Asymptotics of the principal components estimator of large factor models with weak factors

Alexei Onatski*
Economics Department, Columbia University

First draft: November, 2005
This draft: May, 2009

Abstract

We consider large factor models where factors' explanatory power does not strongly dominate the explanatory power of the idiosyncratic terms in finite samples, which is the situation often observed in the empirical applications. To study the principal components (PC) estimator of such a weak factors, we introduce a Pitman-drift-like asymptotic device, which we call weak factors asymptotics. We find the probability limits of the PC estimator under weak factors asymptotics when the idiosyncratic terms can be both cross-sectionally and temporally correlated. We show that the probability limits may be drastically different from the true factors and factor loadings even for factors with substantial explanatory power. For a special case of no cross-sectional and temporal correlation of the idiosyncratic terms, we establish the second order weak factors asymptotics of the PC estimator. The estimator is asymptotically normal with the covariance matrix depending on the strength of the factors and on the ratio of the cross-sectional and the temporal dimensions of the data.

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

*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

1 Introduction

Approximate factor models have recently attracted an increasing amount of attention from researchers in macroeconomics and finance (see Breitung and Eickmeier (2005) for a survey of numerous applications). The most popular technique for estimating factors in such models is the principal components (PC) analysis. Its consistency and asymptotic normality have been shown by Bai (2003). Unfortunately, as Monte Carlo experiments show (see, for example, Boivin and Ng (2006), Uhlig (2008), or Chapter 8 of Bai and Ng (2008)), the finite sample performance of the PC estimator is poor when the explanatory power of factors does not strongly dominate the explanatory power of the idiosyncratic terms. Such a situation is often encountered in practice. Its hallmark is the absence of clearly visible separation of the sample covariance eigenvalues into a group of large eigenvalues representing systematic variation and a group of small eigenvalues representing idiosyncratic variation (see Heaton and Solo (2006) for a related discussion).

This paper shows how and why the principal component estimates for large factor models might not be appropriate. We develop asymptotic approximation to the finite sample biases due to the relatively weak explanatory power of factors. We explicitly link these biases to the covariance structure of the idiosyncratic terms and show that they can be extremely large. For a 1-factor model calibrated to the European macroeconomic data used in Boivin et al. (2008), we find that the PC estimate of the factor is *orthogonal* to the true factor even if the latter explains as much as about 20% of the data's variance. To make things worse, the orthogonal estimate would be taken seriously by an econometrician because it (spuriously) explains 20% of the variation in the data! Our Monte Carlo experiments confirm good approximation quality of our asymptotics in finite samples with relatively weak factors.

Let us describe our main results in more detail. We consider approximate factor models:

$$X_{it} = L_i' F_t + e_{it} \text{ with } i \in \mathbb{N} \text{ and } t \in \mathbb{N}, \tag{1}$$

where F_t and L_i are $k \times 1$ vectors of factors and factor loadings, respectively, and e_{it} are possibly cross-sectionally and temporally correlated idiosyncratic components of X_{it}. The asymptotic identification is achieved by the following standard requirements.

First, the factors are normalized so that $E\left(\frac{1}{T}\sum_{t=1}^{T} F_t F_t'\right) = I_k$. Second, the idiosyncratic terms are only weakly correlated so that:

$$\lim \sup_{n,T\to\infty} \max \operatorname{eval} E\left(\frac{1}{T}ee'\right) < \infty, \tag{2}$$

where $\max \operatorname{eval}(A)$ denotes the maximal eigenvalue of matrix A and e denotes the $n \times T$ matrix with i,t-th elements e_{it}. Finally, the factors are pervasive in the sense that their cumulative loadings on n cross-sectional units rise proportionally to n :

$$\sum_{i=1}^{n} \frac{L_i L_i'}{n} \to S > 0. \tag{3}$$

Let X be an observed $n \times T$ matrix with i,t-th elements X_{it}, and let F and L be unobserved $T \times k$ and $n \times k$ matrices with j-th rows F_j' and L_j', respectively. Then we have: $X = LF' + e$. The PC 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 PC estimator of L, $\hat{L}$, is defined as $X\hat{F}/T$. We would like to study the properties of the PC estimators in the situation when the factors' *finite sample* explanatory power, as measured by $\sum_{i=1}^{n} L_i L_i'$, is weak, that is, only moderately larger than $\max \operatorname{eval} E\left(\frac{1}{T}ee'\right)$.

Note that if we fix model (1) and let n and T go to infinity, assumptions (2) and (3) would imply that, asymptotically, $\sum_{i=1}^{n} L_i L_i'$ is infinitely larger than $\max \operatorname{eval} E\left(\frac{1}{T}ee'\right)$. Hence, such an asymptotics would not provide a useful approximation to the finite samples with relatively weak factors. We will therefore consider a different asymptotics, where models (1) are drifting as n and T tend to infinity so that the *finite sample* explanatory power of factors remains bounded. Formally, we will consider a sequence of models (1) indexed by the cross-sectional dimension n, so that

$$\sum_{i=1}^{n} L_i^{(n)} L_i^{(n)\prime} - D \to 0 \tag{4}$$

as n and $T^{(n)}$ go to infinity proportionately, where $D = \operatorname{diag}(d_1, ..., d_k)$ with $d_1 > ... > d_k > 0$.[1]

The above asymptotic device is similar to the well-known Pitman drift (see, for

[1]Note that for the model indexed by a particular n, we still have $\sum_{i=1}^{N} \frac{L_i^{(n)} L_i^{(n)\prime}}{N} \longrightarrow S > 0$ as $N \to \infty$.

example, Davidson and MacKinnon (2004)), used in the asymptotic power comparisons of consistent tests. Since for any fixed alternative, the asymptotic power of any consistent test equals 1, no sensible finite sample power comparison is made by considering asymptotics under a fixed alternative. A much more interesting asymptotics is obtained when the alternatives drift towards the null as the sample size rises. In the unit roots literature, for example, such an analysis is called local-to-unity asymptotics (see Stock (1994)). By analogy, the reader can call asymptotics (4) local-to-non-pervasiveness asymptotics or, to avoid complicated terms, weak factors asymptotics.

This paper answers the question: what is the first and the second order weak factors asymptotics of the PC estimators of the factors and factor loadings. We develop the first order asymptotics under the assumption that the matrix of the idiosyncratic terms can be represented as $e^{(n)} = A^{(n)}\varepsilon^{(n)}B^{(n)}$, where $A^{(n)}$ and $B^{(n)}$ are relatively unrestricted $n \times n$ and $T^{(n)} \times T^{(n)}$ matrices and $\varepsilon^{(n)}$ is an $n \times T^{(n)}$ matrix with i.i.d. entries with mean zero, variance σ^2 and finite fourth moment. Similar assumptions have been previously made in Onatski (2005), Bai and Ng (2005) and Harding (2006). The assumption allows the idiosyncratic terms to be non-trivially correlated both cross-sectionally and over time. We discuss its relation to economic models in Section 2.

Our main results refer to the $k \times k$ matrix $Q \equiv (F'F)^{-1} F'\hat{F}$ of the OLS coefficients in the regression of the PC estimates $\hat{F}$ on the true factors F. Theorem 1 below shows[2] that the probability limit of Q under the weak factors asymptotics is a diagonal matrix with the diagonal elements strictly smaller than one. It further describes the diagonal elements of $\operatorname{plim} Q$ as specific functions of matrix D from (4), which measures the finite sample strength of factors; of σ^2, which scales the variance of the idiosyncratic terms; and of the limiting empirical eigenvalue distributions of the matrices $A^{(n)}$ and $B^{(n)}$, which encode the degree of the cross-sectional and temporal correlation of the idiosyncratic terms.

We show that for a wide range of cases when σ^2 is not negligible relative to D and when the empirical eigenvalue distributions of $A^{(n)}$ and $B^{(n)}$ are relatively dispersed, the probability limit of Q equals zero, which means that the PC estimates have no relation to the true factors whatsoever! In contrast, under the "strong factors

[2] All theorems in the paper, and Theorem 1 in particular, describe the asymptotics of the factor loadings estimates in addition to that of the factor estimates.

asymptotics" where the factors' explanatory power is asymptotically infinitely larger than that of the idiosyncratic terms, $\operatorname{plim} Q$ is the identity matrix so that the principal components estimator $\hat{F}$ is consistent for F.

We extend our analysis to the second order weak factors asymptotics only in the special case when $A^{(n)}$ and $B^{(n)}$ are identity matrices and the elements of $\varepsilon^{(n)}$ are Gaussian. For this reason, our second order asymptotic results, described in Theorems 2 and 3, are of little empirical relevance. They are, however, interesting from a theoretical point of view.

Under the above restrictive assumptions, we find that the PC estimates of the factors at particular time periods (or factor loadings corresponding to specific cross-sectional units) are inconsistent but asymptotically jointly normal, and we find explicit formulae for the corresponding asymptotic biases and covariance matrices. As D tends to infinity so that the finite sample cumulative effects of the factors on the cross-sectional units becomes larger and larger, the biases disappear. Moreover, our second order 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 PC 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 of the normalized factor loadings 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.

Let us now describe some related literature. An alternative approach to modeling weak factors has been recently proposed in DeMol et al. (2008). The authors of that paper replace (2) by a weaker assumption: $\limsup_{n,T\to\infty} \max \operatorname{eval} E\left(\frac{1}{n^{1-\alpha}T}ee'\right) < \infty$, where $0 < \alpha \leq 1$. Dividing the data by $n^{(1-\alpha)/2}$ and redefining L and e as $n^{(\alpha-1)/2}L$ and $n^{(\alpha-1)/2}e$, respectively, we see that such a modeling strategy is equivalent to maintaining (2) but assuming that $\sum_{i=1}^{n} \frac{L_i L_i'}{n^\alpha} \to S > 0$ with $0 < \alpha \leq 1$. In this paper we study the case $\alpha = 0$, which is not considered by DeMol et al. (2008). When $0 < \alpha < 1$, the PC estimator remains consistent but its rate of convergence decreases relative to the strong factor case: $\alpha = 1$.

In the statistical literature, the weak factors asymptotics of the PC estimators have been recently studied by Johnstone and Lu (2007) and by Paul (2007).[3] For a 1-factor model with i.i.d. Gaussian factor and i.i.d Gaussian idiosyncratic terms, Johnstone and Lu (2007) show that the one-dimensional analog of our Q remains separated from one as n and T go to infinity proportionately. Paul (2007) quantifies the amount of the inconsistency pointed out by Johnstone and Lu (2007) 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 (2007) finds the asymptotic distribution of the eigenvectors corresponding to the k largest eigenvalues.

In contrast to Johnstone and Lu (2007) and Paul (2007), our first order asymptotic analysis does not require the idiosyncratic terms be i.i.d. and Gaussian. Allowing the idiosyncratic terms to be correlated is crucial for macroeconomic and financial applications. Further, our second order asymptotic analysis uses a substantially different machinery than the proofs of Paul (2007), which allows us to relax his requirement that factors are i.i.d. Gaussian.

The rest of the paper is organized as follows. In Section 2 we state our assumptions and describe the first order asymptotic results. In Section 3 we describe the second order asymptotic results. Monte Carlo analysis is given in Section 4. Section 5 concludes. All proofs are relegated to the Technical Appendix available from the author's web site at http://www.columbia.edu/~ao2027.

2 First order asymptotics

As explained in the Introduction, we study finite samples of increasing dimensions from a sequence of approximate factor models (1). Finite samples of the cross-sectional size n and temporal size $T^{(n)}$ are summarized in $n \times T^{(n)}$ matrices $X^{(n)}$, which can be represented as:

$$X^{(n)} = L^{(n)} F^{(n)\prime} + A^{(n)} \varepsilon^{(n)} B^{(n)}, \tag{5}$$

where the parameters of the representation satisfy Assumptions 1, 2, and 3, described below.

[3] Although these papers study the weak factor asymptotics they have a different motivation and do not use "weak factor" terminology.

Assumption 1: *There exist a positive constant* c *and a* $k \times k$ *diagonal matrix* $D \equiv \operatorname{diag}(d_1, ..., d_k)$, $d_1 > ... > d_k > 0$, *such that, as* n *tends to infinity:*
i) $n/T^{(n)} \to c$,
ii) $\frac{1}{T} F^{(n)\prime} F^{(n)} \xrightarrow{p} I_k$,
iii) $L^{(n)\prime} L^{(n)} \xrightarrow{p} D$.

Part i) of the assumption requires that n and $T^{(n)}$ be comparable even asymptotically. It implies that, when n tends to infinity, $T^{(n)}$ also tends to infinity. Such a simultaneous divergence to infinity stands in contrast to the classical assumption of fixed n and rising T. Parts ii) and iii) of the assumption describe the asymptotic normalization of factors and the weak factors asymptotics, discussed in the Introduction, respectively. Part iii) slightly generalizes (4) by replacing the ordinary convergence by the convergence in probability. In particular, we allow both factors and factor loadings be random.

In what follows, we will omit the superscript (n) from notations $T^{(n)}$, $X^{(n)}$, $L^{(n)}$, $F^{(n)}$, $A^{(n)}$, $\varepsilon^{(n)}$ and $B^{(n)}$ to make them easier to read. Let us denote the normalized loadings $L(L'L)^{-1/2}$ as $\mathcal{L}$. Note that the columns of $\mathcal{L}$, denoted as $\mathcal{L}_1, ..., \mathcal{L}_k$, represent a system of orthonormal vectors in $\mathbb{R}^n$. Let $\mathcal{L}_{k+1}, ..., \mathcal{L}_n$ be any vectors that complement $\mathcal{L}_1, ..., \mathcal{L}_k$ to an orthonormal basis in $\mathbb{R}^n$. Similarly, let us denote the normalized factors $F(F'F)^{-1/2}$ as $\mathcal{F}$. Let $\mathcal{F}_1, ..., \mathcal{F}_k$ be the columns of $\mathcal{F}$ and let $\mathcal{F}_{k+1}, ..., \mathcal{F}_T$ complement them to an orthonormal basis in $\mathbb{R}^T$. Finally, define Δ as $(L'L)^{1/2}(F'F)^{1/2}$ and let us denote the smallest eigenvalue of $E(\Delta\Delta'|L)$ as λ_L and the smallest eigenvalue of $E(\Delta'\Delta|F)$ as λ_F.

Assumption 2: i) ε *is an* $n \times T$ *matrix with i.i.d. entries* ε_{it} *independent from* F *and* L *and such that* $\mathrm{E}\,\varepsilon_{it} = 0, \mathrm{E}\,\varepsilon_{it}^2 = \sigma^2$ *and* $\mathrm{E}\,\varepsilon_{it}^4 < \infty$,
ii) $A = [a_1\mathcal{L}_1, ..., a_n\mathcal{L}_n]$, where $a_i = 1$ for $i \leq k$, $\sum_{i=1}^n a_i^2 = n$ and $\max\limits_{j=1,...,n} a_i^2 < \frac{\lambda_L}{\sigma^2 T} + 1$,
iii) $B' = [b_1\mathcal{F}_1, ..., b_T\mathcal{F}_T]$, where $b_i = 1$ for $i \leq k$, $\sum_{i=1}^T b_i^2 = T$ and $\max\limits_{j=1,...,T} b_i^2 < \frac{\lambda_F}{\sigma^2 n} + 1$

Matrices A and B introduce the cross-sectional and temporal correlation in the idiosyncratic terms $e \equiv A\varepsilon B$. Although non-trivial, the pattern of covariances between different elements of matrix e remains restrictive. Denoting the $nT \times 1$ vector of stacked columns of e as $\operatorname{vec} e$, we have: $E\left(\operatorname{vec} e\,(\operatorname{vec} e)'\right) = B'B \otimes AA'$. How well $B'B \otimes AA'$ can approximate more interesting covariance structures depends on the details of these structures. For a general discussion of the quality of approximations with Kronecker products see Van Loan and Pitsianis (1993).

In economic applications, the covariance matrix of $\operatorname{vec} e$ exactly equals the Kroneker product of two matrices only in special cases. For example, in the spirit of Forni and Lippi (1999, 2001), consider an n-industry constant-returns economy, where the productions X_{it} in industries $i = 1, ..., n$ at time t are given by the equations:

$$\begin{pmatrix} 1 & w_{12} & \dots & w_{1n} \\ w_{21} & 1 & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 1 \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{nt} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} F_t + \begin{pmatrix} b_1(L)\varepsilon_{1t} \\ b_2(L)\varepsilon_{2t} \\ \vdots \\ b_n(L)\varepsilon_{nt} \end{pmatrix},$$

where F_t is a demand common shock, $b_i(L)\varepsilon_{it}$ are autocorrelated idiosyncratic productivity shocks, and ε_{it} are i.i.d. innovations to these shocks. For such a model, w_{ji} is the quantity of the i-th product necessary as a means of production to produce one unit of the j-th output. Inverting the input-output matrix W, we obtain:

$$X_t = \Lambda F_t + W^{-1}\varepsilon_t b(L),$$

where $b(L) \equiv \operatorname{diag}(b_1(L), b_2(L), ..., b_n(L))$. In the special case when all $b_i(L)$ with $i = 1, ..., n$ are the same so that all the productivity shocks have the same dynamics described by a filter $b_0 + b_1 L + b_2 L^2 + ...$, we can write: $e = A\varepsilon B$, where $A = W^{-1}$ and B is such that $B_{ij} = 0$ for for $i > j$ and $B_{ij} = b_{j-i}$ for for $i \leq j$.[4]

Note that, for the above application, matrix B is Toeplitz. Such a form of B is a consequence of the stationarity of the data. In this paper, we do not constrain B be Toeplitz. In fact, matrices A and B may have similar properties. Hence, we allow as much non-stationarity in the time dimension as there is in the cross-sectional dimension.

There are several reasons to restrict A and B as in parts ii) and iii) of Assumption 2. First, the main justification of the PC estimator of approximate factor models is that for a fixed model, the factor loadings and the factors asymptotically span the subspace of the first k principal eigenvectors of $E(XX'|L)$ and $E(X'X|F)$, respectively. For the weak factors asymptotics this is not necessarily so. Conditions ii) and iii) of Assumption 2 reconcile the weak factors asymptotics with the PC method by implying that $E(XX'|L) = \mathcal{L}E(\Delta\Delta'|L)\mathcal{L}' + \sigma^2 T \sum_{i=1}^{n} a_i^2 \mathcal{L}_i \mathcal{L}_i'$ and $E(X'X|F) = \mathcal{F}E(\Delta'\Delta|F)\mathcal{F}' + \sigma^2 n \sum_{i=1}^{T} b_i^2 \mathcal{F}_i \mathcal{F}_i'$ so that L and F span the subspaces of the corresponding k principal eigenvectors.

[4]Here we assume that the innovations ε_{it} for $t = 0, -1, -2, ...$ equal zero.

The constraints $\max_{j=1,...,n} a_i^2 < \frac{\lambda_L}{\sigma^2 T} + 1$ and $\max_{j=1,...,T} b_i^2 < \frac{\lambda_F}{\sigma^2 n} + 1$ make sure that L and F span the principal eigenspaces, that is, the spaces corresponding to the k largest eigenvalues, as opposed to some other eigenspaces. The constraints $a_i = 1$ for $i \leq k$ and $\sum_{i=1}^n a_i^2 = n$ make possible the separate identification of σ^2 and $\{a_i, i = 1, ..., n\}$ from $E(XX'|L)$. Similarly, the constraints $b_i = 1$ for $i \leq k$ and $\sum_{i=1}^T b_i^2 = T$ make possible the separate identification of σ^2 and $\{b_i, i = 1, ..., T\}$ from $E(X'X|F)$.

The second reason to impose constraints ii) and iii) is related to the first one. By making the spaces spanned by L and F the principal eigenspaces of $E\left(\frac{1}{T}XX'|L\right)$ and $E\left(\frac{1}{n}X'X|F\right)$, we identify them even in finite samples, which allows us to unambiguously discuss the finite sample performance of the PC estimator. Without such an identification, no information about F and L can be extracted from a finite sample. Indeed, for any $\tilde{L}$ and $\tilde{F}$, we always can represent $LF' + e$ as $\tilde{L}\tilde{F}' + \tilde{e}$ with $\tilde{e} = \left(LF' - \tilde{L}\tilde{F}'\right) + e$ and justify such a representation by saying that the future observations will reveal that the loadings and factors are consistent with $\tilde{L}$ and $\tilde{F}$ rather than with L and F.

Third, parts ii) and iii) of Assumption 2, while being restrictive, do allow for non-trivial cross-sectional and temporal correlation of the idiosyncratic terms. In addition, they allow us to use large random matrix machinery, which facilitates our proofs. Our next assumption supplies additional asymptotic requirements which allow us to use large random matrix theory results established in Zhang (2006), Paul and Silverstein (2008) and Onatski (2005).

To formulate the next assumption we introduce new notation. Let $\lambda_1(M) \geq ... \geq \lambda_n(M)$ be the eigenvalues of a generic $n \times n$ symmetric matrix M. We define the eigenvalue distribution function for M as

$$F^M(x) = 1 - \frac{1}{n}\#\{i \leq n : \lambda_i(M) > x\}, \tag{6}$$

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

Assumption 3: *As* n and T *go to infinity:*

i) $F^{AA'}$ *and* $F^{B'B}$ *weakly converge to probability distribution functions with bounded support* $\mathcal{F}^A$ *and* $\mathcal{F}^B$, *respectively,*

ii) $u\left(F^{AA'}\right) \to u(\mathcal{F}^A) > 0$ *and* $u\left(F^{B'B}\right) \to u(\mathcal{F}^B) > 0,$

iii) $\liminf_{\delta\to 0} \frac{1}{\delta}\int_{u(\mathcal{F}^A)-\delta}^{u(\mathcal{F}^A)} d\mathcal{F}^A(\lambda) = k_A > 0$ *and* $\liminf_{\delta\to 0} \frac{1}{\delta}\int_{u(\mathcal{F}^B)-\delta}^{u(\mathcal{F}^B)} d\mathcal{F}^B(\lambda) = k_B > 0.$

The cdf's $F^{AA'}$ and $F^{B'B}$ can be interpreted as the empirical distributions of sequences $\{a_i^2, i = 1, ..., n\}$ and $\{b_i^2, i = 1, ..., T\}$, respectively. Therefore, parts i) and ii) of the assumption would be satisfied if $\{a_i^2, i = 1, ..., n\}$ and $\{b_i^2, i = 1, ..., T\}$ are random samples from $\mathcal{F}^A$ and $\mathcal{F}^B$. Note that sequences $\{\sigma^2 a_i^2, i = 1, ..., n\}$ and $\{\sigma^2 b_i^2, i = 1, ..., T\}$ represent the eigenvalues of the covariance matrices $E\left(\frac{ee'}{T}|L\right)$ and $E\left(\frac{e'e}{n}|F\right)$, respectively. Therefore, part iii) of the assumption implies clustering of the largest eigenvalues of these covariance matrices. In particular, it does not allow a few linear cross-sectional or temporal combinations of idiosyncratic terms to have "unusually" large variation, which would make these combinations difficult to separate from combinations of the genuine factors in finite samples.

Now, let $\hat{F}$, $\hat{L}$ and $\hat{\mathcal{L}} \equiv \hat{L}\left(\hat{L}'\hat{L}\right)^{-1/2}$ be the PC estimators of factors, factor loadings and the normalized factor loadings, respectively. That is, $\hat{F}$ equals $\sqrt{T}$ times the matrix of the principal k eigenvectors[5] of $\frac{1}{T}X'X$ and $\hat{L}$ equals $\frac{1}{T}X\hat{F}$. Our first theorem describes the probability limits under the weak factors asymptotics of the coefficients in the projection decompositions

$$\begin{aligned} \hat{F} &= FQ + F^{\perp} \text{ and} && (7) \\ \hat{\mathcal{L}} &= \mathcal{L}P + \mathcal{L}^{\perp}, && (8) \end{aligned}$$

where $Q = (F'F)^{-1} F'\hat{F}$ and $P = \mathcal{L}'\hat{\mathcal{L}}$. We will also describe the probability limit of $\hat{L}'\hat{L}$, whose diagonal elements measure the cumulative effects of factors on the cross-sectional units.

The formulation of the theorem refers to solutions of the system:

$$\begin{cases} v = w\left(c\int \frac{au}{u-a} d\mathcal{F}^A(a)\right)^{-1} \\ u = w\left(\int \frac{bv}{v-b} d\mathcal{F}^B(b)\right)^{-1} \end{cases}, \quad (9)$$

with unknown real $v > u\left(\mathcal{F}^B\right)$, unknown real $u > u\left(\mathcal{F}^A\right)$ and real non-negative parameter w. As shown in Onatski (2005), there exists $\bar{w} > 0$ (equal to the probability limit of the largest eigenvalue of $A\varepsilon BB'\varepsilon'A'$) such that for $w < \bar{w}$, system

[5] To eliminate the indeterminacy due to the fact that eigenvectors are defined up to a sign, we require that the sign is chosen so that the scalar products of the eigenvectors and the corresponding factors are non-negative.

(9) has no solutions; for $w = \bar{w}$, the system has exactly one solution, which we will denote as $\bar{u}$ and $\bar{v}$; and for $w > \bar{w}$, the system has two solutions (u_{1w}, v_{1w}) and (u_{2w}, v_{2w}) such that $v_{2,w} > v_{1,w}$ and $u_{2,w} > u_{1,w}$. For each $x > \bar{w}\left(1 - \bar{u}^{-1}\right)\left(1 - \bar{v}^{-1}\right)$, let us define $w(x)$, $u(x)$ and $v(x)$ as w, u_{2w} and v_{2w}, respectively, which satisfy[6] $x = w\left(1 - (u_{2,w})^{-1}\right)\left(1 - (v_{2,w})^{-1}\right)$.

Theorem 1: *Let Assumptions 1-3 hold and q be such that $\frac{d_i}{\sigma^2} > \bar{w}\left(1 - \bar{u}^{-1}\right)\left(1 - \bar{v}^{-1}\right)$ for $i \leq q$, and $\frac{d_i}{\sigma^2} \leq \bar{w}\left(1 - \bar{u}^{-1}\right)\left(1 - \bar{v}^{-1}\right)$ for $i > q$. For any $i \leq q$, define $w_i = w\left(\frac{d_i}{\sigma^2}\right)$, $u_i = u\left(\frac{d_i}{\sigma^2}\right)$, $v_i = v\left(\frac{d_i}{\sigma^2}\right)$, $r_{ui} = \int\left(\frac{a}{u_i - a}\right)^2 d\mathcal{F}^A(a) / \int \frac{a}{u_i - a} d\mathcal{F}^A(a)$ and $r_{vi} = \int\left(\frac{b}{v_i - b}\right)^2 d\mathcal{F}^B(b) / \int \frac{b}{v_i - b} d\mathcal{F}^B(b)$ Then, as n (and by Assumption 1 i) also T) goes to infinity:*

i) *The matrix coefficient Q from (7) converges to a diagonal matrix with non-negative diagonal elements strictly smaller than one so that:*

$$Q_{ii}^2 \xrightarrow{p} \left[1 + \frac{1 + r_{vi}}{v_i\left(1 - r_{ui} r_{vi}\right)}\left(r_{ui} + \frac{v_i - 1}{u_i - 1}\right)\right]^{-1} \text{ for } i \leq q \text{ and}$$
$$Q_{ii}^2 \xrightarrow{p} 0 \text{ for } i > q,$$

ii) *The matrix coefficient P from (8) converges to a diagonal matrix with non-negative diagonal elements strictly smaller than one so that:*

$$P_{ii}^2 \xrightarrow{p} \left[1 + \frac{1 + r_{ui}}{u_i\left(1 - r_{ui} r_{vi}\right)}\left(r_{vi} + \frac{u_i - 1}{v_i - 1}\right)\right]^{-1} \text{ for } i \leq q \text{ and}$$
$$P_{ii}^2 \xrightarrow{p} 0 \text{ for } i > q,$$

iii) *Matrix $\hat{L}'\hat{L}$ converges to a diagonal matrix so that*

$$\left(\hat{L}'\hat{L}\right)_{ii} \xrightarrow{p} \sigma^2 w_i \text{ for } i \leq q \text{ and}$$
$$\left(\hat{L}'\hat{L}\right)_{ii} \xrightarrow{p} \sigma^2 \bar{w} \text{ for } i > q.$$

[6] It is easy to solve for $\bar{w}$, $\bar{u}$, $\bar{v}$, $w(x)$, $u(x)$ and $v(x)$ numerically for any distribution functions $\mathcal{F}^A$ and $\mathcal{F}^B$. It is because $\left(c \int \frac{au}{u-a} d\mathcal{F}^A(a)\right)^{-1}$ and $\left(\int \frac{bv}{v-b} d\mathcal{F}^B(b)\right)^{-1}$ are strictly concave functions of u and v and $w\left(1 - (u_{2w})^{-1}\right)\left(1 - (v_{2w})^{-1}\right)$ is a strictly increasing function of w. A matlab code, which implements the solution is available from the author upon request.

Let us illustrate the results of the theorem using an example. Suppose data are generated by the following 1-factor model:

$$\begin{aligned} X_{it} &= \sqrt{dT}\mathcal{L}_{1i}\mathcal{F}_{1t} + e_{it}, \text{ where} \\ e_{it} &= \rho_1 e_{i-1,t} + c_1\xi_{it} \text{ and} \\ \xi_{it} &= \rho_2\xi_{it-1} + c_2\eta_{it}, \ \eta_{it} \sim \text{iid}N(0,1) \end{aligned} . \tag{10}$$

Here $\sqrt{d}\mathcal{L}_1$ is the vector of loadings, $\sqrt{T}\mathcal{F}_1$ is the factor, and the idiosyncratic terms e_{it} follow auto-regressions both temporally and cross-sectionally.

Note that $\text{vec}\,(e)$ is an $nT \times 1$ Gaussian vector with covariance matrix $T_2 \otimes T_1$, where T_1 and T_2 are Toeplitz matrices with i, j-th entries equal to $c_1^2\rho_1^{|i-j|}/\left(1-\rho_1\right)$ and $c_2^2\rho_2^{|i-j|}/\left(1-\rho_2\right)$, respectively. Therefore, e can be represented in the form $A\varepsilon B$, where ε is an $n \times T$ matrix with iid$N(0,1)$ entries, $A = \mathcal{L}\mathcal{A}_0$ with $\mathcal{L}$ being an $n \times n$ matrix of eigenvectors of T_1 and $\mathcal{A}_0$ being the diagonal matrix of the square roots of the corresponding eigenvalues of T_1, and $B = \mathcal{B}_0\mathcal{F}'$ with $\mathcal{F}$ being a $T \times T$ matrix of eigenvectors of T_2 and $\mathcal{B}_0$ being the diagonal matrix of the square roots of the corresponding eigenvalues of T_2.

As is required by Assumption 2, we will assume that $\mathcal{L}_1$ and $\mathcal{F}_1$ are the first columns of $\mathcal{L}$ and $\mathcal{F}$, respectively. Constants c_1 and c_2 will be chosen so that the normalization required by Assumption 2 holds. As to Assumption 3, note that $AA' = T_1$ and $B'B = T_2$. The form of the limiting empirical distribution of eigenvalues of Toeplitz matrices, as their dimensionality grows, is well known (see, for example, Grenander and Szego, 1958). For the special case of the Toeplitz matrices T_1 and T_2, Assumption 3 is satisfied for $|\rho_1| < 1$ and $|\rho_2| < 1$.

To begin with, we calibrate our example (10) so that the generated data resemble those from Boivin et al. (2008). Boivin et al. (2008) perform a factor analysis of European quarterly macroeconomic time series. Even after extracting seven factors from that data, the residuals remain highly auto-correlated. We will set $\rho_2 = 0.9$ to represent strong serial correlation of such residuals. We will also introduce a mild degree of the cross-sectional correlation by setting $\rho_1 = 0.5$. Finally, we will set $n/T = 2$ so that the cross-sectional dimension is twice as large as the time series dimension. The actual data in Boivin et al. (2008) have $n = 245$ and $T = 111$.

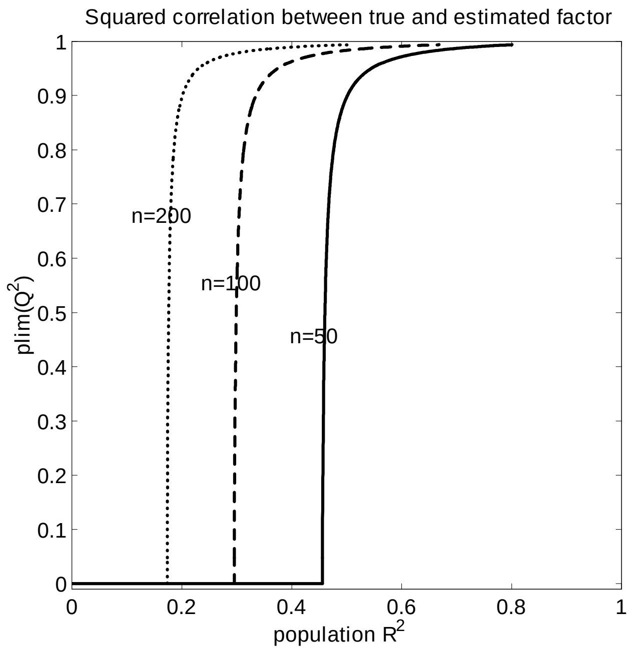


Figure 1: Plots of $p\lim Q^2$ against $\frac{d}{d+n}$ for $n = 50, 100$ and 200.

For such a calibrated example, we have computed the probability limits of Q^2, P^2 and $\hat{L}'\hat{L}$ described in Theorem 1 as functions of d.[7] Let us define the population R^2 of the factor as $R^2 \equiv \frac{E\,\mathrm{tr}\,LF'FL'/T}{E\,\mathrm{tr}\,XX'/T} = \frac{d}{d+n}$. The sample counterpart of this quantity is the actual R^2 from fitting 1-factor model to the data. It equals $\frac{\hat{L}'\hat{L}}{\mathrm{tr}\,XX'/T} \approx \frac{p\lim \hat{L}'\hat{L}}{p\lim \hat{L}'\hat{L}+n}$. We will call the latter approximation the sample R^2. Figure 1 shows the plots of $\mathrm{plim}\, Q^2$ versus the population $R^2 = \frac{d}{d+n}$ for $n = 50, 100$ and 200. Figure 2 shows the the plots of the sample R^2 versus the population $R^2 = \frac{d}{d+n}$ for $n = 50, 100$ and 200.

We see that the PC estimators of factors remain orthogonal to the true factors until the population R^2 becomes very large. The Pitman drift asymptotics, which we use in this paper to model weak factors, implies that as n tends to infinity, the population R^2 tends to zero for any fixed d. This fact explains the left shift of the plots in Figure 1 as n grows. For $n = 200$, even if a single factor explains 17% of variance in the data, the corresponding PC estimate will be orthogonal to it. For $n = 100$, a single factor must explain more than 30% of the data's variance before PC estimate becomes sensible. For $n = 50$, the explanatory power required to obtain sensible PC estimate becomes extremely large: 45% of the data's variance!

[7] The variance of ε_{it}, σ^2, equals 1.

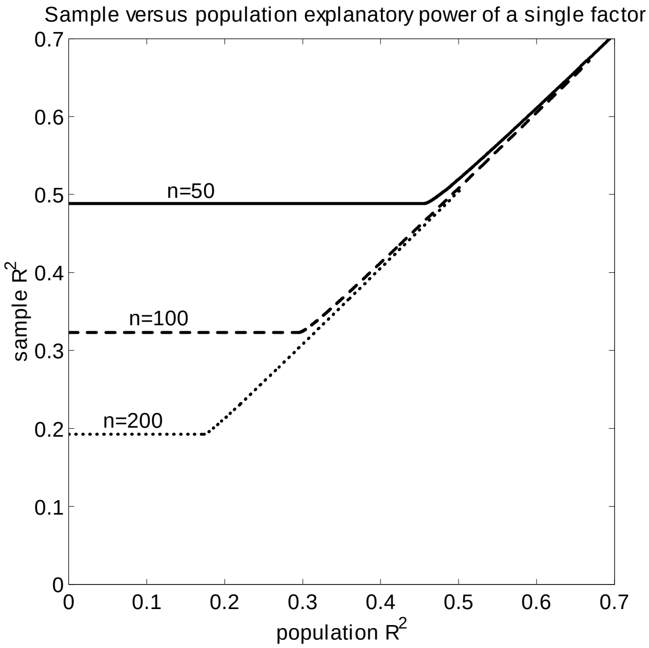


Figure 2: Plots of $\frac{p\lim \hat{L}'\hat{L}}{p\lim \hat{L}'\hat{L}+n}$ against $\frac{d}{d+n}$ for $n = 50, 100$ and 200.

To make a PC analyst even more miserable, Figure 2 reveals that the sample R^2 from fitting a single factor to the data may be very large even in cases when the factors are *de facto* very weak. According to our asymptotic analysis, in the extreme case when there are no factors in the data at all ($d = 0$), the sample R^2 would be around 0.19 for $n = 200$, around 0.32 for $n = 100$ and around 0.49 (!) for $n = 50$. These results accord well with Uhlig (2008) who is surprised to find a large R^2 (around 0.20) from fitting a 1-factor model to cross-sectionally independent (hence, there are no factors in the data) but temporally persistent data, which mimics Boivin et al. (2008).

The only positive news for the PC analysis with weak factors contained in Figures 1 and 2 is that, once the population R^2 goes above certain threshold, the quality of the PC estimator quickly improves. The correlation between the true and the estimated factor quickly approaches 1, and the sample R^2 becomes very well approximated by the population R^2.

For the idiosyncratic covariance structures which are different from (10), the probability limits of Q^2, P^2 and $\hat{L}'\hat{L}$ will vary. Theorem 1 explains how. Tables 1 and 2 illustrate the nature of this variation. The rows and columns of the tables correspond to different choices of the cross-sectional and temporal AR(1) coefficients ρ_1 and ρ_2 in (10), respectively.

Table 1: Threshold for the population R^2 (in percentage points) of a single factor, below which the PC estimates of factors and factor loadings are orthogonal to the true factors and true factor loadings. The population R^2 is measured by $d/(d+n)$, $n = 100$. Different choices of ρ_1 and ρ_2 are in the rows and columns, respectively.

$\rho_1 \backslash \rho_2$	0	0.2	0.4	0.6	0.8	0.9	0.95
0	1.4	2.0	3.6	6.8	15.0	27.7	44.2
0.2	1.7	2.3	3.9	7.1	15.3	27.9	44.4
0.4	2.6	3.1	4.8	8.0	16.3	28.8	45.0
0.6	4.3	4.9	6.6	9.9	18.3	30.6	46.4
0.8	9.1	9.7	11.4	14.9	23.1	34.8	49.5
0.9	17.1	17.7	19.4	22.7	30.4	40.9	53.9
0.95	29.4	29.8	31.3	34.2	40.6	49.3	59.9

Table 1 reports the threshold $\bar{w}\left(1-\bar{u}^{-1}\right)\left(1-\bar{v}^{-1}\right)/\left(\bar{w}\left(1-\bar{u}^{-1}\right)\left(1-\bar{v}^{-1}\right)+n\right)$ for the population R^2, below which the PC estimates of factors and factor loadings are orthogonal to the true factors and true factor loadings. Table 2 reports the asymptotic approximation $\bar{w}/\left(\bar{w}+n\right)$ to the sample R^2 from fitting a single factor to the data which have, in fact, no factors in them. For both Table 1 and Table 2, we set $n = 100$. We see that both the threshold for the population R^2 and the sample R^2 in the absence of factors quickly rise when the amount of the cross-sectional and temporal idiosyncratic correlation increase. This means that the PC estimator would be highly inaccurate even for very influential factors as long as the amount of the idiosyncratic correlation is relatively high.

Table 2: The sample R^2 (in percentage points) from fitting one factor when there are no factors. The sample R^2 is approximated by $\text{plim}\hat{L}'\hat{L}/(\text{plim}(\hat{L}'\hat{L})+n)$, $n = 100$. Different choices of ρ_1 and ρ_2 are in the rows and columns, respectively.

$\rho_1 \backslash \rho_2$	0	0.2	0.4	0.6	0.8	0.9	0.95
0	5.5	6.1	7.6	10.8	18.8	30.9	46.4
0.2	5.8	6.3	7.9	11.0	19.0	31.1	46.6
0.4	6.7	7.2	8.7	11.8	19.8	31.7	47.1
0.6	8.6	9.1	10.5	13.6	21.5	33.2	48.1
0.8	13.6	14.0	15.4	18.3	25.9	36.9	50.9
0.9	21.6	22.0	23.2	25.9	32.8	42.6	55.0
0.95	33.3	33.6	34.7	36.9	42.6	50.6	60.7

In the very special case when A and B are identity matrices so that the idiosyncratic terms lack cross-sectional and temporal correlation, the formulas of Theorem 1 considerably simplify. We obtain such a simplified formulas and extend our analysis to the second order asymptotics in the next section.

3 Second order asymptotics

In this section, we study the second order weak factors asymptotics. To establish the second order asymptotic results, we will use the following assumptions which are stronger than the corresponding assumptions in the previous section.

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

Assumption 1a i) strengthens Assumption 1 i) by requiring that the convergence $n/T^{(n)} \to c$ is faster than $n^{-1/2}$. Such a requirement eliminates any possible effects of this convergence on our second order asymptotic results. In our opinion, the 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 $L^{(n)\prime}L^{(n)}$ to D is made for the same reason.

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 second order 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 the rows of $F^{(n)}$, $F^{(n)}_{t\cdot}$ with $t = 1, ..., T^{(n)}$, 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.

Assumption 2a: i) $\varepsilon^{(n)}$ *is an* $n \times T^{(n)}$ *matrix with i.i.d.* $N(0, \sigma^2)$ *entries* ε_{it} *independent from* F *and* L,
ii) $A^{(n)} = I_n$,
iii) $B^{(n)} = I_T$.

This assumption is essential for our derivations of the second order asymptotics. It is too restrictive for our second order results to be of empirical relevance. However, the results are interesting from the theoretical point of view. Since matrices $A^{(n)}$ and $B^{(n)}$ that we consider in this section are trivial, we do not need any analog of Assumption 3. As in the previous section, 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 $\operatorname{span}(F)$ as its invariant subspace. Similarly, denote the matrix of the first q columns of $\hat{\mathcal{L}}$ as $\hat{\mathcal{L}}_{1:q}$ 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 conditional on L is invariant with respect to multiplication from the left by any orthogonal matrix having $\operatorname{span}(\mathcal{L})$ as its invariant subspace. We establish 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 1a and 2a hold and let, in addition,* $\phi_{ijst} = 0$ *when* $(i,j) \neq (s,t)$ *and* $(i,j) \neq (t,s)$. *Then,*

i) $T\hat{F}_{1:q} = F\tilde{Q} + F_q^{\perp}$ *with* $\tilde{Q} = \tilde{Q}^{(1)} + \frac{1}{\sqrt{T}}\tilde{Q}^{(2)}$, *where* $\tilde{Q}^{(1)}$ *is diagonal with* $\tilde{Q}^{(1)}_{ii} = \sqrt{\frac{d_i^2 - \sigma^4 c}{d_i(d_i+\sigma^2)}}$, *and* $\operatorname{vec}\tilde{Q}^{(2)}$ *is an asymptotically zero mean Gaussian vector with* $\operatorname{Acov}\left(\tilde{Q}^{(2)}_{ij}, \tilde{Q}^{(2)}_{st}\right)$ *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) $\hat{\mathcal{L}}_{1:q}{=}\mathcal{L}\tilde{P} + \mathcal{L}_q^{\perp}$ *with* $\tilde{P} = \tilde{P}^{(1)} + \frac{1}{\sqrt{T}}\tilde{P}^{(2)}$, *where* $\tilde{P}^{(1)}$ *is diagonal with* $\tilde{P}_{ii}^{(1)} = \sqrt{\frac{d_i^2-\sigma^4 c}{d_i(d_i+\sigma^2 c)}}$, *and* $\mathrm{vec}\,\tilde{P}^{(2)}$ *is an asymptotically zero mean Gaussian vector with* $\mathrm{Acov}\left(\tilde{P}_{ij}^{(2)}, \tilde{P}_{st}^{(2)}\right)$ *given by the following formulae:*

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

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

iii) *Matrix* $\hat{L}_{1:q}'\hat{L}_{1:q} = W^{(1)} + \frac{1}{\sqrt{T}}W^{(2)}$, *where* $W^{(1)}$ *is a diagonal matrix with* $W_{ii}^{(1)} = \frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2 c\right)}{d_i}$ *and* $\mathrm{vec}\,W^{(2)}$ *is an asymptotically zero mean Gaussian vector with* $\mathrm{Acov}\left(W_{ij}^{(2)}, W_{st}^{(2)}\right)$ *given by the following formulae:*

a) $\phi_{iiss}\frac{\left(d_i^2-\sigma^4 c\right)\left(d_s^2-\sigma^4 c\right)}{d_i d_s} + 2\delta_{is}\sigma^2\left(2d_i + c\sigma^2 + \sigma^2\right)\frac{d_i^2-\sigma^4 c}{d_i^2}$ *if* $i = j$ *and* $s = t$, *where* δ_{is} *denotes the Kronecker delta.*

b) 0 *otherwise.*

Theorem 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 the following theorem:

Theorem 3: *Suppose the assumptions of Theorem 2 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* $\bar{F}_{\tau_1\cdot}, ..., \bar{F}_{\tau_r\cdot}$. *Similarly, 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* $\left\{\hat{F}_{\tau_g i} - \tilde{Q}_{ii}^{(1)} F_{\tau_g i}: \ g = 1, ..., r; \ i = 1, ..., q\right\}$ *are asymptotically jointly mean-zero Gaussian. The asymptotic covariance between* $\hat{F}_{\tau_s i} - \tilde{Q}_{ii}^{(1)} F_{\tau_s i}$ *and* $\hat{F}_{\tau_f p} - \tilde{Q}_{pp}^{(1)} F_{\tau_f p}$ *equals*

$$\sum\nolimits_{s=1}^{k} \bar{F}_{\tau_g s} \bar{F}_{\tau_f s} \operatorname{Avar}\left(\tilde{Q}_{si}^{(2)}\right) + \left(\delta_{gf} - \sum\nolimits_{s=1}^{k} \bar{F}_{\tau_g s} \bar{F}_{\tau_f s}\right)\left(1 - \left(\tilde{Q}_{ii}^{(1)}\right)^2\right),$$

when $i = p$ *and*

$$-\bar{F}_{\tau_g p} \bar{F}_{\tau_f i} \operatorname{Acov}\left(\tilde{Q}_{pi}^{(2)}, \tilde{Q}_{ip}^{(2)}\right), \ \text{when } i \neq p.$$

ii) *Random variables* $\left\{\sqrt{T}\left(\hat{\mathcal{L}}_{j_g i} - \tilde{P}_{ii}^{(1)} \mathcal{L}_{j_g i}\right), \ g = 1, ..., r; \ i = 1, ..., q\right\}$ *are asymptotically jointly mean-zero Gaussian. The asymptotic covariance between* $\sqrt{T}\left(\hat{\mathcal{L}}_{j_g i} - \tilde{P}_{ii}^{(1)} \mathcal{L}_{j_g i}\right)$ *and* $\sqrt{T}\left(\hat{\mathcal{L}}_{j_f p} - \tilde{P}_{pp}^{(1)} \mathcal{L}_{j_f p}\right)$ *equals*

$$\sum\nolimits_{s=1}^{k} \bar{\mathcal{L}}_{j_g s} \bar{\mathcal{L}}_{j_f s} \operatorname{Avar}\left(\tilde{P}_{si}^{(2)}\right) + \left(\delta_{gf} - \sum\nolimits_{s=1}^{k} \bar{\mathcal{L}}_{j_g s} \bar{\mathcal{L}}_{j_f s}\right)\left(1 - \left(\tilde{P}_{ii}^{(1)}\right)^2\right),$$

when $i = p$ *and*

$$-\bar{\mathcal{L}}_{j_g p} \bar{\mathcal{L}}_{j_f i} \operatorname{Acov}\left(\tilde{P}_{pi}^{(2)}, \tilde{P}_{ip}^{(2)}\right), \ \text{when } i \neq p.$$

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. A similar interpretation holds for $\bar{\mathcal{L}}_{j_1 \cdot}, ..., \bar{\mathcal{L}}_{j_r \cdot}$.

Theorem 2 can be compared to Theorem 1 of Bai (2003). He finds that, under the strong-factor asymptotics, $\sqrt{n}\left(\hat{F}_{t\cdot} - H' F_{t\cdot}\right) \xrightarrow{d} N(0, \Omega)$, 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 or if c is close to 1. Note that the multiplier $\tilde{Q}_{ii}^{(1)}$, 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 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 2: *Suppose that, in addition to the assumptions of Theorem 2, the factors $F_{t\cdot}$ are i.i.d. standard multivariate normal random variables. Then, for any $i \leq q$:*

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

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

ii) $\hat{L}_{1:q}'\hat{L}_{1:q}=W^{(1)}+\frac{1}{\sqrt{T}}W^{(2)}$, *where* $W^{(1)}$ *is a diagonal matrix with* $W_{ii}^{(1)}=\frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2 c\right)}{d_i}$ and vec $W^{(2)}$ *is an asymptotically zero mean Gaussian vector with* Acov $\left(W_{ij}^{(2)},W_{st}^{(2)}\right)$ *given by the following simplified formulae:*

$$\begin{aligned}\text{Acov}\left(W_{ij}^{(2)},W_{st}^{(2)}\right) &= 2\left(d_i+\sigma^2\right)^2\left(1-\frac{\sigma^4 c}{d_i^2}\right)\text{ if } i=j=s=t \text{ and}\\ \text{Acov}\left(W_{ij}^{(2)},W_{st}^{(2)}\right) &= 0 \textit{ otherwise.}\end{aligned}$$

Note that when factors are i.i.d. Gaussian random variables, the PC 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. According to Theorem 13.5.1 of Anderson (1984), in such a case:

$$\sqrt{T}\left(\hat{\mathcal{L}}_{\cdot i}-\mathcal{L}_{\cdot i}\right)\rightarrow N\left(0,\Pi\right),\tag{11}$$

where

$$\Pi_{gf}=\sum_{\substack{s=1\\ s\neq i}}^{n}\mathcal{L}_{gs}\mathcal{L}_{fs}\frac{\left(d_i+\sigma^2\right)\left(d_s+\sigma^2\right)}{\left(d_i-d_s\right)^2}\tag{12}$$

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 (12) as

$$\Pi_{gf} = \sum_{\substack{s=1 \\ s\neq i}}^{k} \mathcal{L}_{gs}\mathcal{L}_{fs} \frac{(d_i + \sigma^2)(d_s + \sigma^2)}{(d_i - d_s)^2} + \left(\delta_{gf} - \sum_{s=1}^{k} \mathcal{L}_{gs}\mathcal{L}_{fs}\right) \frac{\sigma^2 (d_i + \sigma^2)}{d_i^2}. \tag{13}$$

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 (11) 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 2, we see that the probability limit of $\hat{\mathcal{L}}_{j_s i}$ equals $\mathcal{L}_{j_s i}$ multiplied by $0 \leq \tilde{P}_{ii}^{(1)} < 1$. 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}. Moreover, the bias of the PC estimator of factor loadings, as measured by the difference between matrix $\tilde{P}$ and the identity matrix, disappears as should be the case, intuitively.

Note that in a less special case when factors are not i.i.d but follow a Gaussian vector autoregression, we could have improved on the PC estimator by using maximum likelihood, based on Kalman filtering as in Doz et al. (2006). The authors of that paper use Monte Carlo experiments to show that the PC estimator is dominated by such a Kalman filtering procedure even if Assumption 2a is violated and the idiosyncratic terms are cross-sectionally and temporally correlated.[8] Although there is little hope that the quasi maximum likelihood would be consistent under weak factors asymptotics (recall that when factors are i.i.d. Gaussian, the PC estimator is the maximum likelihood and is inconsistent by Theorem 2), it is still possible that it

[8] In such a case, the procedure can be interpreted as quasi maximum likelihood (QML). Doz et al. (2006) establish the consistency of QML under strong factor asymptotics. For an early study of the properties of QML in the context of a dynamic exact factor model, see Doz and Lenglart (1999)

may reduce weak-factor-related bias of the PC, at least in some situations. We leave a study of such a possibility for future research.

4 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. First, we simulate data having 1-factor structure where $\{F_t, t = 1, ..., T\}$ is an AR(1) Gaussian process with AR coefficient 0.5 and variance 1 and where factor loadings $\{L_i, i = 1, ..., n\}$ are i.i.d. $\sqrt{d/n}N(0,1)$ random variables. For the idiosyncratic terms, we take the empirical distributions $F^{AA'}$ and $F^{B'B}$ as in our example (10) with $\rho_A = 0.5$ and $\rho_B = 0.9$ and we set ε_{it} be $i.i.dN(0,1)$ random variables.

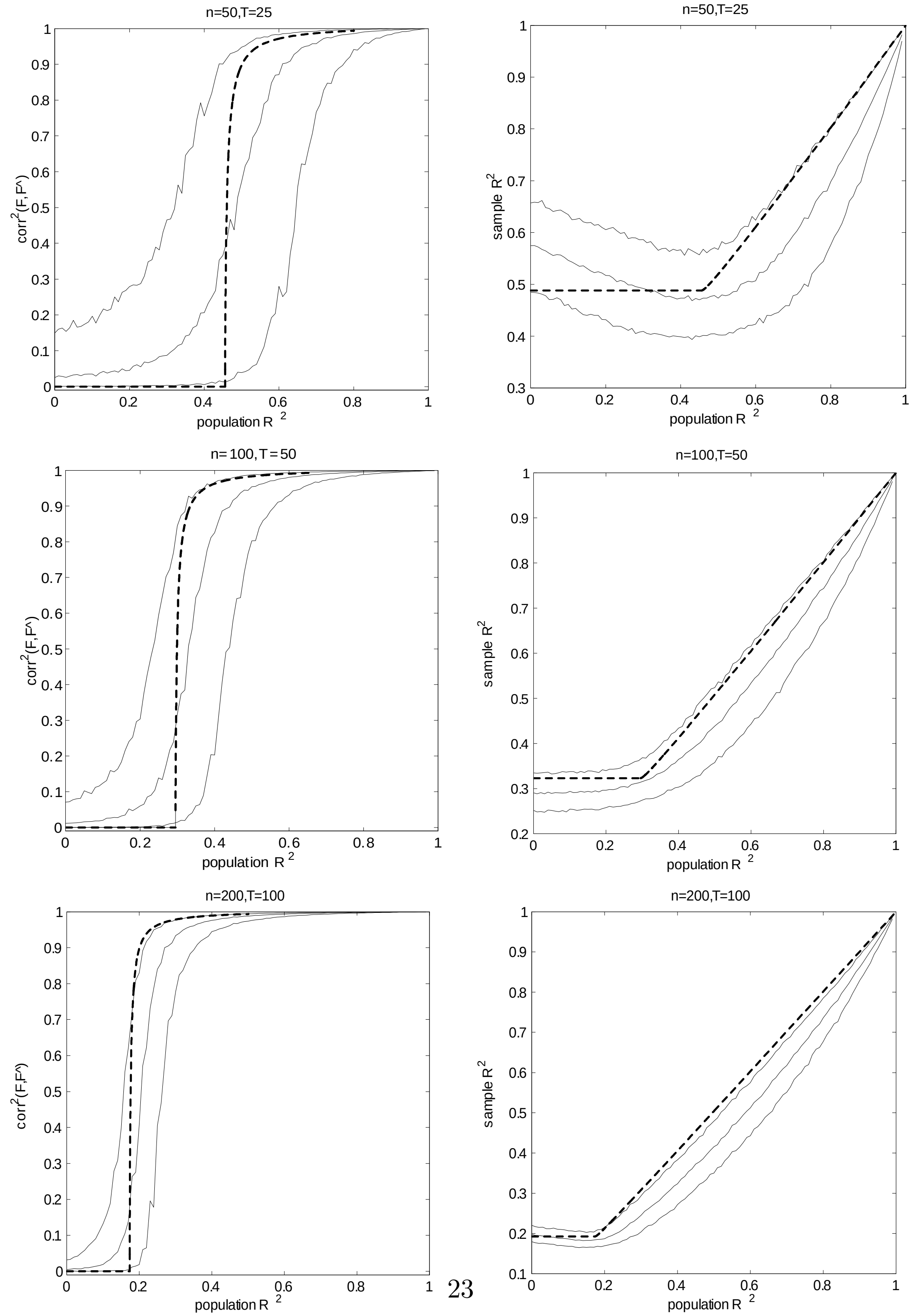
We simulate 1,000 replications of the data for $d/(d+n)$ on the grid 0:0.01:0.99. For each simulation, the matrices $\mathcal{F} \equiv [\mathcal{F}_1, ..., \mathcal{F}_T]$ and $\mathcal{L} \equiv [\mathcal{L}_1, ..., \mathcal{L}_T]$ used to define the idiosyncratic terms (see Assumption 2) are chosen randomly from the set of all matrices which complement $\mathcal{F}_1 \equiv (F'F)^{-1/2} F$ and $\mathcal{L}_1 \equiv (L'L)^{-1/2} L$ to the orthonormal basis.

The left panel of Figure 3 shows the theoretical probability limits from Theorem 1 (dashed lines) and the Monte Carlo medians and 10% and 90% percentiles (solid lines) of the square of the regression coefficient Q in the regression of $\hat{F}$ on F as functions of the population R^2, which equals $d/(d+n)$. The right panel of Figure 3 shows the theoretical probability limits and the Monte Carlo medians and 10% and 90% percentiles of the sample R^2 as functions of the population R^2.

We see that, overall, the theoretical limits do a good job of approximating the corresponding finite sample relationships. For very small n and T, the Monte Carlo distributions are very dispersed. They become less dispersed and the theoretical approximations become more and more informative as n and T rise. The theoretical limits for the squared correlation coefficient between the true and the estimated factor are under-predicting for relatively small values of the population R^2 and over-predicting for relatively large values of the population R^2. The degree of the over- and under-prediction diminishes as n and T rise.

We repeated the above MC experiment with $\rho_A = 0.5$ and $\rho_B = 0.5$ (instead of $\rho_A = 0.5$ and $\rho_B = 0.9$). Qualitatively, the results remain the same although the MC distribution becomes less dispersed for the same values of n and T. Next,

Figure 3: Square of the correlation coefficient between true and estimated factor (left panel) and the sample R^2 from fitting a single factor (right panel). Theoretical values: dashed line, 10, 50, and 90 MC percentiles: solid lines. Upper panel: $n = 50, T = 25$, middle panel: $n = 100, T = 50$, lower panel: $n = 200, T = 100$.



we again set $\rho_A = 0.5$ and $\rho_B = 0.9$, but consider a relatively fat-tailed and a skewed distribution for ε_{it}. Precisely, we consider Student's t distribution with five degrees of freedom, normalized to have unit variance, and the centered chi-squared distribution with one degree of freedom, normalized to have unit variance. For such a non-normal distributions, the results are very similar to those shown in Figure 3, and we do not report them to save space. Finally, we have repeated our MC experiment with $\rho_A = 0.5$ and $\rho_B = 0.9$, but without respecting Assumption 2 in that the matrices $\mathcal{F} \equiv [\mathcal{F}_1, ..., \mathcal{F}_T]$ and $\mathcal{L} \equiv [\mathcal{L}_1, ..., \mathcal{L}_T]$ used to define the idiosyncratic terms in Assumption 2 are chosen independent from F and L. Again, the results obtained were very similar to those reported in Figure 3 and we do not report them here.

To assess the finite sample quality of our second-order asymptotic results, we perform three different MC 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 2. According to that theorem, the regression coefficient should be equal to $\tilde{Q}^{(1)} + \frac{1}{\sqrt{T}}\tilde{Q}^{(2)}$. Note that the theoretical lines do not start from $d = 0.1$. It is because our second order formulae are valid for d larger than the threshold $\bar{w}\left(1 - \bar{u}^{-1}\right)\left(1 - \bar{v}^{-1}\right)$, 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. $N\left(0, 1\right)$, $\sigma^2 = 1$, 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

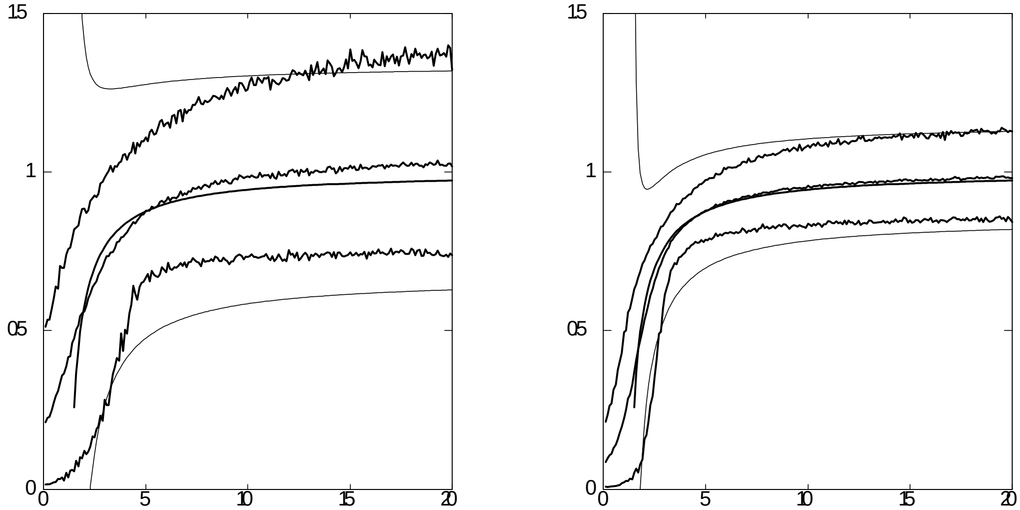


Figure 3: Monte Carlo and theoretical means and 5% and 95% quantiles of the regression coefficient in the regression of $\hat{F}$ on F as functions of d. Horizontal axis: d. Left panel: $n = 40, T = 20$; right panel: $n = 200, T = 100$.

the 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 2), the dashed lines correspond to the 95% confidence ellipses of the classical asymptotic distribution (see equation 13), 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

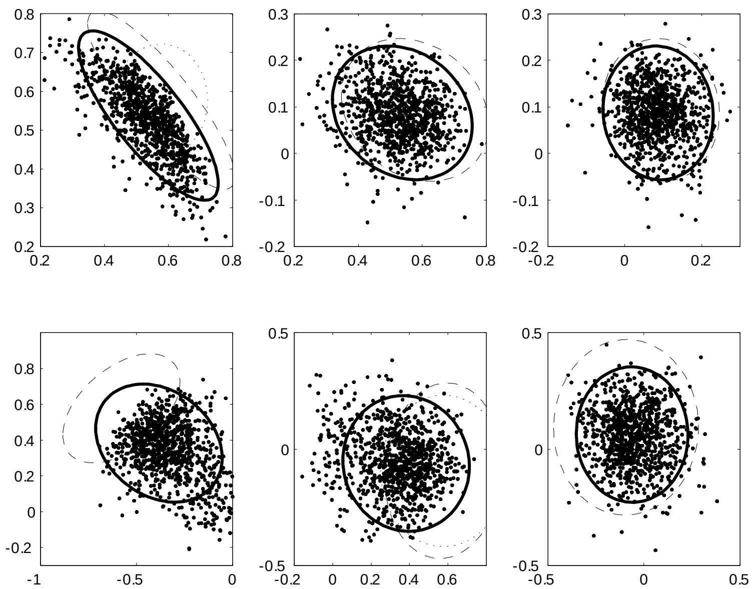


Figure 4: Monte Carlo draws and 95% asymptotic confidence ellipsoids for (from left to right) (1st, 2nd), (2nd, 3rd), (3rd, 4th) components of the normalized vectors of factor loadings. Upper panel: loadings correspond to the first factor. Lower panel: loadings correspond to the second factor. Solid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

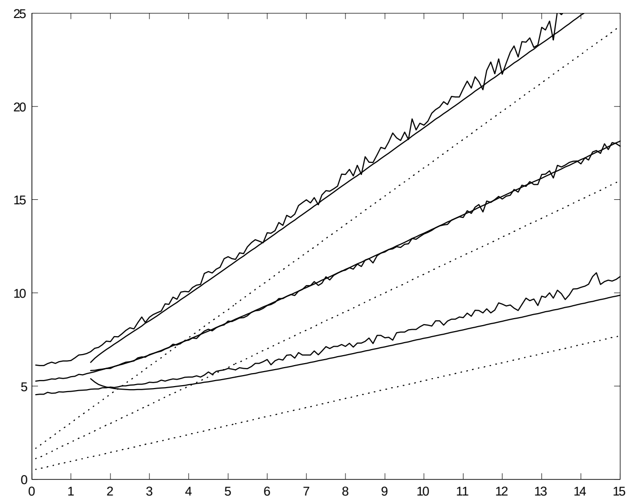


Figure 5: Monte Carlo and asymptotic means and 5% and 95% quantiles of the eigenvalue distribution. Smooth solid lines: our asymptotics. Dotted lines: classical asymptotics. Horizontal axis: the cumulative effect d of the factor. $n = 40$, $T = 20$.

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 than the classical distribution.

5 Conclusion

In this paper we have introduced a weak factors asymptotics framework which allows us to assess the finite sample properties of the PC estimator in the situation when factors' explanatory power does not strongly dominates the explanatory power of the cross-sectionally and temporally correlated idiosyncratic terms. We have shown that the principal components estimators of factors and factor loadings are inconsistent and found explicit formulae for the amount of the inconsistency. For the special case when there are no cross-sectional and temporal correlation in the idiosyncratic terms, we have shown that the PC estimators, although inconsistent, are asymptotically normal, and we have found explicit formulae for the asymptotic covariance matrix of the estimators. Our Monte Carlo analysis suggests that our asymptotic formulae work well even for relatively small n and T.

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, J. and S. Ng (2005) "Determining the Number of Factors in Approximate Factor Models, Errata", mimeo.

[5] Bai, J. and S. Ng (2008) "Large Dimensional Factor Analysis", Foundations and Trends in Econometrics: Vol. 3: No 2, pp 89-163.

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

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

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

[9] Davidson, R. and J. G. MacKinnon (2004) Econometric Theory and Methods, Oxford University Press, New York, Oxford.

[10] DeMol, C., Giannone, D. and L. Reichlin (2008) "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?", Journal of Econometrics 146, pp. 318-328

[11] Doz, C., Giannone, D. and L. Reichlin (2006) "A quasi maximum likelihood approach for large approximate dynamic factor models", manuscript, Universite Cergy-Pontoise

[12] Doz, C. and F. Lenglart (1999) "Analyse factorielle dynamique: test du nombre de facteurs, estimation et application à l'enquête de conjoncture dans l'industrie", Annales d'Économie et de Statistique 54, pp. 91-127.

[13] Forni, M. and Lippi, M. (1999) "Aggregation of linear dynamic microeconomic models", Journal of Mathematical Economics 31, pp. 131-158

[14] Forni, M. and Lippi, M. (2001) "The generalized dynamic factor model: representation theory", Econometric Theory 17, pp. 1113-1141

[15] Grenander, U. and G. Szerö (1958) Toeplitz Forms and Their Applications. University of California Press, Berkeley.

[16] Harding, M. (2006) "Structural estimation of high-dimensional factor models", unpublished manuscript, Stanford university.

[17] Heaton, C. and Solo, V. (2006) "Estimation of approximate factor models : is it important to have a large number of variables?" Macquarie economics research papers, Issue 5, p.1-31

[18] Johnstone, I.M., and A. Y. Lu (2007) "Sparse principal components analysis", Journal of the American Statistical Association. To appear.

[19] Onatski, A. (2005) "Determining the number of factors from the empirical distribution of eigenvalues", forthcoming in the Review of Economics and Statistics.

[20] Paul, D. (2007) “Asymptotics of sample eigenstructure for a large dimensional spiked covariance model”, Statistica Sinica 17, pp. 1617-1642

[21] Paul, D., and Silverstein, J. (2008) “No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix”, working paper, North Carolina State University.

[22] Stock, J. (1994) Unit Roots, Structural Breaks and Trends, in Handbook of Econometrics, vol.4, eds. Engle, R.F. and D.L. McFadden, pp.2740-2831.

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

[24] Van Loan, C. F. and Pitsianis, N. P. (1993) “Approximation with Kronecker products”, in M. S. Moonen and G. H. Golub, editors, Linear Algebra for Large Scale and Real Time Applications, Kluwer Publications, pp. 293–314.

[25] Zhang, L. (2006). Spectral Analysis of Large Dimensional Random Matrices. Ph. D. Thesis. National University of Singapore.