R^{\perp}_{1i}$, and $\sqrt{n}\left(\hat{\mathcal{L}}_{j_f p} - R^{(1)}_{pp}\mathcal{L}_{j_f p}\right) = \sum_{1 \le s \le k} \mathcal{L}_{j_f s}\sqrt{n}\left(R_{sp} - R^{(1)}_{sp}\right) + y\sqrt{n}R^{\perp}_{1p} + z\sqrt{n}R^{\perp}_{2p}$. These two formulae together with (14), (15), and the formulae for the asymptotic covariance of entries of $\sqrt{n}\left(R - R^{(1)}\right)$ and of the first two rows of $\sqrt{n}R^{\perp}$ established above and in Theorem 2 imply the formula for the asymptotic covariance matrix claimed by Theorem 4.

Part ii of the theorem follows from part ii of Theorem 2 and the fact that the entries of the first row of $\sqrt{n}\tilde{R}^{\perp}$, where $\tilde{R}^{\perp}$ is defined similarly to $R^{\perp}$, converge in distribution. This fact can be established similarly to the analogous fact for $\sqrt{n}R^{\perp}$.$\square$

7 Conclusion

In this paper we have shown that the principal components estimators of factors and factor loadings are inconsistent but asymptotically normal as n and T approach infinity proportionally when the cumulative effects of the normalized factors on the cross-sectional units are assumed to be bounded, as opposed to increasing in n. We have found explicit formulae for the amount of the inconsistency and for the asymptotic covariance matrix of the estimators. Our Monte Carlo analysis suggests that our asymptotic formulae work well even for such small samples as $n = 40$, $T = 20$.

Our assumption that the cumulative effects of the factors are bounded contrasts the usual assumption of the unbounded effects made in the approximate factor models. This conflict should not preclude using our results in the empirical applications of such models. Our formulae simply provide an alternative asymptotic approximation to the finite sample distributions of interest to the applications. As we have shown, our asymptotic approximations converge to those proposed by Bai (2003) when the assumed bounds on the cumulative effects of the factors increase. Hence, in the applications where factors have very large cumulative effects in the sample investigated, our asymptotic approximation should work similarly to Bai's. On the other hand, when factors do not have large cumulative effects in the sample investigated, our results will provide a better approximation than results based on the assumption of strong asymptotic domination of factors over the idiosyncratic influences.

We obtained all our results under a strong assumption of i.i.d. noise. Such an assumption substantially reduces the generality of approximate factor models. Therefore, we see the main contribution of this paper in pointing out some situations in which the "strong factor" approximation of Bai (2003) may not perform well. Providing a general asymptotic theory for principal components estimation under weak factors is left for future research. In this paper we have not explored implications of our results for econometric practice. The implications of our results for diffusion index forecasting is the topic that I am currently working on.

8 Appendix

Definition of covariance function Ω from Lemma 1:

For $\tau = (s, t, j)$, $\tau_1 = (s_1, t_1, j_1)$, and integers p_1 and p_2 such that $1 \leq p_1 \leq p_2 \leq 2$, we define Ω as follows.

$$\begin{aligned}
\Omega^{(p_1,p_2)}(\tau, \tau_1) &= \frac{c}{4}(d_s + d_t)(d_{s_1} + d_{t_1})\,\phi_{sts_1t_1} \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)^{p_1}} \int \frac{d\mathcal{F}_c(\lambda)}{(x_{j_1} - \lambda)^{p_2}} \\
\Omega^{(p_1,3)}(\tau, \tau_1) &= \frac{c}{4}(d_s + d_t)\sqrt{d_{s_1}}\phi_{sts_1t_1} \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)^{p_1}} \int \frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda} \\
\Omega^{(3,3)}(\tau, \tau_1) &= \frac{c}{4}\sqrt{d_s d_{s_1}}\phi_{sts_1t_1} \int \frac{d\mathcal{F}_c(\lambda)}{x_j - \lambda} \int \frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda}
\end{aligned}$$

if $(s_1, t_1) \neq (s, t)$ and $(s_1, t_1) \neq (t, s)$;

$$\begin{aligned}
\Omega^{(p_1,p_2)}(\tau, \tau_1) &= \left[\frac{c}{4}(d_s + d_t)^2 \phi_{stst} - (1 + \delta_{st})\, d_s d_t\right] \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)^{p_1}} \int \frac{d\mathcal{F}_c(\lambda)}{(x_{j_1} - \lambda)^{p_2}} \\
&\quad + \left[(1 + \delta_{st})\left(\sigma^4 c^2 + d_s d_t\right) + \sigma^2 c\left(d_s + d_t + 2\delta_{st}\sqrt{d_s d_t}\right)\right] \\
&\quad \cdot \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)^{p_1}(x_{j_1} - \lambda)^{p_2}} \\
\Omega^{(p_1,3)}(\tau, \tau_1) &= \left[\frac{c}{4}(d_s + d_t)\sqrt{d_s}\phi_{stst} - (1 + \delta_{st})\sqrt{d_s} d_t\right] \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)^{p_1}} \int \frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda} \\
&\quad + \left[(1 + \delta_{st})\sqrt{d_s} d_t + \sigma^2 c\left(\sqrt{d_s} + \delta_{st}\sqrt{d_t}\right)\right] \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)^{p_1}(x_{j_1} - \lambda)} \\
\Omega^{(3,3)}(\tau, \tau_1) &= \left(\frac{c}{4} d_s \phi_{stst} - (1 + \delta_{st})\, d_t\right) \int \frac{d\mathcal{F}_c(\lambda)}{x_j - \lambda} \int \frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda} \\
&\quad + \left((1 + \delta_{st})\, d_t + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{(x_j - \lambda)(x_{j_1} - \lambda)}
\end{aligned}$$

if $(s_1, t_1) = (s, t)$; and

$$\begin{aligned}
\Omega^{(p_1,p_2)}(\tau,\tau_1) &= \Omega^{(p_1,p_2)}((t,s,j),(s_1,t_1,j_1)) \\
\Omega^{(p_1,3)}(\tau,\tau_1) &= \Omega^{(p_1,3)}((t,s,j),(s_1,t_1,j_1)) \\
\Omega^{(3,3)}(\tau,\tau_1) &= \left(\frac{c}{4}\phi_{stst} - (1+\delta_{st})\right)\sqrt{d_s d_t}\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda} \\
&\quad + \left((1+\delta_{st})\sqrt{d_s d_t} + \delta_{st}\sigma^2 c\right)\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)(x_{j_1}-\lambda)}
\end{aligned}$$

if $(s_1, t_1) = (t, s)$.

Proof of Lemma 1:

Suppose $x_0 \neq \lambda_i$, $i = 1, ..., n$ and x_0 satisfies (7). Let v be an eigenvector of $M_n^{(1)}(x_0)$ corresponding to the unit eigenvalue. Define $z_i = \frac{1}{x_0-\lambda_i}\Psi_{i\cdot}v$ and let $z = (z_1, ..., z_n)'$. We have: $\Psi' z = M_n^{(1)}(x_0)\, v = v$, and hence, $z = \frac{1}{x_0-\lambda_i}\Psi_{i\cdot}\Psi' z$, which proves that z is an eigenvector of $\frac{1}{T}\hat{X}\hat{X}'$ corresponding to eigenvalue x_0. Since the eigenvalues of $\frac{1}{T}\hat{X}\hat{X}'$ and $\frac{1}{T}XX'$ coincide, x_0 must be an eigenvalue of $\frac{1}{T}XX'$ which proves the "if" statement of the Lemma. The "only if" statement of the Lemma has been established in Section 4.□

Proof of Lemma 2:

We, first, formulate and prove the key technical lemma of this paper. Let $g_j(\lambda)$, $j = 1, ..., J$, be analytic functions of real variable λ on an open interval $(\bar{a}, \bar{b})$ containing the support of the Marčenko-Pastur distribution, that is the set $\{0, [a, b]\}$ if $c > 1$, and the segment $[a, b]$ if $c \geq 1$. Further, let $\varsigma^{(n)}$ be an array of $n \times m$ matrices with i.i.d. standard normal entries independent of $\lambda_1, ..., \lambda_n$. In what follows we will omit the superscript n in $\varsigma^{(n)}$ to simplify notations. Finally, denote the set of triples $\{(j, s, t) : 1 \leq j \leq J, 1 \leq s \leq t \leq m\}$ as Θ_1. Then, we have the following

Lemma 5: *Let Assumptions 2 and 3 hold. Then, the joint distribution of random variables* $\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_j(\lambda_i)(\varsigma_{is}\varsigma_{it} - \delta_{st}); (j, s, t) \in \Theta_1\right\}$ *weakly converges to a multivariate normal distribution as* $n \to \infty$. *The covariance between components* (j, s, t) *and* (j_1, s_1, t_1) *of the limiting distribution is equal to* 0 *when* $(s, t) \neq (s_1, t_1)$, *and to* $(1 + \delta_{st})\int g_j(\lambda)g_{j_1}(\lambda)d\mathcal{F}_c(\lambda)$ *when* $(s, t) = (s_1, t_1)$.

Proof: To prove this lemma we will need two well known results, which we formulate below as two additional lemmas.

Lemma 6: *(McLeish (1974)) Let* $\{X_{n,i}, \mathcal{F}_{n,i}; i = 1, 2, ..., n\}$ *be a martingale difference array on the probability triple* $(\Omega, \mathcal{F}, P)$. *If the following conditions are satisfied: a) Lindeberg's condition: for all* $\varepsilon > 0$, $\sum_i \int_{|X_{n,i}|>\varepsilon} X_{n,i}^2 dP \to 0, n \to \infty$; *b)* $\sum_{i=1}^{n} X_{n,i}^2 \xrightarrow{p} 1$, *then* $\sum_{i=1}^{n} X_{n,i} \xrightarrow{d} N(0, 1)$.

Proof: This is a consequence of Theorem (2.3) of McLeish (1974). Two conditions of the theorem, i) $\max_{i\leq n}|X_{n,i}|$ is uniformly bounded in L_2 norm, and ii) $\max_{i\leq n}|X_{n,i}| \xrightarrow{p} 0$, are replaced here by the Lindeberg condition. As explained in McLeish (1974), since for any ε, $\max_{i\leq n} X_{n,i}^2 \leq \varepsilon^2 + \sum_i X_{n,i}^2 I(|X_{n,i}|>\varepsilon)$ and since $P\{\max_{i\leq n}|X_{n,i}|>\varepsilon\} = P\left\{\sum_i X_{n,i}^2 I(|X_{n,i}|>\varepsilon)>\varepsilon^2\right\}$, both conditions i) and ii) follow from the Lindeberg condition.□

Lemma 7: *(Hall and Heyde) Let $\{X_{n,i},\mathcal{F}_{ni};1\leq i\leq n\}$ be a martingale difference array and define $V_{n,j}^2 = \sum_{i=1}^j E\left(X_{n,i}^2|\mathcal{F}_{n,i-1}\right)$ and $U_{n,j}^2 = \sum_{i=1}^j X_{n,i}^2$ for $1\leq j\leq n$. Suppose that the conditional variances $V_{n,n}^2$ are tight, that is $\sup_n P\left(V_{n,n}^2>\varepsilon\right)\to 0$ as $\varepsilon\to\infty$, and that the conditional Lindeberg condition holds, that is for all $\varepsilon>0$, $\sum_i E\left[X_{n,i}^2 I(|X_{n,i}|>\varepsilon)|\mathcal{F}_{n,i-1}\right]\xrightarrow{p} 0$. Then $\max_j\left|U_{n,j}^2-V_{n,j}^2\right|\xrightarrow{p} 0$.*

Proof: This is a shortened version of Theorem 2.23 in Hall and Heyde (1980).□

Returning to the proof of Lemma 5, let real numbers a_1 and b_1 be such that $[a_1,b_1]$ is included in $\left(\bar{a},\bar{b}\right)$, but itself includes the support of the Marčenko-Pastur law. Define functions $h_j(\lambda), j = 1,...,J$, so that $h_j(\lambda)=g_j(\lambda)$ for $\lambda\in[a_1,b_1]$, and $h_j(\lambda)=0$ otherwise. Note that $|h_j(\lambda)|<B$ for any $j=1,...,J$ and any λ, where B is a constant larger than $\max_{j=1,...,J}\sup_{\lambda\in[a_1,b_1]}|g_j(\lambda)|$. Note also that since, as shown in Bai, Silverstein and Yin (1988), λ_1 almost surely converges to b, $P\{\exists j\leq J, i\leq n \text{ such that } h_j(\lambda_i)\neq g_j(\lambda_i)\}\to 0$ as $n\to\infty$.

Consider random variables $X_{n,i}=\frac{1}{\sqrt{n}}\sum_{(j,s,t)\in\Theta_1}\gamma_{jst}h_j(\lambda_i)\left(\varsigma_{is}\varsigma_{it}-\delta_{st}\right)$, where γ_{jst} are some constants. Let $\mathcal{F}_{n,i}$ be sigma-algebra generated by $\lambda_1,...,\lambda_n$ and $\varsigma_{js};1\leq j\leq i,1\leq s\leq m$. Clearly, $\{X_{n,i},\mathcal{F}_{n,i};i=1,2,...,n\}$ form a martingale difference array. Let K be the number of different triples $(j,s,t)\in\Theta_1$. Consider an arbitrary order in Θ_1. In Hölder's inequality $\sum_{r=1}^K a_r b_r\leq \left(\sum_{r=1}^K (a_r)^p\right)^{1/p}\left(\sum_{r=1}^K (b_r)^q\right)^{1/q}$, which holds for $a_r>0, b_r>0, p>1, q>1$, and $(1/p)+(1/q)=1$, take $a_r=\left|\frac{1}{\sqrt{n}}\gamma_{jst}h_j(\lambda_i)\left(\varsigma_{is}\varsigma_{it}-\delta_{st}\right)\right|$, where (j,s,t) is the r-th triple in Θ_1, $b_r=1$, and $p=2+\delta$ for some $\delta>0$. Then, the inequality implies that $|X_{n,i}|^{2+\delta}\leq K^{1+\delta}B^{2+\delta}\sum_{(j,s,t)\in\Theta_1}\left|\gamma_{jst}\frac{\varsigma_{is}\varsigma_{it}-\delta_{st}}{\sqrt{n}}\right|^{2+\delta}$. Recalling that ς_{is} are i.i.d. standard normal random variables, we have: $\sum_i E|X_{n,i}|^{2+\delta}$ tends to zero as $n\to\infty$, which means that the Lyapunov condition holds for $X_{n,i}$. As is well known, Lyapunov's condition implies Lindeberg's condition. Hence, condition a) of McLeish's proposition is satisfied for $X_{n,i}$.

Now, let us consider $\sum_{i=1}^n X_{n,i}^2$. Since convergence in mean implies convergence in probability, the conditional Lindeberg condition is satisfied for $X_{n,i}$ because the unconditional Lindeberg condition is satisfied as checked above. Further, in notations of Hall and Heyde's proposition, we have $V_{n,n}^2=\frac{1}{n}\sum_{i=1}^n E(\sum_{\substack{(j,s,t)\in\Theta_1,\\(j_1,s_1,t_1)\in\Theta_1}}\gamma_{jst}\gamma_{j_1s_1t_1}h_j(\lambda_i)h_{j_1}(\lambda_i)\left(\varsigma_{is}\varsigma_{it}-\delta_{st}\right)\left(\varsigma_{is_1}\varsigma_{it_1}-\delta_{s_1t_1}\right)|\mathcal{F}_{n,i-1})$. It is straightforward to check that the latter expression is equal to

$\sum_{\substack{1\le j\le J\\1\le j_1\le J}}\left[\left(\sum_{1\le s\le t\le m}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\frac{1}{n}\sum_{i=1}^{n}h_j(\lambda_i)h_{j_1}(\lambda_i)\right].$

Consider now $\tilde{V}_{n,n}^2=\sum_{\substack{1\le j\le J\\1\le j_1\le J}}\left[\left(\sum_{1\le s\le t\le m}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\frac{1}{n}\sum_{i=1}^{n}g_j(\lambda_i)g_{j_1}(\lambda_i)\right]$. Since $P\left(\tilde{V}_{n,n}^2\neq V_{n,n}^2\right)\to 0$ as $n\to\infty$, $\tilde{V}_{n,n}^2$ and $V_{n,n}^2$ must converge in probability to the same limit, or must both diverge. But, by Theorem 1.1 of Bai and Silverstein (2004), $\frac{1}{n}\sum_{i=1}^{n}g_j(\lambda_i)g_{j_1}(\lambda_i)-\int g_j(\lambda)g_{j_1}(\lambda)d\mathcal{F}_{\frac{n}{T}}(\lambda)$ converges in probability to zero. Therefore, since $\mathcal{F}_{\frac{n}{T}}(\lambda)$ weakly converge to $\mathcal{F}_c(\lambda)$ as $n\to\infty$, we have

$$\tilde{V}_{n,n}^2\xrightarrow{p}\Sigma\equiv\sum_{\substack{1\le j\le J\\1\le j_1\le J}}\left[\left(\sum_{1\le s\le t\le m}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\int g_j(\lambda)g_{j_1}(\lambda)d\mathcal{F}_c(\lambda)\right].\tag{16}$$

Hence, $V_{n,n}^2$ also converges in probability to Σ. In particular, $V_{n,n}^2$ is tight and Hall and Heyde's proposition applies. From Hall and Heyde's proposition, we know that $\sum_{i=1}^{n}X_{n,i}^2$ must converge to the same limit as $V_{n,n}^2$. Therefore, using McLeish's result, we get $\sum_{i=1}^{n}X_{n,i}\xrightarrow{d}N(0,\Sigma)$.

Let us now define $Y_{n,i}=\sum_{(j,s,t)\in\Theta_1}\gamma_{jst}g_j(\lambda_i)\frac{\varsigma_{is}\varsigma_{it}-\delta_{st}}{\sqrt{n}}$. Since $P\left(\sum_{i=1}^{n}Y_{n,i}\neq\sum_{i=1}^{n}X_{n,i}\right)\to 0$ as $n\to\infty$, we have $\sum_{i=1}^{n}Y_{n,i}\xrightarrow{d}N(0,\Sigma)$. Finally, Lemma 5 follows from the latter convergence, the Cramer-Wold result (see White (1999), p.114), and definition of Σ (16).□

Now we turn to the proof of Lemma 2. To save the space, we will only study the convergence of $N_n^{(1)}(x)$. The joint convergence of $\left\{N_n^{(p)}(x);p=1,2,3\right\}$ can be demonstrated using similar ideas. We will prove the convergence of $N_n^{(1)}(x)$ by first checking the convergence of the finite dimensional distributions $\left\{N_{n,st}^{(1)}(x_j),(s,t,j)\in\Theta\right\}\xrightarrow{d}\left\{N_{st}^{(1)}(x_j),(s,t,j)\in\Theta\right\}$, where Θ denotes the set of all integer triples (s,t,j) satisfying $1\le s,t\le k$ and $1\le j\le J$, and, second, by demonstrating the tightness of all entries of $N_n^{(1)}(x)$.

Note that the distribution of $N_n^{(1)}(x)$ will not change if we substitute $O_{1:k}$ and $O\tilde{\varepsilon}_{1:k}$ in the definition of Ψ by $\xi(\xi'\xi)^{-1/2}$ and $\sigma\eta$, where ξ and η are two independent $n\times k$ matrix with i.i.d. standard normal entries independent from η,F, and $\lambda_1,...,\lambda_n$. Indeed, the substitution of $O\tilde{\varepsilon}_{1:k}$ by $\sigma\eta$ is justified by Assumption 2. As to the other substitution, note that the columns of $\xi(\xi'\xi)^{-1/2}$ are orthogonal and of unit length. Further, the joint distribution of elements of $\xi(\xi'\xi)^{-1/2}$ is invariant with respect to multiplication from the left by any orthogonal matrix. Hence, this distribution coincides with the joint distribution of the elements of the first k columns of random orthogonal matrix having Haar invariant distribution. But the latter is the joint distribution of elements of $O_{1:k}$. In the rest of the proof, we, therefore, will make the substitutions and redefine $N_n^{(1)}(x)$ accordingly.

It is straightforward to check that $N_n^{(1)}(x)=\sum_{v=1}^{10}S^{(v)}(x)$, where

$$S^{(1)}(x) = \left(\frac{F'F}{T}\right)^{1/2}(L'L)^{1/2}\left(\frac{\xi'\xi}{n}\right)^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{i.}'\xi_{i.}-I_k}{h(x,\lambda_i)}\right)\left(\frac{\xi'\xi}{n}\right)^{-1/2}(L'L)^{1/2}\left(\frac{F'F}{T}\right)^{1/2},$$
$$S^{(2)}(x) = \left(\frac{F'F}{T}\right)^{1/2}(L'L)^{1/2}\left(\frac{\xi'\xi}{n}\right)^{-1}(L'L)^{1/2}\sqrt{\frac{n}{T}}\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2}-I_k\right)\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)},$$
$$S^{(3)}(x) = \sqrt{\frac{n}{T}}\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2}-I_k\right)(L'L)^{1/2}\left(\frac{\xi'\xi}{n}\right)^{-1}(L'L)^{1/2}\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)},$$
$$S^{(4)}(x) = (L'L)^{1/2}\sqrt{n}\left(I_k-\left(\frac{\xi'\xi}{n}\right)\right)\left(\frac{\xi'\xi}{n}\right)^{-1}(L'L)^{1/2}\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)},$$
$$S^{(5)}(x) = \sqrt{n}\,(L'L-D)\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)},$$
$$S^{(6)}(x) = \sigma\sqrt{\frac{n}{T}}\left(\frac{F'F}{T}\right)^{1/2}(L'L)^{1/2}\left(\frac{\xi'\xi}{n}\right)^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{i.}'\eta_{i.}}{h(x,\lambda_i)}\right),$$
$$S^{(7)}(x) = \sigma\sqrt{\frac{n}{T}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{i.}'\xi_{i.}}{h(x,\lambda_i)}\right)\left(\frac{\xi'\xi}{n}\right)^{-1/2}(L'L)^{1/2}\left(\frac{F'F}{T}\right)^{1/2},$$
$$S^{(8)}(x) = \sigma^2\left(\frac{n}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{i.}'\eta_{i.}-I_k}{h(x,\lambda_i)},$$
$$S^{(9)}(x) = \sigma^2\sqrt{n}\left(\frac{n}{T}-c\right)I_k\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)},$$
$$S^{(10)}(x) = -\left(D+\sigma^2 cI_k\right)\sqrt{n}\left(\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)}\right)$$

By Theorem 1 of Bai and Silverstein (2004), $\sqrt{n}\left(\int\frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{n}\frac{1}{x-\lambda_i}\right)\xrightarrow{p}0$ for any $x\in[\theta_1,\theta_2]$. Our assumption that $n/T-c=o\left(1/\sqrt{n}\right)$ and the definition of Marčenko-Patur law imply that $\sqrt{n}\left(\int\frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda}-\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}\right)\xrightarrow{p}0$, and hence $\sqrt{n}\left(\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)}\right)\xrightarrow{p}0$. The latter convergence result together with the facts that $F'F/T\xrightarrow{p}I_k$, $\xi'\xi/n\xrightarrow{p}I_k$, $L'L-D=o(\sqrt{n})$, and $n/T-c=o(\sqrt{n})$ imply that $\left\{\sum_{v=1}^{10}S_{st}^{(v)}(x_j);(s,t,j)\in\Theta\right\}$ and $\left\{\sum_{v=1}^{10}\tilde{S}_{st}^{(v)}(x_j);(s,t,j)\in\Theta\right\}$ weakly converge to the same limit or do not converge together, where

$$\tilde{S}^{(1)}(x) = D^{1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{i.}'\xi_{i.}-I_k}{h(x,\lambda_i)}\right)D^{1/2},$$
$$\tilde{S}^{(2)}(x) = D\sqrt{c}\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2}-I_k\right)\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda},$$
$$\tilde{S}^{(3)}(x) = \sqrt{c}\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2}-I_k\right)D\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda},$$
$$\tilde{S}^{(4)}(x) = D^{1/2}\sqrt{n}\left(I_k-\left(\frac{\xi'\xi}{n}\right)\right)D^{1/2}\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda},$$
$$\tilde{S}^{(5)}(x) = 0,$$
$$\tilde{S}^{(6)}(x) = \sigma\sqrt{c}D^{1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{i.}'\eta_{i.}}{h(x,\lambda_i)}\right),$$
$$\tilde{S}^{(7)}(x) = \sigma\sqrt{c}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{i.}'\xi_{i.}}{h(x,\lambda_i)}\right)D^{1/2},$$
$$\tilde{S}^{(8)}(x) = \sigma^2 c\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{i.}'\eta_{i.}-I_k}{h(x,\lambda_i)},$$
$$\tilde{S}^{(9)}(x) = \tilde{S}^{(10)}(x) = 0.$$

Let us, first, consider the limit of $\left\{\tilde{S}_{st}^{(2)}(x_j)+\tilde{S}_{st}^{(3)}(x_j),(s,t,j)\in\Theta\right\}$. Since $\left(\frac{F'F}{T}\right)^{1/2}=I+\frac{1}{2}\left(\frac{F'F}{T}-I\right)+o_p\left(\frac{1}{\sqrt{T}}\right)$, using Assumption 3, we get $\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2}-I_k\right)\xrightarrow{d}\frac{1}{2}\Phi$. The latter convergence and the definition of $\tilde{S}^{(2)}(x)$, $\tilde{S}^{(3)}(x)$, and Φ imply that $\left\{\tilde{S}^{(2)}(x_j)+\tilde{S}^{(3)}(x_j),1\le j\le J\right\}\xrightarrow{d}\left\{\frac{\sqrt{c}}{2}(D\Phi+\Phi D)\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda},1\le j\le J\right\}$, and, hence, $\left\{\tilde{S}_{st}^{(2)}(x_j)+\tilde{S}_{st}^{(3)}(x_j),(s,t,j)\in\Theta\right\}$ weakly converge to $\left\{Z^{(1)}_{stj}, (s,t,j) \in \Theta\right\}$ having joint zero-mean Gaussian distribution such that

$$\operatorname{cov}\left(Z^{(1)}_{stj}, Z^{(1)}_{s_1t_1j_1}\right) = \frac{c}{4}\left(d_s + d_t\right)\left(d_{s_1} + d_{t_1}\right)\phi_{sts_1t_1}\int \frac{d\mathcal{F}_c(\lambda)}{x_j - \lambda}\int \frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda}. \tag{17}$$

Now, let us consider the limit of $\left\{\sum_{v\neq 2,3} \tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\}$. By definition, we have: $\sum_{v\neq 2,3} \tilde{S}^{(v)}_{st}(x_j) = \sqrt{d_s d_t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\xi_{is}\xi_{it}-\delta_{st}}{h(x_j,\lambda_i)} - \sqrt{d_s d_t}\int \frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it}-\delta_{st}}{\sqrt{n}} + \sigma\sqrt{cd_s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\eta_{it}}{h(x_j,\lambda_i)} + \sigma\sqrt{cd_t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{it}\eta_{is}}{h(x_j,\lambda_i)} + \sigma^2 c\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{is}\eta_{it}-\delta_{st}}{h(x_j,\lambda_i)}$. Since $[\xi,\eta]$ is an $n\times 2k$ matrix with i.i.d. standard normal entries, Lemma 5 and the above decomposition imply that $\left\{\sum_{v\neq 2,3} \tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\}$ weakly converge to $\left\{Z^{(2)}_{stj}, (s,t,j) \in \Theta\right\}$ having joint normal distribution such that $\operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right) = 0$ if $(s,t) \neq (s_1,t_1)$ and $(s,t) \neq (t_1,s_1)$ and $\operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right)$ is equal to

$$\begin{aligned}\operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right) &= \left[(1+\delta_{st})\left(\sigma^4c^2 + d_sd_t\right) + \sigma^2 c\left(d_s + d_t + 2\delta_{st}\sqrt{d_sd_t}\right)\right]\cdot \\ &\quad \cdot\int \frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)(x_{j_1}-\lambda)} - (1+\delta_{st})\, d_sd_t\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda}\end{aligned} \tag{18}$$

otherwise.

Finally, since $\left\{\tilde{S}^{(2)}_{st}(x_j) + \tilde{S}^{(3)}_{st}(x_j), (s,t,j) \in \Theta\right\}$ are, by definition, independent from $\left\{\sum_{v\neq 2,3} \tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\}$, $\left\{Z^{(1)}_{stj}, (s,t,j) \in \Theta\right\}$ must be independent from $\left\{Z^{(2)}_{stj}, (s,t,j) \in \Theta\right\}$ and $\left\{\sum_{v=1}^{10} \tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\} \xrightarrow{d} \left\{Z^{(1)}_{stj} + Z^{(2)}_{stj}; (s,t,j) \in \Theta\right\}$, having joint zero-mean Gaussian distribution such that $\operatorname{cov}\left(Z^{(1)}_{stj} + Z^{(2)}_{stj}, Z^{(1)}_{s_1t_1j_1} + Z^{(2)}_{s_1t_1j_1}\right) = \operatorname{cov}\left(Z^{(1)}_{stj}, Z^{(1)}_{s_1t_1j_1}\right) + \operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right)$. (17) and (18) imply that the joint distribution of $Z^{(1)}_{stj} + Z^{(2)}_{stj}$ is equal to that of $\left\{N^{(1)}_{st}(x_j); (s,t,j) \in \Theta\right\}$.

Now we have to prove the tightness of all entries of $N^{(1)}_n(x) = \sum_{v=1}^{10} S^{(v)}(x)$. Since product and sum are continuous mappings from $C\left[\theta_1,\theta_2\right]^2$ to $C\left[\theta_1,\theta_2\right]$, it is enough to prove the tightness of every entry of each matrix entering definition of $S^{(v)}(x)$, $v = 1, ..., 10$. Assumption 3 and the facts that $F'F/T \xrightarrow{p} I_k$, $\xi'\xi/n \xrightarrow{p} I_k$, $L'L - D = o\left(\sqrt{n}\right)$, and $n/T - c = o\left(\sqrt{n}\right)$ imply the tightness of every entry of each of the matrices $\left(\frac{F'F}{T}\right)^{1/2}$, $(L'L)^{1/2}$, $\sqrt{n}\left(L'L - D\right)$, $\left(\frac{\xi'\xi}{n}\right)^{-1/2}$, $\left(\frac{\xi'\xi}{n}\right)^{-1}$, $\sqrt{\frac{n}{T}}I$, $\sqrt{n}\left(\frac{n}{T} - c\right)I$, $\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2} - I_k\right)$, and $\sqrt{n}\left(I_k - \left(\frac{\xi'\xi}{n}\right)\right)$ considered as (constant) elements of $C\left[\theta_1,\theta_2\right]$. Therefore, we only need to prove the tightness of entries of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it}-\delta_{st}}{h(x,\lambda_i)},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\eta_{it}}{h(x,\lambda_i)},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{is}\eta_{it}-\delta_{st}}{h(x,\lambda_i)} \tag{19}$$

of $\sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}$ and of $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right)$.

Since ξ and η are, by definition, two independent $n \times k$ matrices with i.i.d. standard normal entries, to prove the tightness of the sequences of sums in (19), it is enough to prove the tightness of the first sum for all $1 \leq s \leq t \leq k$. We will use Theorem 12.3 of Billingsley (1968), p. 95. Condition i) of the theorem is equivalent in our context to the assumption of the tightness of the sum at $x = \theta_1$. Lemma 5 implies that this assumption is satisfied. We will verify condition ii) of Theorem 12.3 by proving the moment condition (12.51) of Billingsley (1968). We have $\frac{E\left(\sum_{i=1}^{n}\left(h(x_1,\lambda_i)^{-1}-h(x_2,\lambda_i)^{-1}\right)(\xi_{is}\xi_{it}-\delta_{st})\right)^2}{n(x_1-x_2)^2} \leq E\left(\sum_{i=1}^{n} (h(x_1,\lambda_i)h(x_2,\lambda_i))^{-1} (\xi_{is}\xi_{it} - \delta_{st})\right)^2 /n \leq \frac{16}{n(\theta_1-b)^4} E\left(\sum_{i=1}^{n} (\xi_{is}\xi_{it} - \delta_{st})\right)^2 = \frac{16}{(\theta_1-b)^4}(1+\delta_{st})$, where the first inequality follows from the fact that $\left|\frac{1}{h(x_1,\lambda_i)} - \frac{1}{h(x_2,\lambda_i)}\right| \leq \frac{|x_2-x_1|}{h(x_1,\lambda_i)h(x_2,\lambda_i)}$. Hence, $\sup_{n;x_1,x_2\in[\theta_1,\theta_2]} E\left(\sum_{i=1}^{n}\left(h(x_1,\lambda_i)^{-1} - h(x_2,\lambda_i)^{-1}\right)(\xi_{is}\xi_{it} - \delta_{st})\right)^2 /n\,(x_1-x_2)^2$ is finite and the moment condition (12.51) of Billingsley (1968) is satisfied. In a more complete proof (in which the tightness of the elements of $N_n^{(2)}(x)$ is demonstrated), we also need to check Billingsley's moment condition when $h(\cdot,\cdot)$ is replaced by $h^2(\cdot,\cdot)$. We can use the above reasoning and inequality $\left|\frac{1}{h^2(x_1,\lambda_i)} - \frac{1}{h^2(x_2,\lambda_i)}\right| \leq \frac{|x_2-x_1|(h(x_1,\lambda_i)+h(x_2,\lambda_i))}{h^2(x_1,\lambda_i)h^2(x_2,\lambda_i)} \leq \frac{32\theta_2|x_2-x_1|}{(\theta_1-b)^4}$ to perform such a check.

Similarly, conditions of Theorem 12.3 of Billingsley (1968) are satisfied for $\sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}$. Condition i) is satisfied because, as has been shown above, $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right) \xrightarrow{p} 0$ for any $x \in [\theta_1, \theta_2]$. Condition ii) is satisfied because $E\left(\sum_{i=1}^{n} \frac{1}{nh(x_1,\lambda_i)h(x_2,\lambda_i)}\right)^2 \leq \frac{16}{(\theta_1-b)^4}$.

To prove the tightness of $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right)$, we adopt the argument on page 563 of Bai and Silverstein (2004). In notations of Bai and Silverstein (2004), $\hat{M}_n(\cdot) \to -\frac{1}{2\pi i}\int \frac{1}{x-z}\hat{M}_n(z)dz$ is a continuous mapping of $C\left(\mathcal{C}, R^2\right)$ into $C[\theta_1,\theta_2]$. Since, $\hat{M}_n(\cdot)$ is tight, $-\frac{1}{2\pi i}\int \frac{1}{x-z}\hat{M}_n(z)dz$, and subsequently $n\left(\int \frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{n}\frac{1}{x-\lambda_i}\right)$, form a tight sequence. But $\sup_{x\in[\theta_1,\theta_2]} \sqrt{n}\left(\int \frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda} - \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}\right) \xrightarrow{p} 0$ because, by assumption, $n/T - c = o\left(1/\sqrt{n}\right)$. Therefore, $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{n}\frac{1}{x-\lambda_i}\right)$ is tight too. Finally, the latter tightness and the fact that $P\left\{\sum_{i=1}^{n} \frac{1}{\sqrt{n}}\left(\frac{1}{x-\lambda_i} - \frac{1}{h(x,\lambda_i)}\right) \neq 0\right\} \to 0$ imply that sequence $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right)$ must be tight.□

A derivation of the explicit formula for x_{0j}.

Recall that x_{0j} was defined as the solution to equation $\left(d_j + \sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} = 1$, and it is the probability limit of $\mu_j(\frac{1}{T}XX')$. Changing the roles of factors and factor loadings, it is straightforward to show that y_{0j} defined as the solution to $\left(cd_j + \sigma^2\frac{1}{c}\right)\int \frac{d\mathcal{F}_{\frac{1}{c}}(\lambda)}{y-\lambda} = 1$ must be the probability limit of $\mu_j(\frac{1}{n}X'X)$. But $\mu_j(\frac{1}{T}XX') = \frac{n}{T}\mu_j(\frac{1}{n}X'X)$. Hence, $x_{0j} = cy_{0j}$ and $\frac{d_j+\sigma^2}{c}\int \frac{\mathcal{F}_{\frac{1}{c}}(d\lambda)}{\frac{1}{c}x_{0j}-\lambda} = 1$. Now, it is straightforward to check that $f_{\frac{1}{c}}(\lambda) = c^2 f_c(c\lambda)$ and $\mathcal{F}_{\frac{1}{c}}$ does not have mass at zero if $c > 1$ and has mass at zero equal to $1-c$ if $c<1$. Therefore, we have $c\left(d_j+\sigma^2\right)\left(\int\frac{\mathcal{F}_c(d\lambda)}{x_{0j}-\lambda}-\frac{1-\frac{1}{c}}{x_{0j}}\right)=1$. Substituting $\int\frac{\mathcal{F}_c(d\lambda)}{x_{0j}-\lambda}$ by $\left(d_j+\sigma^2c\right)^{-1}$ in the latter equation, we get $1=c\left(d_j+\sigma^2\right)\left(\left(d_j+\sigma^2c\right)^{-1}-\frac{1-\frac{1}{c}}{x_{0j}}\right)$, which implies that $x_{0j}=\frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2c\right)}{d_i}$.

Proof of Lemma 3:

Let $R(z,\varkappa)=\left(A(\varkappa)-zI_k\right)^{-1}$ be the resolvent of $A(\varkappa)$ defined for all complex z not equal to any of the eigenvalues of $A(\varkappa)$. We will denote $R(z,0)$ as $R(z)$. Let Γ be a positively oriented circle in the complex plane with center at a_j and radius r_0. The second Neumann series for the resolvent $R(z,\varkappa)=R(z)+\sum_{n=1}^{\infty}(-\varkappa)^n R(z)\left(A^{(1)}R(z)\right)^n$ (see Kato (1980), p.67, for a definition of the second Neumann series) is uniformly convergent on Γ for $\varkappa<\min_{z\in\Gamma}\left(\left\|A^{(1)}\right\|\|R(z)\|\right)^{-1}=r_0/\left\|A^{(1)}\right\|$, where the last equality follows from the fact that $\|R(z)\|=r_0^{-1}$ for any $z\in\Gamma$. Therefore, formula (1.19) of Kato (1980) implies that, for $|\varkappa|<r_0/\left\|A^{(1)}\right\|$, there is exactly one eigenvalue, $a_j(\varkappa)$, inside the circle Γ. Formulae (3.6)[7] and (2.32) of Kato (1980) imply the inequality stated in part i of Lemma 3.

We now turn to the proof of part ii. According to Kato (1980), p.67, projection $P_j(\varkappa)$ can be represented as $P_j(\varkappa)=-\frac{1}{2\pi i}\int_\Gamma R(z,\varkappa)\,dz$. Substituting the second Neumann series for the resolvent in this formula, we obtain

$$P_j(\varkappa)=P_j-\frac{1}{2\pi i}\sum_{n=1}^{\infty}(-\varkappa)^n\int_\Gamma R(z)\left(A^{(1)}R(z)\right)^n dz \tag{20}$$

where $P_j\equiv P_j(0)$ and the series absolutely converges for $|\varkappa|<\frac{r_0}{\left\|A^{(1)}\right\|}$. Kato (1980), page 76, shows that $\frac{1}{2\pi i}\int_\Gamma R(z)A^{(1)}R(z)\,dz=-P_jA^{(1)}S_j-S_jA^{(1)}P_j$. This equality and (20) imply that $P_j(\varkappa)=P_j-\varkappa\left(P_jA^{(1)}S_j-S_jA^{(1)}P_j\right)-\frac{1}{2\pi i}\sum_{n=2}^{\infty}(-\varkappa)^n\int_\Gamma R(z)\left(A^{(1)}R(z)\right)^n dz$. Therefore, we have:

$$\left\|\frac{1}{\varkappa}\left(P_j(\varkappa)-P_j\right)+P_jA^{(1)}S_j+S_jA^{(1)}P_j\right\|\le\frac{|\varkappa|\left\|A^{(1)}\right\|^2}{r_0\left(r_0-|\varkappa|\left\|A^{(1)}\right\|\right)} \tag{21}$$

for any $|\varkappa|<r_0/\left\|A^{(1)}\right\|$.

Since A is diagonal with decreasing elements along the diagonal, e_j is an eigenvector of A corresponding to the eigenvalue a_j. By definition of $P_j(\varkappa)$, $e_j(\varkappa)\equiv\frac{P_j(\varkappa)e_j}{\|P_j(\varkappa)e_j\|}$ must be an eigenvector of $A(\varkappa)$ corresponding to the eigenvalue $a_j(\varkappa)$. Consider an identity $\frac{1}{\varkappa}\left(e_j(\varkappa)-e_j\right)+S_jA^{(1)}e_j=\left(\frac{1}{\varkappa}\left(P_j(\varkappa)e_j-e_j\right)+S_jA^{(1)}e_j\right)+\frac{1}{\varkappa}e_j(\varkappa)\left(1-\|P_j(\varkappa)e_j\|\right)$. Using (21) and the fact that $S_je_j=0$,

[7] Note the difference in notations. Kato's r_0 is ours $r_0/\left\|A^{(1)}\right\|$.

for the first term on right hand side of the identity we have:

$$\left\| \frac{1}{\varkappa} \left(P_j \left(\varkappa \right) e_j - e_j \right) + S_j A^{(1)} e_j \right\| \leq \frac{|\varkappa| \left\| A^{(1)} \right\|^2}{r_0 \left(r_0 - |\varkappa| \left\| A^{(1)} \right\| \right)}. \tag{22}$$

Using the fact that $P_j(\varkappa)$ is a projection operator so that $\|P_j(\varkappa) e_j\| \leq 1$ and $P_j(\varkappa)^2 = P_j(\varkappa)$, for the second term on right hand side of the identity we have:

$$\left\| \frac{1}{\varkappa} e_j \left(\varkappa \right) \left(1 - \| P_j \left(\varkappa \right) e_j \| \right) \right\| \leq \frac{1}{|\varkappa|} \left(1 - \| P_j \left(\varkappa \right) e_j \|^2 \right) = |\varkappa| \left\| \frac{1}{\varkappa} \left(P_j \left(\varkappa \right) e_j - e_j \right) \right\|^2. \tag{23}$$

But, form (22), $\left\| \frac{1}{\varkappa} \left(P_j(\varkappa) e_j - e_j \right) \right\|^2 \leq 2 \left\| S_j A^{(1)} e_j \right\|^2 + \frac{2|\varkappa|^2 \left\| A^{(1)} \right\|^4}{r_0^2 \left(r_0 - |\varkappa| \left\| A^{(1)} \right\| \right)^2} \leq \frac{\left\| A^{(1)} \right\|^2}{2r_0^2} + \frac{2|\varkappa|^2 \left\| A^{(1)} \right\|^4}{r_0^2 \left(r_0 - |\varkappa| \left\| A^{(1)} \right\| \right)^2}$. Combining the above identity, (22), (23), and the latter inequality, we obtain: $\left\| \frac{1}{\varkappa} \left(e_j(\varkappa) - e_j \right) + S_j A^{(1)} e_j \right\| \leq \frac{|\varkappa| \left\| A^{(1)} \right\|^2 \left(3r_0^2 - 4r_0 |\varkappa| \left\| A^{(1)} \right\| + 5|\varkappa|^2 \left\| A^{(1)} \right\|^2 \right)}{2r_0^2 \left(r_0 - |\varkappa| \left\| A^{(1)} \right\| \right)^2} \leq \frac{2|\varkappa| \left\| A^{(1)} \right\|^2}{\left(r_0 - |\varkappa| \left\| A^{(1)} \right\| \right)^2}$, where the last inequality follows from the fact that $r_0 > |\varkappa| \left\| A^{(1)} \right\|$. This proves statement ii) of the lemma.□

Proof of Lemma 4:

Since $f_n(x) \xrightarrow{d} f_0(x)$, $\{f_n(x)\}$ is tight and, hence, for any $\varepsilon > 0$, we can choose a compact K such that $P(f_n(x) \in K) > 1 - \frac{\varepsilon}{2}$ for all n. By the Arzelà-Ascoli theorem (see, for example, Billingsley (1999), p.81), for any positive ε_1, we have $K \subset \{f : |f(\theta_1)| \leq r\}$ for large enough r and $K \subset \{f : w_f(\delta(\varepsilon_1)) \leq \varepsilon_1\}$ for small enough $\delta(\varepsilon_1)$, where $w_f(\delta)$ is the modulus of continuity of function f, defined as $w_f(\delta) = \sup_{|s-t| \leq \delta} |f(s) - f(t)|$, $0 < \delta \leq \theta_2 - \theta_1$. Let us choose $N(\varepsilon, \varepsilon_1)$ so that for any $n > N(\varepsilon, \varepsilon_1)$, $P(|x_n - x_0| > \delta(\varepsilon_1)) < \frac{\varepsilon}{2}$. Then, for $n > N(\varepsilon, \varepsilon_1)$, we have: $P(|f_n(x_n) - f_n(x_0)| > \varepsilon_1) = P$
$(|f_n(x_n) - f_n(x_0)| > \varepsilon_1 \text{ and } |x_n - x_0| \leq \delta(\varepsilon_1)) + P(|f_n(x_n) - f_n(x_0)| > \varepsilon_1 \text{ and } |x_n - x_0| > \delta(\varepsilon_1)) \leq$
$P(f_n(x) \notin K) + P(|x_n - x_0| > \delta(\varepsilon_1)) < \varepsilon$, which proves the lemma.□

Derivation of the explicit formula for $\nu_j'(1)$:

By definition, $\nu_j'(1) = \left(\mu_j' \left(M_0^{(1)}(m_j) \right) \right)^{-1} = \left(-\left(d_j + \sigma^2 c \right) \int \frac{d\mathcal{F}_c(\lambda)}{(m_j - \lambda)^2} \right)^{-1}$. The latter expression can be simplified as follows. Consider m_j as a function of $d_j : m_j = \left(d_j + \sigma^2 \right) \left(d_j + \sigma^2 c \right) / d_j$. Note that since $m_j = x_{0j}$, and x_{0j} is defined as the solution to equation $\left(d_j + \sigma^2 c \right) \int \frac{d\mathcal{F}_c(\lambda)}{m_j - \lambda} = 1$, we must have:

$$\left(d_j + \sigma^2 c \right) \int \frac{d\mathcal{F}_c(\lambda)}{m_j - \lambda} = 1 \tag{24}$$

Differentiating both sides of (24) with respect to d_j, we get: $\left(d_j + \sigma^2 c \right)^{-1} - \left(d_j + \sigma^2 c \right) \int \frac{d\mathcal{F}_c(\lambda)}{(m_j - \lambda)^2} \left(1 - \frac{\sigma^4 c}{d_j^2} \right) =$

0. Solving this equation for the integral, we get:

$$\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2} = \frac{d_j^2}{\left(d_j^2-\sigma^4 c\right)\left(d_j+\sigma^2 c\right)^2}, \tag{25}$$

and therefore $-\nu_j'(1) = \frac{\left(d_j^2-\sigma^4 c\right)\left(d_j+\sigma^2 c\right)}{d_j^2}$.

A proof of the fact that the asymptotic joint distribution of $\sqrt{n}\left(R_{\cdot j}-R_{\cdot j}^{(1)}\right), j=1,...,q$ and $\sum_{s=1}^4 \varkappa_j \tilde{A}_j^{(s)}, j=1,...,q$ are the same:

Representation (12) implies that $\sqrt{n}\left(R_{\cdot j}-R_{\cdot j}^{(1)}\right) = \sum_{s=1}^4 A_j^{(s)} + o_p(1)$, where
$A_j^{(1)} = N_n^{(3)}(x_{nj})\, w_{nj}\, \|\Psi' y_{\cdot j}\|$,
$A_j^{(2)} = \sqrt{n}\left(D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{x_{nj}-\lambda} - D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{m_j-\lambda}\right) w_{nj}\, \|\Psi' y_{\cdot j}\|$,
$A_j^{(3)} = D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{m_j-\lambda}\sqrt{n}\left(w_{nj}-e_j\right)\|\Psi' y_{\cdot j}\|$,
$A_j^{(4)} = D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{m_j-\lambda} e_j \sqrt{n}\left(\|\Psi' y_{\cdot j}\| - p\lim \|\Psi' y_{\cdot j}\|\right)$.

Consider, first $A_j^{(3)}$ and $A_j^{(2)}$. Using the Taylor expansion of function $x^{-1/2}$ around probability limit of $\|\Psi' y_{\cdot j}\|^{-2}$, we get: $\sqrt{n}\left(\|\Psi' y_{\cdot j}\| - p\lim\|\Psi' y_{\cdot j}\|\right) = -\frac{1}{2} p\lim \|\Psi' y_{\cdot j}\|^3 \sqrt{n}\left(\|\Psi' y_{\cdot j}\|^{-2} - p\lim\|\Psi' y_{\cdot j}\|^{-2}\right) + o\left(\sqrt{n}\left(\|\Psi' y_{\cdot j}\|^{-2} - p\lim\|\Psi' y_{\cdot j}\|^{-2}\right)\right)$. As has been shown in the proof of Theorem 2, $\|\Psi' y_{\cdot j}\| \xrightarrow{p} \left(\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2}\right)^{-1/2}$ and, with high probability for large enough n, $\|\Psi' y_{\cdot j}\| = \left(w_{nj}' \hat{M}_n^{(2)}(x_{nj})\, w_{nj}\right)^{-1/2}$. Combining these facts with formulae (24) and (25) and using the Taylor expansion of $N_n^{(2)}(x_{nj})$ around m_j and Lemma 2, we get the following decomposition $A_j^{(3)} = \varrho_j e_j \left(w_{nj}+e_j\right)' \hat{M}_n^{(2)}(x_{nj})\sqrt{n}\left(w_{nj}-e_j\right) + \varrho_j e_j N_{n,jj}^{(2)}(x_{nj}) - 2\varrho_j e_j\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^3}\sqrt{n}\left(x_{nj}-m_j\right) + o_p(1)$, where $\varrho_j = -0.5\left(d_j^2-\sigma^4 c\right)^{3/2}\left(d_j+\sigma^2 c\right)^{1/2} d_j^{-5/2}$. Further, using Taylor expansion of function $\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}$ around $x = m_j$, $A_j^{(2)}$ can be transformed into $-D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2}\sqrt{n}\left(x_{nj}-m_j\right) w_{nj}\,\|\Psi' y_{\cdot j}\| + o_p(1)$.

The formulae obtained for $A_j^{(3)}$ and $A_j^{(2)}$ imply that we have the following representation $\sqrt{n}\left(R_{\cdot j}-R_{\cdot j}^{(1)}\right) = \sum_{s=1}^4 \hat{A}_j^{(s)} + o_p(1)$, where
$\hat{A}_j^{(1)} = N_n^{(3)}(x_{nj})\, w_{nj}\,\|\Psi' y_{\cdot j}\|$,
$\hat{A}_j^{(2)} = \varrho_j e_j N_{n,jj}^{(2)}(x_{nj})$,
$\hat{A}_j^{(3)} = -\left(D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2} w_{nj}\,\|\Psi' y_{\cdot j}\| + 2\varrho_j e_j\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^3}\right)\sqrt{n}\left(x_{nj}-m_j\right)$,
$\hat{A}_j^{(4)} = \left(D^{1/2}\int \frac{d\mathcal{F}_c(\lambda)}{m_j-\lambda}\|\Psi' y_{\cdot j}\| + \varrho_j e_j\left(w_{nj}+e_j\right)' \hat{M}_n^{(2)}(x_{nj})\right)\sqrt{n}\left(w_{nj}-e_j\right)$.

Statement ii) of Lemma 3 and Lemma 2 imply that

$$\sqrt{n}\left(w_{nj}-e_j\right) = -\tilde{S}(x_{nj})\, N_n^{(1)}(x_{nj})\, e_j + o_p(1), \tag{26}$$

where $\tilde{S}(x) = \left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}\right)^{-1} \mathrm{diag}\left((d_1-d_j)^{-1}, ..., \underbrace{0}_{j\text{-th position}}, ..., (d_k-d_j)^{-1}\right)$. Using the same argument as that in the derivation of the explicit formula for $\nu_j'(1)$ given in the previous section of the Appendix, we obtain

$$\int \frac{d\mathcal{F}_c(\lambda)}{(m_i-\lambda)^3} = \frac{\left(d_i^3 + c^2\sigma^6\right) d_i^3}{(d_i + c\sigma^2)^3 \left(d_i^2 - c\sigma^4\right)^3}, \tag{27}$$

$$\int \frac{d\mathcal{F}_c(\lambda)}{(m_i-\lambda)^4} = \frac{\left(d_i^6 + c^4\sigma^{12} + c\sigma^4 d_i^4 + 4c^2\sigma^6 d_i^3 + c^3\sigma^8 d_i^2\right) d_i^4}{(d_i + c\sigma^2)^4 \left(d_i^2 - c\sigma^4\right)^5}. \tag{28}$$

Finally, the definitions of $\hat{A}_j^{(s)}$ and x_{nj}, the facts that $\|\Psi' y_{\cdot j}\| \xrightarrow{p} \left(\left(d_j + \sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2}\right)^{-1/2}$, $w_{jn} \xrightarrow{p} e_j$, $x_{nj} \xrightarrow{p} m_j$, and $\hat{M}_n^{(2)}(x_{nj}) \xrightarrow{p} M_0^{(2)}(m_j)$, Lemma 4, and formulae (11), (26), (24), (25), and (27) imply that the distribution limit of $\left\{\sum_{s=1}^4 \hat{A}_j^{(s)}, j = 1, ..., q\right\}$ must be the same as that of $\left\{\sum_{s=1}^4 \varkappa_j \tilde{A}_j^{(s)}, j = 1, ..., q\right\}$, where $\varkappa_j$ and $\tilde{A}_j^{(s)}$ are as defined in the proof of Theorem 2.□

References

[1] Anderson, T.W. (1971) The statistical analysis of time series, John Wiley & Sons, New York, Chichister, Brisbane, Toronto.

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

[3] Bai, J. (2003) "Inferential Theory for Factor Models of Large Dimensions", *Econometrica* 71, 135-171.

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

[5] Bai, Z.D., and J.W. Silverstein (2004) "CLT of linear spectral statistics of large dimensional sample covariance matrices", *Annals of Probability* 32(1A) pp. 553-605

[6] Bai, Z., Silverstein, J., and Yin, Y. (1988) "A note on the largest eigenvalue of a large dimensional sample covariance matrix", *Journal of Multivariate Analysis* 26, pp 166-168.

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

[8] Baik, J. and J. Silverstein (2006) "Eigenvalues of large sample covariance matrices of spiked population models", *Journal of Multivariate Analysis* 97, 1382-1408.

[9] Bernanke, B., Boivin, J., and Eliasz, P. (2005) "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach", *The Quarterly Journal of Economics* 120, pp. 387-422.

[10] Billingsley, P. (1968) Convergence of probability measures, John Wiley & Sons, New York, Chichester, Brisbane, Toronto.

[11] Billingsley, P. (1999) Convergence of probability measures, second edition, Wiley, New York.

[12] Boivin J. and S. Ng (2006) "Are More Data Always Better for Factor Analysis?", *Journal of Econometrics* 132, p. 169-194.

[13] Connor, G. and R. Korajczyk (1988) "Risk and return in an equilibrium APT", *Journal of Financial Econometrics* 21, pp. 255-289.

[14] Forni, M. and Reichlin, L. (1998) "Let's get real: a factor analytical approach to disaggregated business cycle dynamics", *Review of Economic Studies*, 65, pp 453-473.

[15] Geman, S. (1980) "A limit theorem for the norm of random matrices", *Annals of Probability*, 8, 252-261.

[16] Hall, P., and C.C. Heyde (1980) Martingale limit theory and its application, New York : Academic Press.

[17] Hoyle, D. C. and M. Rattray (2007) "Statistical mechanics of learning multiple orthogonal signals: Asymptotic theory and fluctuation effects", *Physical Review E*, 75, 016101.

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

[19] Johnstone, I.M., and A. Y. Lu (2004) "Sparse principal components analysis", manuscript available at http:// www-stat.stanford.edu/ people/ faculty/ johnstone/ weblist/ publist.html

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

[21] Marčenko, V.A., and L.A. Pastur (1967) "Distribution of eigenvalues for some sets of random matrices", Math. USSR-Sbornik, vol. 1, no. 4, 457-483

[22] McLeish, D.L. (1974) "Dependent Central Limit Theorems and Invariance Principles", *Annals of Probability*, Vol. 2, No. 4, p.620-628.

[23] Milman V.D. (1988) "The heritage of P. Levy in geometrical functional analysis", *Astérisque*, 157-158, pp. 273-301.

[24] Paul, D. (2006) "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model", to appear in Statistica Sinica, available at www3.stat.sinica.edu.tw/statistica/ as preprint article no. SS-05-131

[25] Reimann, P., C. Van den Broeck and G. J. Bex (1996) "A Gaussian scenario for unsupervised learning", *Journal of Physics, A*, 29, pp. 3521-3535.

[26] Silverstein, J. W. (1990) "Weak convergence of random functions defined by the eigenvectors of sample covariance matrices", *Annals of Probability*, 18, 1174-1194.

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

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

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

[30] White, H. (1999) Asymptotic theory for econometricians, revised edition, Academic press, San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo.