Spurious Factor Analysis

Alexei Onatski* and Chen Wang†

*Faculty of Economics, University of Cambridge.

†Department of Statistics and Actuarial Science,
University of Hong Kong.

September 14, 2018

Abstract

This paper draws parallels between the Principal Components Analysis of factorless high-dimensional nonstationary data and the classical spurious regression. We show that a few of the princpal components of such data absorbs nearly all the data variation. The corresponding scree plot suggests that the data contains a few factors, which is collaborated by the standard panel information criteria. Furthermore, the Dickey-Fuller tests of the unit root hypothesis applied to the estimated 'idiosyncratic terms' often reject, creating an impression that a few factors are responsible for most of the nonstationarity in the data. We warn empirical researchers of these peculiar effects and suggest to always compare the analysis in levels with that in differences.

KEY WORDS: Spurious regression, principal components, factor models, Karhunen-Loève expansion.

1 Introduction

Researchers applying factor analysis to nonstationary macroeconomic panels face a choice: keep the data in levels or first-difference them. If all the nonstationarity is due to the factors, no differencing is necessary. A simple principal components estimator of the factors is consistent (e.g. Bai, 2004). Otherwise, the standard advice is to extract the factors from the first-differenced data, and then, accumulate them to obtain estimates of the factors in levels (e.g. Bai and Ng, 2004).

Even though risk-averse econometricians would probably prefer to difference the data, both strategies are used in practice. For a recent example, Engel et al. (2015) estimate factors in a panel of log exchange rates without the differencing. They assume "that the factors component soaks up a common unit root component" in the data. Banerjee et al. (2017) give several reasons for such an assumption in macroeconomic applications. In particular, they cite a very high rejection rate of the hypothesis of a unit root in the estimated idiosyncratic components of the 114 nonstationary monthly US macroeconomic series for the 1959-2014 period (see McCracken and Ng, 2015). In contrast, Barigozzi et al. (2018) claim that "the assumption of I(0) idiosyncratic components is empirically not supported by typical macroeconomic datasets". Moreover, they point out that this assumption implies too much cointegration between macroeconomic variables.

This paper is intended as a warning to the empirical researches tempted by the arguments advocating factor estimation in levels. We show theoretically that a few principal components of a *factorless* nonstationary panel must 'explain' an extremely high portion of the data variation. Moreover, the Dickey-Fuller tests on the estimated idiosyncratic terms are strongly oversized, supporting the stationarity hypothesis where, in fact, the null of nonstationarity is true.

We are not the first to point out the high explanatory power of a few of the principal components of factorless persistent data. Uhlig (2009), discussing Boivin et al. (2009), generates artificial cross-sectionally *independent* AR(1) data with the autoregressive coefficients matching the first-order autocorrelations of the 243 macroeconomic series used in Boivin et al. (2009). Then he plots the fraction of variation explained against the number of factors for both actual and artificial data (see Figure 1), and noted that the two plots "look surprisingly and uncomfortably alike". In particular, five estimated factors explain about 75% of the actual data variation, but at the same time, five estimated factors, that must be spurious by construction, 'explain' about 60% of the simulated data variation.

Uhlig (2009) attributes the high explanatory power of the spurious factors to the fact that the simulated data are considerably autocorrelated. Many of the simulated series' first-order autocorrelation coefficients are close to unity. In a finite sample (in his setting, 83 observations), the series may appear to be correlated, which will be picked up by the principal components. Although this explanation is intuitive, Uhlig admits that it is "perhaps tricky to formalize".

In this paper, we do such a formalization. Similarly to the spurious regression framework, we set our theoretical analysis in the context of a panel of cross-

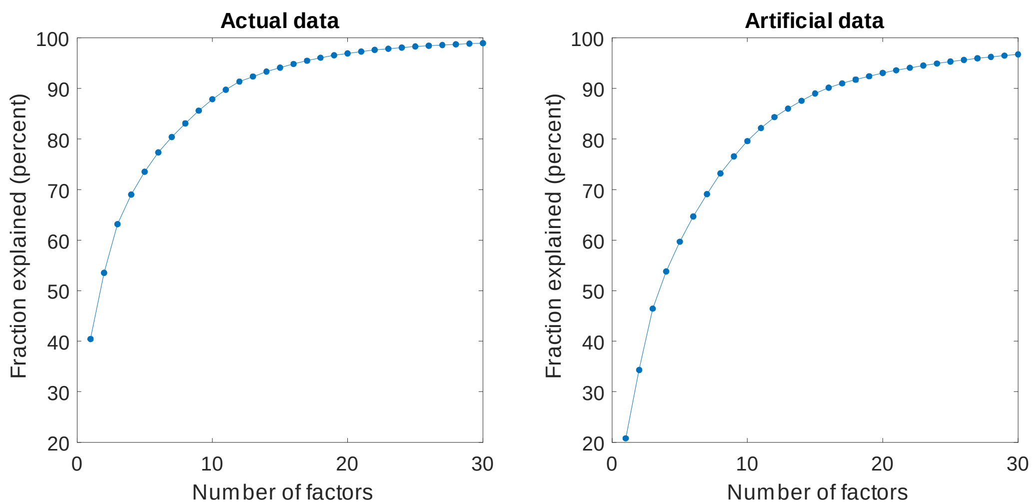

Figure 1: Factor contribution to the overall variance. Left panel: actual Boivin et al.'s (2009) data. Right panel: factorless simulated data with similar autocorrelation properties.

sectionally *independent* difference stationary processes. The differenced processes are assumed to have continuously differentiable spectral densities which are positive at zero frequency. The cross-sectional and temporal dimensions of the panel are assumed to increase to infinity at arbitrary rates.

We prove that under such an asymptotic regime the fraction of the data variation explained by the first principal component converges in probability to $6/\pi^2 \approx 0.61$ even though the data are cross-sectionally independent. The first three principal components together asymptotically explain $100\% \sum_{j=1}^{3} 6/(j\pi)^2 \approx 83\%$ of the variation in the *factorless* nonstationary data.

We show from a theoretical standpoint that, in our setting, the standard panel information criteria (e.g. Bai, 2004) are very sensitive to the choice of the *a priori* maximum number of factors. For empirically relevant choices and data sizes, the criteria will often detect two or three 'factors'. We provide Monte Carlo evidence supporting this claim.

The peculiar results of the principal component analysis of cross-sectionally independent nonstationary data can be linked to the Karhunen-Loève expansion of the demeaned Wiener process (e.g. Shorack and Wellner, 1986). Such a process approximates each cross-sectional component of our demeaned nonstationary data asymptotically.

The expansion represents the process in the form of an infinite sum of trigonometric functions with uncorrelated random coefficients whose variances are quickly decaying. Hence, much of the variation in a nonstationary panel is captured by a few of the trigonometric functions corresponding to the first terms in the Karhunen-Loève expansion. In an unusual sense, cross-sectionally independent nonstationary data do have a factor structure with the trigonometric functions as factors and the random coefficients in the Karhunen-Loève expansion as loadings.

Phillips (1998) points out that the "prototypical spurious regressions, in which unit root nonstationary time series are regressed on deterministic functions," reproduces the underlying Karhunen-Loève representation of the Wiener process. In the similar spirit, the spurious factor analysis, i.e. the principal components analysis of the cross-sectionally independent difference-stationary data, picks up the common Karhunen-Loève structure of the cross-sectional units.[1]

This intuition immediately suggests that the Dickey-Fuller tests of the hypothesis of a unit root in the estimated spurious idiosyncratic terms must be oversized. Indeed, when researchers apply the test to an estimated idiosyncratic term, they ignore the fact that the estimate is, essentially, the residual from a regression of a nonstationary series on a few slowly varying trigonometric functions.

These functions are similar to the deterministic polynomial trends. Hence, the intercept-only Dickey-Fuller statistic computed on the basis of estimated idiosyncratic terms asymptotically behaves similarly to the intercept-only Dickey-Fuller statistic for the regression that includes several deterministic polynomial time trends. This leads to a substantial size distortion and a potential confused conclusion that the factors soak up all or most of the nonstationarity in the data.

One may wonder whether the principal components method delivers consistent estimates of the *genuine* pervasive factors contaminated by nonstationary factorless noise. Unfortunately, as we show in Section 3.3, the answer is negative. In general, the principal components estimator will be inconsistent. Such an inconsistency was pointed out informally in the previous literature (e.g. Bai and Ng (2004, p. 1143)). We now confirm this by providing a formal proof.

All in all, the results of the principal components analysis of the levels of nonstationary data may be very misleading. We recommend to always compare the first differences of factors estimated from the levels with factors estimated from the first-differenced data. A mismatch indicates a spurious factor analysis in levels. Whether the estimation in differences is always preferable to the estimation

[1] The notion of spurious factors considered in this paper is not directly related to the spurious factors in asset returns that received much recent research attention (see Bryzgalova (2018) and references therein).

in levels is a separate question which we leave to future research.

The remainder of the paper is structured as follows. In Section 2, we formally introduce our setting and present our main results. Section 3 analyzes the estimated spurious factors and studies implications for the Dickey-Fuller tests applied to estimated idiosyncratic terms. Section 4 discusses ways to detect spurious results and concludes. Technical proofs are given in the Appendix.

2 Setting and main results

Consider an n-dimensional difference stationary process

$$X_t = X_0 + \sum_{s=1}^{t} \varepsilon_s, \ t \in \mathbb{N}, \tag{1}$$

where X_0 is a vector of initial values and ε_s is a vector process with components ε_{js} given by *independent* linear processes

$$\varepsilon_{js} = \sum_{k=0}^{\infty} \theta_{jk}\xi_{j,s-k}, \ s \in \mathbb{N}. \tag{2}$$

Properties of θ_{jk} and $\xi_{j,s-k}$ are specified in assumptions A1 and A2 below.

Denote the $n \times T$ matrices with t-th columns X_t and ε_t as X and ε, respectively. Then

$$X = X_0 l' + \varepsilon U,$$

where U is the T-dimensional upper triangular matrix of ones and l is the T-dimensional vector of ones. We interpret X as the data. Our goal is to study the workings of the principal components analysis (PCA) of these data as both n and T go to infinity, without any constraints on the relative speed of growth.

In contemporary economic applications, the PCA is often used to estimate factors F and loadings Λ in the factor model for the temporarily demeaned data

$$X - \bar{X} = \Lambda F' + e. \tag{3}$$

See Stock and Watson (2016) for a review of the related literature. The common factors are often interpreted as a few important latent variables affecting a vast number of economic indicators (rows of X). Of course, the above setting is designed to invalidate any such interpretation because the rows of X are independent by construction.

Suppose, however, that a researcher considers representation (3) with r factors, so that F is an unknown $T \times r$ matrix. If it is normalized so that $F'F = I_r$, then its PCA estimator is defined as the matrix $\left[\hat{F}_1, ..., \hat{F}_r\right]$ of the r principal eigenvectors of

$$S \equiv \frac{1}{n}\left(X - \bar{X}\right)'\left(X - \bar{X}\right) = \frac{1}{n} MU'\varepsilon'\varepsilon UM, \tag{4}$$

where M is the projector matrix on the space orthogonal to l. The corresponding principal eigenvalues $\lambda_1 \geq ... \geq \lambda_r$ estimate the explanatory power of the factors. Precisely, $\lambda_j / \operatorname{tr} S$ is interpreted as the fraction of the data variation explained by the j-th factor.

To study the asymptotic behavior of λ_j and $\hat{F}_j$, we make the following assumptions.

Assumption A1. *Random variables ξ_{jt} with $j \in \mathbb{N}$ and $t \in \mathbb{Z}$ are independent and such that, for all $j \in \mathbb{N}$ and $t \in \mathbb{Z}$, $E\xi_{jt} = 0$, $E\xi_{jt}^2 = 1$, and $\sup_{j\in\mathbb{N}} \left|E\xi_{jt}^4 - 3\right| \leq \kappa_4 < \infty$.*

Assumption A2. *The linear filters in (2) are one-summable uniformly in $j \in N$, i.e. there exists $B < \infty$* s.t. $\sup_{j\in\mathbb{N}} \sum_{k=0}^{\infty} (1+k)\left|\theta_{jk}\right| \leq B$. *Furthermore, there exists $b > 0$ s.t. $\inf_{j\in\mathbb{N}} \left|\sum_{k=0}^{\infty} \theta_{jk}\right| \geq b$.*

In our derivations below, we rely on the existence of the variance of the entries of the sample covariance matrix. This requires the finite fourth moment of ξ_{jt} and hence, the finite excess kurtosis, as stated in A1. The one-summability assumption in A2 guarantees the existence of continuously differentiable spectral densities

$$f_j(\omega) \equiv \frac{1}{2\pi}\left|\sum_{k=0}^{\infty} \theta_{jk} e^{ik\omega}\right|^2 \tag{5}$$

for the processes ε_{js}, $s \in \mathbb{N}$. The assumption yields explicit bounds on the spectral density and its derivative. Clearly,

$$\max_{\omega} \left|f_j(\omega)\right| \leq B^2/(2\pi). \tag{6}$$

Furthermore, differentiating both sides of (5) with respect to ω and using A2, we obtain

$$\max_{\omega} \left|f_j'(\omega)\right| \leq B^2/\pi. \tag{7}$$

The part of A2 that requires uniform boundedness from below of $\left|\sum_{k=0}^{\infty} \theta_{jk}\right|$

implies that

$$\inf_{j\in\mathbb{N}} |f_j(0)| \geq b^2/(2\pi). \tag{8}$$

We need this assumption to guarantee that the processes ε_{js}, $s \in \mathbb{N}$, do not have MA roots close to unity, which would dramatically decrease the amount of low frequency variation of the entries of X_t.

Theorem 1 *Let "$\xrightarrow{P}$" denote convergence in probability. Under assumptions A1 and A2, for any fixed positive integer k, as $n, T \to \infty$ at arbitrary rates,*
(i) $\lambda_k/(\gamma_n T^2) \xrightarrow{P} (k\pi)^{-2}$, *where* $\gamma_n = 2\pi n^{-1}\sum_{j=1}^{n} f_j(0)$.
(ii) $\lambda_k/\operatorname{tr} S \xrightarrow{P} 6/(k\pi)^2$, *where S is defined in (4).*
(iii) $\left|\hat{F}_k' v_k\right| \xrightarrow{P} 1$, *where* $v_k = (v_{k1}, ..., v_{kT})'$ *with* $v_{kt} = -\sqrt{2/T}\cos[\pi k(t-1/2)/T]$.

A proof of this theorem is given in the Appendix. Figure 2 illustrates statement (i) by showing the asymptotic scree plot for factorless persistent data. The height of the plot is scaled so that the largest eigenvalue equals one. We find the plot uncomfortably similar to those reported in the empirical literature on large factor models.

A typical interpretation of such a plot would be that the data "obviously" contain at least one strong factor, but perhaps two, or even three of them. Theorem 1 (i) shows that such an interpretation is misleading because the plot corresponds to the cross-sectionally independent data, not influenced by any common shocks whatsoever.

Part (ii) of the theorem describes the portion of data variation attributed to the k-th principal component. A naive but standard interpretation of this result would be that the first k factors explain $\sum_{j=1}^{k} 6/(j\pi)^2 \times 100\%$ of the variation in the data. This "explanatory power" is amazingly strong. The first three spurious factors absorb more than 80% of the data variation.

Figure 3 plots the fraction of the data variation "explained" relative to the number of "factors" used. The graph looks qualitatively similar to that constructed by Uhlig (2009). The spurious explanatory power shown in Figure 3 is larger than that in Figure 1. This difference can be attributed to the fact that we study the extreme case of nonstationary data with unit roots, whereas Uhlig (2009) simulates persistent, but stationary data.

Part (iii) of Theorem 1 reveals that the factor estimates converge in probability to deterministic cosine functions in the sense that the angle between the vector of estimates and the vector of uniform grid values of the corresponding cosine function

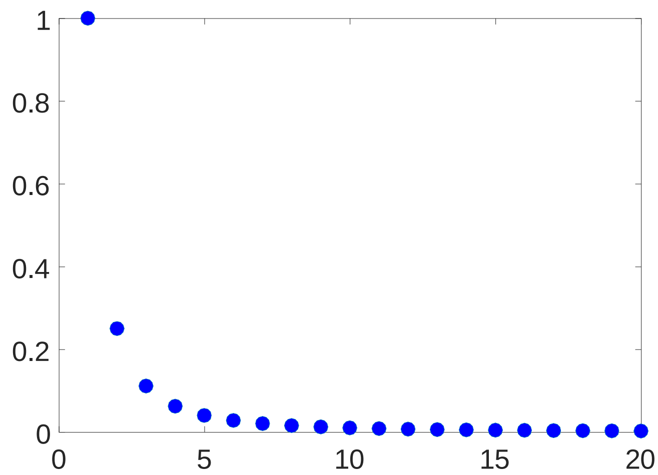

Figure 2: The asymptotic scree plot for factorless persistent data (the first 20 normalized eigenvalues only). The horizontal axis shows the order k of the eigenvalue λ_k. The vertical axis shows the probability limit of λ_k/λ_1.

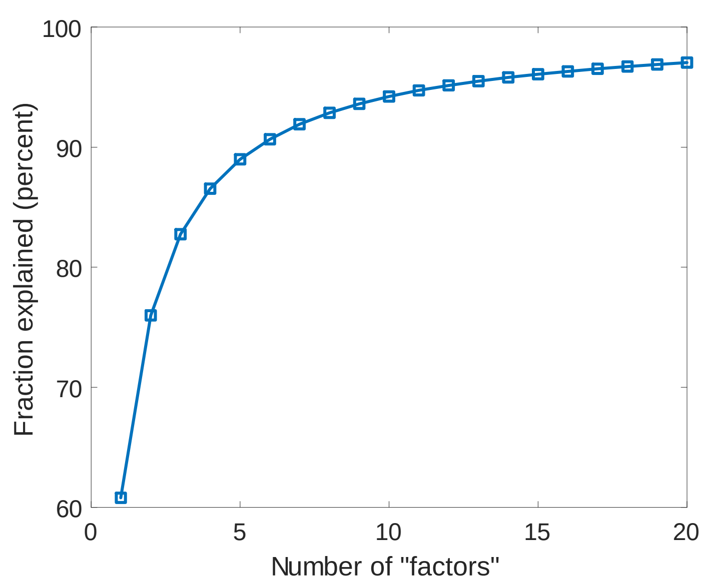

Figure 3: Fraction of the variation in factorless persistent data "explained" by the first several "factors".

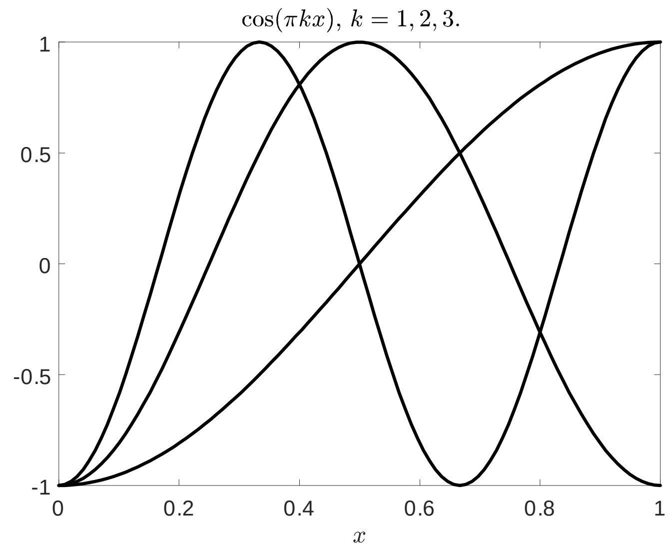

Figure 4: The probability limits of the first three spurious factor estimates.

converges in probability to zero. Figure 4 plots the cosine functions corresponding to the first three "factors". They may be interpreted as the trigonometric versions of the linear, quadratic, and cubic trends.

As mentioned in the introduction, the functions can be linked to the Karhunen-Loève expansion of the demeaned Wiener process. Indeed, let

$$\tilde{W}(x) = W(x) - \int_0^1 W(x)\,\mathrm{d}x,$$

where $W(x)$ is the usual Wiener process on $[0,1]$. As is well known (e.g. Müller and Watson (2008, Thm. 1)), its covariance kernel has eigenfunctions $\sqrt{2}\cos(\pi k x)$, $k = 1, 2, ...$, corresponding to eigenvalues $(\pi k)^{-2}$. Therefore, the Karhunen-Loève expansion of $\tilde{W}(x)$ has the following form

$$\tilde{W}(x) = \sqrt{2}\sum_{k=1}^{\infty} (\pi k)^{-1} \cos(\pi k x)\, z_k, \tag{9}$$

where z_k are i.i.d. standard normal random variables.

Recall that each of the series X_{jt} in our data is a difference stationary process, and the principal component analysis is performed on the demeaned data. As is well-known, functions $Y_{jT}(x) = (f_j(0)\,T)^{-1/2} X_{j[xT]}$ weakly converge in the $D[0,1]$

space to $W(x)$ (e.g. Phillips, 1986) and thus,

$$Y_{jT}(x) - \bar{Y}_{jT} = (f_j(0)\,T)^{-1/2}\left(X_{j[xT]} - \bar{X}_j\right)$$

weakly converge to $\tilde{W}(x)$. Therefore, each of the demeaned standardized series $(f_j(0)\,T)^{-1/2}\left(X_{jt} - \bar{X}_j\right)$ can asymptotically be represented by the Karhunen-Loève expansion of $\tilde{W}(x)$. In particular, functions $\cos(\pi kt/T)$ with $k = 1, 2, ...$ capture much of the variation in each of $X_{jt} - \bar{X}_j$.

Incidentally, in a series of papers, Müller and Watson (2008, 2016, 2017, and 2018) advocate using time series projections on a few of such functions as a powerful tool for answering "low-frequency questions" often arising in economics. Theorem 1 (iii) suggests that the PCA analysis of large nonstationary panels would be akin to extracting such projections. The estimated "loadings" would correspond to the projection coefficients that encode much information about series' low-frequency variability and covariability.

Since the majority of the variation in nonstationary data comes from low frequencies, a few principal components would summarize the data reasonably well. For example, Müller and Watson (2008, p. 980) report that projecting on $\cos(\pi kt/T)$ with $k = 1, ..., 13$ "almost completely capture[s] the lower than business cycle variability in postwar macroeconomic time series..." Such a situation would lead to scree plots similar to that on Figure 2, which, in their turn, may lead to confused conclusions about the existence of a few common factors or shocks driving the data dynamics.

Although independent nonstationary unit root data do not contain factors in the usual sense, they admit an unusual factor-like structure. The "factors" are deterministic functions that summarize variation of nonstationary processes in principal directions, and their loadings are random variables with variance proportional to the long-run variance of the corresponding series. From a technical viewpoint, an important difference with the standard factor structure is that the number of the eigenvalues of the sample covariance matrix that go to infinity as $n, T \to \infty$ is not finite. This directly follows from Theorem 1 (i).

As a consequence, the "factor" and the "idiosyncratic" eigenvalues of the sample covariance matrix do not separate, even asymptotically. This makes theoretical analysis of the eigenvalue and eigenvector behavior not as straightforward as in the usual case of the asymptotic separation. In recent studies, Koltchinskii and Lounichi (2016, 2017) develop new methods for such a theoretical analysis. One of their key findings is that the accuracy of the match between high-dimensional sample, $\hat{\Sigma}$, and the population, Σ, covariance matrices crucially depends on the so-called effective rank of Σ, defined as $r(\Sigma) = \operatorname{tr}\Sigma / \|\Sigma\|$, where $\|\Sigma\|$ denotes the spectral norm. The lower $r(\Sigma)$, the better the match.

Separation of the eigenvalues into a few exploding and remaining bounded ones is not the only way to make $r(\Sigma)$ small. Having a fast decaying eigenvalues, as in the case of the (conditional) autocovariance matrix of random walk, is another way. Koltchinskii and Lounichi's methods are derived for the i.i.d. Gaussian data. In this paper, although we assume that the data are cross-sectionally independent, the autocovariance matrices of individual series do not coincide. Furthermore, we do not assume Gaussianity. This calls for different methods which we develop in the proof of Theorem 1.

3 Spurious factors

In this section, we answer three questions. First, what is the number of "factors" in factorless persistent data detected by information criteria proposed in Bai (2004)? Second, how oversized are the standard Dickey-Fuller tests of unit root in the "idiosyncratic" component of the factorless persistent data? Third, are the true pervasive factors in the data with persistent idiosyncratic terms estimated consistently by the principal components without the first differencing?

3.1 The number of "factors"

Bai (2004) proposes to estimate the number of factors in nonstationary panels by minimizing one of the following functions:

$$IPC_1(k) = V(k) + k\hat{\sigma}^2\alpha_T \frac{n+T}{nT}\log\frac{nT}{n+T}, \text{ or}$$

$$IPC_2(k) = V(k) + k\hat{\sigma}^2\alpha_T \frac{n+T}{nT}\log K_{nT}, \text{ or}$$

$$IPC_3(k) = V(k) + k\hat{\sigma}^2\alpha_T \frac{n+T-k}{nT}\log nT$$

over $k = 0, 1, ..., k_{\max}$, where

$$V(k) = \frac{1}{T}\operatorname{tr} S - \frac{1}{T}\sum_{j=1}^{k}\lambda_j,$$

$\hat{\sigma}^2 = V(k_{\max})$, $\alpha_T = T/(4 \log\log T)$, and $K_{nT} = \min\{n, T\}$.

Let us denote the value k that delivers the minimum of $IPC_j(k)$ as $\hat{k}_j$. Bai's (2004) Theorem 1 gives conditions under which $\hat{k}_j$ is consistent for the true number of factors. One of the theorem's assumptions is the weak temporary dependence of the idiosyncratic terms. By construction, it does not hold for cross-sectionally independent nonstationary data. However, in actual empirical research, one would not know the validity of the assumptions. If the data are nonstationary, it would be natural to apply an IPC criterion.

Observe that if $k_{\max}$ is fixed while $n, T \to \infty$, then $\hat{k}_j \xrightarrow{\mathrm{P}} 0$. It is because, by Theorem 1 (i) and Lemma 6, for any fixed $k \leq k_{\max}$, $V(k)$ is of the probabilistic order T while the penalties are of higher orders.

Such a behavior of IPC_j is encouraging. However, in practice, one would arguably set $k_{\max}$ as some small fraction of $K_{nT} = \min\{n, T\}$, say $k_{\max} = [\delta K_{nT}]$. As the following theorem shows (see Appendix for a proof), in such a case the asymptotic behavior of the IPC_j will change drastically, unless the growth rates of n and T are unbalanced.

Proposition 2 *Under assumptions A1 and A2, as $n, T \to \infty$:*
(i) if $k_{\max}$ is fixed, then $\hat{k}_j \xrightarrow{\mathrm{P}} 0$ for $j = 1, 2, 3$;
(ii) if $k_{\max} = [\delta K_{n,T}]$ with fixed $\delta > 0$, then $\hat{k}_j \xrightarrow{\mathrm{P}} \infty$ with $j = 1, 2$, unless $n^2/\log n = O(T/\log\log T)$;
(iii) if $k_{\max} = [\delta K_{n,T}]$ with fixed $\delta > 0$, then $\hat{k}_3 \xrightarrow{\mathrm{P}} \infty$, unless either $n^2 = O(T \log T/\log\log T)$ or $T \log\log T = O(\log n)$.

The strong sensitivity of IPC_j to the choice of $k_{\max}$ can be circumvented by the use of the logarithmic criteria of the form

$$\log V(k) + k g(n, T).$$

In contrast to IPC_j, the logarithmic criteria do not have the scaling factor $\hat{\sigma}^2$ in the penalty, which therefore does not depend on $k_{\max}$. Bai (2004) shows the consistency of the corresponding $\hat{k}_{\log}$ under his assumptions (not holding in our setting) and when $g(n, T) \to \infty$ while $g(n, T)/\log T \to 0$. Unfortunately, since for any fixed k, $\log V(k) = O_{\mathrm{P}}(\log T)$, we immediately see that penalties satisfying the latter requirement yield $\hat{k}_{\log} \xrightarrow{\mathrm{P}} \infty$.

(n,T)	Content	Source
$(60,52)$	US annual industry-level employment.	Bai (2004)
$(243,83)$	European quarterly macroeconomic data	Boivin et al. (2009)
$(17,104)$	Pre-Euro part of the quarterly OECD exchange rates	Engel et al. (2015)
$(128,710)$	Current version of FRED-MD monthly macroeconomic dataset	McCracken and Ng (2015)
$(86,220)$	US quarterly "real activity dataset"	Stock and Watson (2016)

Table 1: Datasets used in the analysis below.

To assess the finite sample behavior of $\hat{k}_j$, $j=1,2,3$, we do some Monte Carlo (MC) analysis. Specifically, we simulate data on n independent random walks of length T, where the (n,T)-pairs correspond to the dimensions of five actual datasets described in Table 1.

Table 2 reports the obtained MC distributions of $\hat{k}=\hat{k}_1$. Results for $\hat{k}_2$ and $\hat{k}_3$ are similar and not reported. The number of MC replications is set to 10,000. The columns of the table correspond to different choices of $k_{\max}=6,...,15$. The entries of the table are the empirical probabilities (in percent rounded to the nearest integer) of observing a particular value of $\hat{k}$, which is given in the first column.

We see that the MC distributions of $\hat{k}$ concentrate at $\hat{k}=2$ or $\hat{k}=3$ for most of the settings. For example, when $k_{\max}=10$ and $(n,T)=(60,52)$, the MC probability of observing $\hat{k}=3$ equals 91%. For the same $k_{\max}$ and $(n,T)=(243,83)$, this probability becomes 100%. For $(n,T)=(128,710)$ and $(n,T)=(86,220)$, the mode of the MC distributions of $\hat{k}$ shifts to $\hat{k}=1$ (probability 100%) and $\hat{k}=2$ (probability 98%), respectively.

When n is relatively small, as in Engel et al.'s (2015) data, the MC distribution of $\hat{k}$ become very sensitive to the particular choice of $k_{\max}$ from the range $6,...,15$, which further supports the above theoretical claim regarding the sensitivity of the estimate of the number of 'factors' with respect to $k_{\max}$. However overall we see that, for empirically relevant data sizes, IPC criteria would typically estimate a small non-zero number of factors in the factorless persistent data.

3.2 Dickey-Fuller tests for the "idiosyncratic" series

One of the arguments in favour of doing factor analysis in levels discussed in Banerjee et al. (2017) is that the estimated idiosyncratic part of typical macroeconomic

k_{max}	6	7	8	9	10	11	12	13	14	15
$(n, T) = (60, 52)$ as in Bai (2004)										
$\hat{k} = 2$	96	75	43	15	4	1	0			
$\hat{k} = 3$	4	25	57	84	91	79	54	27	9	2
$\hat{k} = 4$			0	1	5	20	46	72	85	77
$\hat{k} = 5$							0	1	6	21
$(n, T) = (243, 83)$ as in Boivin et al. (2009)										
$\hat{k} = 2$	99	76	23	2	0					
$\hat{k} = 3$	1	24	77	98	100	98	84	50	19	4
$\hat{k} = 4$					0	2	16	50	81	96
$(n, T) = (17, 104)$ as in Engel et al. (2015)										
$\hat{k} = 1$	61	29	8	1	0					
$\hat{k} = 2$	39	70	84	68	31	6	0			
$\hat{k} = 3$	0	1	8	31	66	71	30	2	0	
$\hat{k} = 4$				0	3	23	64	46	2	
$\hat{k} = 5$						0	6	49	39	0
$\hat{k} = 6$							0	3	52	9
$\hat{k} = 7$									7	47
$\hat{k} = 8$									0	39
$\hat{k} = 9$										5
$(n, T) = (128, 710)$ as in FRED-MD dataset, McCracken and Ng (2015)										
$\hat{k} = 1$	100	100	100	100	100	97	88	67	41	20
$\hat{k} = 2$				0	0	3	12	33	59	80
$(n, T) = (86, 220)$ as in "real activity dataset", Stock and Watson (2016)										
$\hat{k} = 1$	91	61	27	7	2	0	0	0		
$\hat{k} = 2$	9	39	73	93	98	99	96	87	69	47
$\hat{k} = 3$				0	0	1	4	13	31	53

Table 2: The Monte Carlo distribution of the number of factors estimated using IPC_1 criterion. The probabilities in columns are measured in percent rounded to the nearest integer. The empty cells correspond to zero MC observations. The data are n independent random walks of length T each. The number of MC replications is 10,000.

data looks stationary in applications. The hypothesis of a unit root in the estimated idiosyncratic components can often be easily rejected. As pointed out in the introduction, such a rejection may be due to the standard unit root tests being seriously oversized.

Intuitively, the "idiosyncratic" terms of factorless persistent data are well approximated by the residuals from the OLS regression of the data on slowly varying trigonometric functions. Much of the variation of stochastically trending series is absorbed by such functions. Hence, the standard Dickey-Fuller null hypothesis over-states the low frequency content of the "idiosyncratic" terms, which leads to the over-rejection of the unit root null.

Formally, suppose that r factors are estimated by principal components and subtracted from the data to yield the estimated idiosyncratic series

$$\hat{e}_{jt} = X_{jt} - \bar{X}_j - \hat{\Lambda}_j' \hat{F}_t. \tag{10}$$

Further, suppose one runs the autoregression

$$\hat{e}_{jt} = \alpha_j + \beta_j \hat{e}_{jt-1} + \epsilon_{jt}, t = 2, ..., T. \tag{11}$$

Even if X_{jt}, $t = 1, ..., T$ is a pure random walk series, the asymptotic distributions of the corresponding Dickey-Fuller statistics are not the usual Dickey-Fuller ones.

A derivation of the asymptotic distribution of the OLS estimate $\hat{\beta}_j$ would require knowledge of the behavior of $\hat{F}_t$ beyond the consistency result reported in Theorem 1 (iii). We leave such a derivation for future research.

To support our intuition regarding the size distortion of the Dickey-Fuller test, we perform the following Monte Carlo experiment. For each of the empirically relevant sample sizes (n, T) reported in Table 1, we simulate n independent Gaussian random walks of length T. Then, we extract $0, 1, ..., 6$ "factors" from the simulated data and run the Dickey-Fuller regression (intercept only) on the remaining "idiosyncratic" series.

Table 3 reports the actual size of the Dickey-Fuller test. When no factors are extracted, the actual size equals the nominal one, which is set to 5%. However, when some factors are extracted, the tests become substantially over-sized. The size distortion becomes extreme when 6 factors are extracted, with the actual size becoming close to 100%.

	Number of "factors" extracted						
(n,T)	0	1	2	3	4	5	6
$(60,52)$	5	18.9	43.0	68.4	87.0	95.9	99.1
$(243,83)$	5	17.9	40.7	65.3	84.5	94.8	98.6
$(17,104)$	5	18.2	42.2	68.6	87.9	96.2	99.3
$(128,710)$	5	18.1	40.7	65.0	83.8	93.9	97.9
$(86,220)$	5	18.5	42.0	67.2	84.9	94.6	98.5

Table 3: The actual size of the 5% size Dickey-Fuller test (intercept only) based on the t-statistic, applied to the first component (in the cross-sectional order) of the "idiosyncratic" series. The series are obtained by subtracting a few "factors" from the pure random walk data of dimensions n and T. The number of MC replications is 10,000.

3.3 Inconsistency of the estimates of genuine factors

Suppose now that the data are not factorless. Instead they are given by

$$Y = \Lambda F' + X,$$

where X is as above and $\Lambda F'$ is the factor component. Unfortunately, the PC estimator of the genuine factors will be inconsistent, in general. Indeed, consider for concreteness the case of a single difference-stationary factor generated by

$$\begin{aligned} F_{1t} &= F_{10} + \sum_{s=1}^{t} \varepsilon_{Fs},\ t \in \mathbb{N}, \text{ where} \\ \varepsilon_{Fs} &= \sum_{k=0}^{\infty} \theta_{Fk} \xi_{F,s-k},\ s \in \mathbb{N} \end{aligned}$$

with ξ_{Ft} and θ_{Fk} satisfying Assumptions A1 and A2, similarly to ξ_{jt} and θ_{jk}.

Further, let the factor be pervasive in the usual sense that $\Lambda'\Lambda/n \to L > 0$ as $n \to \infty$. Denote $(F_{11}, ..., F_{1T})'$ as F_1, and its demeaned and normalized version $MF_1/\|MF_1\|$ as $\check{F}_1$. The principal eigenvector of $S_Y \equiv \frac{1}{n}\left(Y - \bar{Y}\right)'\left(Y - \bar{Y}\right)$, which we denote as $\tilde{F}_1$, can be interpreted as an estimator of $\check{F}_1$. The proof of the following proposition can be found in the Appendix.

Proposition 3 *There exist $\delta > 0$ and $p > 0$ such that the probability of event $\left(\tilde{F}_1'\check{F}_1\right)^2 < 1 - \delta$ remains no smaller than p for all sufficiently large n and T.*

The proposition implies that the probability that the angle between the demeaned and normalized genuine factor $\check{F}_1$ and its estimate $\tilde{F}_1$ is larger than a fixed positive number remains positive for all sufficiently large n and T. We interpret such a situation as the inconsistency of the PC estimate of the factor.

Remark. Proposition 3 considers unconditional probabilities. There certainly exist realizations of F_{1t}, $t \in \mathbb{N}$ such that the corresponding estimate $\tilde{F}_1$ approaches the direction of $\check{F}_1$ as $n, T \to \infty$. However, according to the proposition, the set of such realizations has measure less than one.

4 Discussion and conclusion

As we have seen above, factor analysis applied directly to large nonstationary panels may be spurious. This raises a question: how may one distinguish spurious from genuine results? A possible strategy consists of comparing factor estimates from the data in levels to those from the differenced data.

If all the nonstationarity in the data comes from factors, then under assumptions of Bai (2004) the PCA estimates $\hat{F}$ are consistent (up to a non-degenerate linear transformation) for the true factors F. Similarly, under assumptions of Bai and Ng (2004), the estimates $\hat{f}$ of the factors in the differenced data are consistent for ΔF. In such a case, $\Delta\hat{F}$ should be well aligned with $\hat{f}$. In contrast, a poor alignment would signal spurious results.

This strategy can be formalized as follows. Let P_l and P_d be the projections on the r-dimensional subspaces of $\mathbb{R}^{T-1}$ spanned by $\Delta\hat{F}$ and $\hat{f}$, respectively. The quality of alignment between those spaces can be measured by the first r eigenvalues of $P_l P_d$, which we denote as $\rho_{1r}^2 \geq ... \geq \rho_{rr}^2$. They may be interpreted as squared cosines of the principal angles between the spaces, or alternatively, as the squared sample canonical correlations between $\Delta\hat{F}$ and $\hat{f}$ (e.g. Hotelling, 1936). Observing ρ_{rr}^2 substantially below unity indicates a problem.

A simple choice of the threshold for ρ_{rr}^2 would be a fixed number, say $(0.5)^2$. This corresponds to the r-th canonical correlation between $\Delta\hat{F}$ and $\hat{f}$ equal to 0.5, which many researchers would probably consider to be convincingly small. For example, Stock and Watson (2016, ch. 6.4) view a canonical correlation of 0.6 between the spaces spanned by factors extracted from a large macro dataset and by selected macroeconomic indicators as a clear evidence of a mismatch.

We illustrate the proposed strategy by applying it to the Boivin et al. (2009) data, which motivated Uhlig (2009) (and us) to question the validity of the factor analysis of highly persistent series. Boivin et al.'s (2009) preferred number of

r	1	2	3	4	5	6	7	8
ρ_{rr}^2	0.99	0.01	0.80	0.12	0.14	0.08	0.08	0.00

Table 4: Values of ρ_r^2 for different choices of the number of factors r in Boivin et al. (2009).

factors is $r = 5$. We extract five factors from the data in levels, difference them and compute the five largest squared canonical correlations with five factors estimated from the differenced data. Their values turn out to be 1.00, 1.00, 0.98, 0.70, and 0.14. The smallest one, $\rho_{55}^2 = 0.14$, is below $(0.5)^2$, which clearly indicates a problem.

It is possible that low ρ_{55}^2 reflects a misspecification of the true number of factors r. If $r < 5$, there is no reason to expect that the fifth largest canonical correlation would be large. Furthermore, if $r > 5$, then since PCA estimates factors only up to a linear transformation, the low ρ_{55}^2 may reflect PCA picking up different five-dimensional factor subspaces from the level and from the differenced data.

To handle a possible misspecification of the number of factors, we compute ρ_{rr}^2 for a wide range of r from 1 to 8, which is the maximum number of factors considered by Boivin et al. (2009). Table 4 reports the results.

We see that $\rho_{rr}^2 \leq 0.14$ for all $r \geq 4$ from the considered range. Hence, it is unlikely that the problem comes from the underspecification of the number of factors. On the contrary, the table suggests that the true number of factors may be three. This is despite the fact that the fourth and the fifth principal components together 'explain' more than 10% of the data variation (see Figure 1). Such an apparent explanatory power may well be spurious.

An alternative interpretation of the results in Table 4 may be that, although the fourth and fifth factors are not spurious, they are weak. If so, the PCA estimator would contain much noise that picks up some features of the idiosyncratic dependence. When the data are differenced, those features change, which may lead to a mismatch between PCA estimators based on the level and differenced data. We leave a detailed analysis of this interesting possibility for future research.

To conclude, this paper warns empirical researches that a very high explanatory power of a few principal components of nonstationary data does not necessarily indicates the presence of factors. Even if such data are cross-sectionally independent, the first k principal components must explain $\sum_{j=1}^{k} 6/(j\pi)^2 \times 100\%$ of the variation, asymptotically. Unfortunately, relying on the standard criteria for the determination of the number of factors would lead to results that are very sensitive the choice of the maximum number of factors. Moreover, checking the stationarity of the PCA residuals using the Dickey-Fuller tests may spuriously favour the stationarity hypothesis. To detect these potential problems, we propose to always compare the PCA estimates obtained from the data in levels and in first differences. A mismatch signals a problem that necessitates further analysis.

5 Appendix

5.1 Eigenstructure of MU'

In preparation for the proof of Theorem 1, this section studies the singular value decomposition of MU'. Note that

$$UMU' = \begin{pmatrix} 0 & 0 \\ 0 & U_{T-1} \end{pmatrix} M \begin{pmatrix} 0 & 0 \\ 0 & U'_{T-1} \end{pmatrix},$$

where U_{T-1} is the $T-1$-dimensional upper triangular matrix of ones. Denoting the $T-1$-dimensional vector of ones as l_{T-1}, we obtain

$$UMU' = \begin{pmatrix} 0 & 0 \\ 0 & U_{T-1}\left(I_{T-1} - l_{T-1}l'_{T-1}/T\right)U'_{T-1} \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 \\ 0 & W \end{pmatrix}.$$

We have

$$W^{-1} = \left(U'_{T-1}\right)^{-1}\left(I_{T-1} + l_{T-1}l'_{T-1}\right)\left(U_{T-1}\right)^{-1}.$$

On the other hand, $\left(U_{T-1}\right)^{-1}$ is a two-diagonal matrix with 1 on the main diagonal and -1 on the super-diagonal. Therefore, W^{-1} is a three-diagonal matrix with 2 on the main diagonal, and -1 on the sub- and super-diagonals. As is well known, e.g. Sargan and Bhargava (1983), the eigenvalues of such a three-diagonal matrix, indexed in the increasing order, are $\mu_k = 2 - 2\cos\left(\omega_k/2\right),\ k = 1, ..., T-1$, where $\omega_k = 2\pi k/T$. The corresponding (normalized) eigenvectors are $\bar{u}_k = \left(\bar{u}_{k1}, ..., \bar{u}_{k,T-1}\right)'$ with $\bar{u}_{kj} = \sqrt{2/T}\sin\left(j\omega_k/2\right)$. This implies that the singular values of MU' (in decreasing order) are

$$\sigma_k = \sqrt{\mu_k^{-1}} = \left(2\sin\left(\omega_k/4\right)\right)^{-1}$$

for $k = 1, ..., T-1$ and $\sigma_T = 0$, and the components of the corresponding normalized right singular vectors are

$$u_{ks} = \sqrt{2/T}\sin((s-1)\,\omega_k/2),\ \ s = 1, ..., T$$

for $k = 1, ..., T-1$; and $u_{Ts} = 1$ for $s = 1$ and $u_{Ts} = 0$ for $s > 1$. Notice that u_{ks}, $s = 1, ..., T$, are proportional to the values at $(s-1)/T$ of the k-th principal eigenfunction of the covariance operator of the Brownian bridge process (e.g. Shorack and Wellner, 1986, pp. 213–214).

To find the k-th left singular vectors v_k with $k < T$, we multiply MU' by $\sigma_k^{-1}u_k$. We have $v_k = 2\sin(\omega_k/4)\,MU'u_k$. On the other hand, the j-th element of $U'u_k$ equals

$$\sqrt{2/T}\,\mathrm{Im}\sum_{s=0}^{j-1} e^{\mathrm{i}s\omega_k/2} = \sqrt{2/T}\,\mathrm{Im}\,\frac{e^{\mathrm{i}j\omega_k/2}-1}{e^{\mathrm{i}\omega_k/2}-1}$$

Therefore, $\sqrt{T/2}$ times the j-th element of $MU'u_k$ equals

$$\mathrm{Im}\,\frac{e^{\mathrm{i}j\omega_k/2}-1}{e^{\mathrm{i}\omega_k/2}-1} - \frac{1}{T}\,\mathrm{Im}\sum_{j=1}^{T}\frac{e^{\mathrm{i}j\omega_k/2}-1}{e^{\mathrm{i}\omega_k/2}-1} = \mathrm{Im}\,\frac{e^{\mathrm{i}j\omega_k/2}}{e^{\mathrm{i}\omega_k/2}-1} = -\frac{\cos((2j-1)\,\omega_k/4)}{2\sin(\omega_k/4)}.$$

Hence,

$$v_{ks} = -\sqrt{2/T}\cos((s-1/2)\,\omega_k/2),\ \ s = 1, ..., T$$

for $k < T$. Clearly, the left singular vector of MU' corresponding to zero singular value equals $v_T = \sqrt{1/T}l_T$.

Remark. From (9), we see that v_{ks} with $s = 1, ..., T$ and $k < T$ are proportional to the values at $(s-1/2)/T$ of the k-th principal eigenfunction of the covariance operator of the demeaned Brownian motion.

We summarize the results of this section in the following lemma.

Lemma 4 *Matrix MU' has the following singular value decomposition $MU' = \sum_{k=1}^{T}\sigma_k v_k u_k'$, where*

$$\sigma_k = \frac{1}{2\sin(\omega_k/4)},\ v_{ks} = -\sqrt{\frac{2}{T}}\cos\frac{(s-1/2)\,\omega_k}{2},\ \ \text{and } u_{ks} = \sqrt{\frac{2}{T}}\sin\frac{(s-1)\,\omega_k}{2}$$

for $k = 1, ..., T-1$, $s = 1, ..., T$, and $\omega_k = 2\pi k/T$. For $k = T$, we have $\sigma_T = 0$, $v_T = \sqrt{1/T}l_T$, and $u_T = e_1$, where e_1 is the first coordinate vector of $\mathbb{R}^T$.

5.2 Proof of Theorem 1

First, we will prove the theorem for $k = 1$. Then, we handle general k by mathematical induction. Since v_q, $q = 1, ..., T-1$, form an orthonormal basis in the space orthogonal to l_T and $\hat{F}_1$ belongs to this space, we have a representation

$$\hat{F}_1 = \sum_{q=1}^{T-1} \alpha_q v_q. \tag{12}$$

To prove part (iii) of the theorem, we have to show that $\alpha_1^2 \xrightarrow{\text{P}} 1$.

Recall that $\hat{F}_1$ is a normalized eigenvector of S corresponding to its largest eigenvalue λ_1. Therefore, representation (12) yields

$$\lambda_1 = \sum_{r,q=1}^{T-1} \alpha_r \alpha_q v_r' S v_q. \tag{13}$$

The idea of the proof consists of, first, showing that the latter sum is dominated by the terms $\alpha_r^2 v_r' S v_r$, and, then, demonstrating that $v_r' S v_r$ is quickly decreasing in r so that the maximum of (13) with respect to α's is achieved when α_1^2 is close to unity whereas α_r^2 with $r > 1$ are close to zero.

In what follows, we use notations introduced in the previous section. Further, $\varepsilon_{j\cdot}$ denotes the j-th row of ε, and $\|\cdot\|$ denotes the spectral norm of a matrix (or vector, in which case it coincides with the Euclidean norm).

By definition (4) of S and Lemma 4, (13) can be written as $\lambda_1 = A'A$, where $A = \sum_{r=1}^{T-1} \alpha_r \sigma_r \varepsilon u_r / \sqrt{n}$. Let R be a fixed positive integer. Split A into two sub-sums, A_1 and A_2, corresponding to $r \leq R$ and $r > R$, respectively. We have

$$\lambda_1 = A'A \leq \|A_1\|^2 + \|A_2\|^2 + 2\|A_1\|\|A_2\|. \tag{14}$$

Consider convenient explicit expressions for $\|A_1\|^2$ and $\|A_2\|^2$:

$$\|A_1\|^2 = \frac{1}{n} \sum_{r,q=1}^{R} \alpha_r \alpha_q \sigma_r \sigma_q u_r' \varepsilon' \varepsilon u_q \text{ and } \|A_2\|^2 = \frac{1}{n} \sum_{j=1}^{n} \left(\sum_{r=R+1}^{T-1} \alpha_r \sigma_r \varepsilon_{j\cdot} u_r \right)^2. \tag{15}$$

We will need the following two lemmas, proven in Sections 5.3 and 5.4.

Lemma 5 *Under assumptions A1 and A2, there exists an absolute constant C such that, for any $j = 1, ..., n$ and any $q, r, p, l = 1, ..., T-1$, we have*

(i) $\left|E\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r\right)-2\pi f_j\left(\omega_q/2\right)\delta_{qr}\right|\leq CB^2/T$, *where* δ_{qr} *is the Kronecker delta;*
(ii) $\left|Cov\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r, u_p'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_l\right)\right|\leq C\left(\delta_{qp}\delta_{rl}+\delta_{ql}\delta_{rp}+\left(1+\kappa_4\right)/T\right)B^4$.

Lemma 6 *Under assumptions A1 and A2, as* $n,T\to\infty$, $\operatorname{tr} S/\left(\gamma_n T^2\right)\xrightarrow{\mathrm{P}}1/6$.

Lemma 5 and the Chebyshev inequality yield

$$u_r'\varepsilon'\varepsilon u_q/n=\frac{1}{n}\sum_{j=1}^{n}2\pi f_j\left(\omega_q/2\right)\delta_{qr}+O_{\mathrm{P}}\left(1/\sqrt{n}\right)+O(1/T), \tag{16}$$

where $O_{\mathrm{P}}\left(1/\sqrt{n}\right)$ and $O(1/T)$ are uniform in $r,q=1,...,T-1$. Further, inequality (7) implies that $f_j\left(\omega_q/2\right)=f_j\left(0\right)+O\left(1/T\right)$ for any $q<R$, where $O\left(1/T\right)$ is uniform in $j\in\mathbb{N}$. Hence,

$$u_r'\varepsilon'\varepsilon u_q/n=\gamma_n\delta_{qr}+O_{\mathrm{P}}\left(1/\sqrt{n}\right)+O\left(1/T\right).$$

This equality, the first part of (15), and Lemma 4 yield

$$\left\|A_1\right\|^2=\sum_{q=1}^{R}\frac{\alpha_q^2\gamma_n}{4\sin^2\left(\omega_q/4\right)}+O_{\mathrm{P}}\left(T^2/\sqrt{n}\right)+O\left(T\right), \tag{17}$$

which implies that

$$\left\|A_1\right\|^2\leq\frac{\alpha_1^2\gamma_n}{4\sin^2\left(\omega_1/4\right)}+\frac{\sum_{q=2}^{R}\alpha_q^2\gamma_n}{4\sin^2\left(\omega_2/4\right)}+O_{\mathrm{P}}\left(T^2/\sqrt{n}\right)+O\left(T\right). \tag{18}$$

Further, since $\sin x\geq 2x/\pi$ for $x\in\left[0,\pi/2\right]$, we have

$$\sum_{q=1}^{R}\alpha_q^2/\sin^2\left(\omega_q/4\right)\leq\sin^{-2}\left(\omega_1/4\right)\leq T^2.$$

Therefore, (17) yields $\left\|A_1\right\|\leq T\left(\gamma_n/4+O_{\mathrm{P}}\left(1/\sqrt{n}\right)+O\left(1/T\right)\right)^{1/2}$. Using the Taylor expansion of the latter square root and recalling that γ_n is bounded and bounded away from zero (see (6) and (8)), we obtain

$$\left\|A_1\right\|\leq T\gamma_n^{1/2}/2+O_{\mathrm{P}}\left(T/\sqrt{n}\right)+O(1). \tag{19}$$

For A_2, the Cauchy-Schwarz inequality and the identity $\sum \alpha_i^2 = 1$ yield

$$\|A_2\|^2 \leq \left(1 - \sum_{i=1}^{R} \alpha_i^2\right) \frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2 . \tag{20}$$

We have

$$E \frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2 = \frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \sigma_r^2 u_r' \Gamma_j u_r \leq \|\Gamma_j\| \sum_{r=R+1}^{T-1} \sigma_r^2,$$

where $\Gamma_j = E\left(\varepsilon_{j\cdot}' \varepsilon_{j\cdot}\right)$ is the Toeplitz matrix of the autocovariances of ε_{jt}. As is well known (e.g. Grenander and Szegö (1958), p. 64),

$$\|\Gamma_j\| \leq 2\pi \max_{\omega \in [0, 2\pi]} f_j(\omega) \leq B^2,$$

where the latter inequality follows from (6). Therefore,

$$E \frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2 \leq B^2 \sum_{r=R+1}^{T-1} \sigma_r^2 = B^2 \sum_{r=R+1}^{T-1} \frac{1}{4 \sin^2\left(\omega_r / 4\right)}.$$

Again using the fact that $\sin x \geq 2x/\pi$ for $x \in [0, \pi/2]$, we obtain

$$E \frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2 \leq \frac{B^2 T^2}{4} \sum_{r=R+1}^{\infty} \frac{1}{r^2}. \tag{21}$$

Furthermore, by Lemma 5 (ii), the variance of $\frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2$ is no larger than

$$\frac{O(1)}{n} \sum_{r=R+1}^{T-1} \frac{1}{16 \sin^4\left(\omega_r / 4\right)} + \frac{O\left(1/T\right)}{n} \left(\sum_{r=R+1}^{T-1} \frac{1}{4 \sin^2\left(\omega_r / 4\right)}\right)^2 = O\left(T^4 / n\right).$$

This estimate of the variance, inequality (21), and the Chebyshev inequality yield

$$\frac{1}{n} \sum_{j=1}^{n} \sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2 \leq \frac{B^2 T^2}{4} \sum_{r=R+1}^{\infty} \frac{1}{r^2} + O_{\mathrm{P}}\left(T^2 / \sqrt{n}\right).$$

Using this in (20), we obtain

$$\|A_2\|^2 \leq \left(1 - \sum_{i=1}^{R} \alpha_i^2\right) \frac{B^2T^2}{4} \sum_{r=R+1}^{\infty} \frac{1}{r^2} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right). \tag{22}$$

Now let $\delta < 1$ be an arbitrarily small positive number. Since $\gamma_n \geq b^2 > 0$ (see (8)), we can choose R so large, and independent from n and T, that

$$\sum_{r=R+1}^{\infty} \frac{1}{r^2} < \delta \frac{\gamma_n}{\pi^2 B^2}. \tag{23}$$

Then, since $\pi/T > \sin(\omega_2/4)$,

$$\frac{B^2T^2}{4} \sum_{r=R+1}^{\infty} \frac{1}{r^2} \leq \delta \frac{\gamma_n}{4\sin^2(\omega_2/4)},$$

so that

$$\|A_2\|^2 \leq \left(1 - \sum_{i=1}^{R} \alpha_i^2\right) \frac{\delta\gamma_n}{4\sin^2(\omega_2/4)} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right). \tag{24}$$

Proceeding as in the derivation of (19) from (18), this time using the inequality

$$\left(1 - \sum_{i=1}^{R} \alpha_i^2\right) \sin^{-2}(\omega_2/4) \leq T^2/4,$$

we obtain

$$\|A_2\| \leq \frac{\delta^{1/2}\gamma_n^{1/2}}{4} T + O_{\mathrm{P}}\left(T/\sqrt{n}\right). \tag{25}$$

Since $\delta < 1$, (24) trivially yields

$$\|A_2\|^2 \leq \left(1 - \sum_{i=1}^{R} \alpha_i^2\right) \frac{\gamma_n}{4\sin^2(\omega_2/4)} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right). \tag{26}$$

Using (18), (19), (25), and (26) in (14), we obtain

$$\lambda_1 \leq \frac{\alpha_1^2\gamma_n}{4\sin^2(\omega_1/4)} + \frac{(1-\alpha_1^2)\gamma_n}{4\sin^2(\omega_2/4)} + \frac{\delta^{1/2}\gamma_n T^2}{4} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right) + O(T). \tag{27}$$

On the other hand, λ_1 must be no smaller than

$$v_1' S v_1 = \sigma_1^2 u_1' \varepsilon' \varepsilon u_1 / n = \frac{\gamma_n}{4\sin^2(\omega_1/4)} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right) + O(T), \tag{28}$$

where the latter equality follows from (16). Hence, overall we must have

$$\frac{(1-\alpha_1^2)\,\gamma_n}{4\sin^2(\omega_1/4)} - \frac{(1-\alpha_1^2)\,\gamma_n}{4\sin^2(\omega_2/4)} \leq \frac{\delta^{1/2}\gamma_n T^2}{4} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right) + O(T)$$

so that

$$\begin{aligned}\left(1-\alpha_1^2\right) &\leq \frac{\delta^{1/2}T^2/4 + \gamma_n^{-1}\left(O_{\mathrm{P}}\left(T^2/\sqrt{n}\right) + O(T)\right)}{\left(4\sin^2(\omega_1/4)\right)^{-1} - \left(4\sin^2(\omega_2/4)\right)^{-1}} \\ &\leq C\delta^{1/2} + \gamma_n^{-1}\left(O_{\mathrm{P}}\left(1/\sqrt{n}\right) + O(1/T)\right),\end{aligned}$$

where C is an absolute constant. Since δ can be chosen arbitrarily small, we conclude that $\alpha_1^2 \xrightarrow{\mathrm{P}} 1$, which establishes statement (iii) of the theorem for $k = 1$.

Statement (i) for $k = 1$ follows from the upper bound (27) and the lower bound (28) on λ_1. Indeed the bounds imply that

$$\left|\lambda_1 - \frac{\gamma_n}{4\sin^2(\omega_1/4)}\right| \leq \delta^{1/2}\gamma_n T^2/4 + O(T) + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right).$$

Since $4\sin^2(\omega_1/4) = \pi^2/T^2 + o\left(T^{-2}\right)$ and δ can be chosen arbitrarily small, this yields

$$\lambda_1 = \gamma_n T^2/\pi^2 + o_{\mathrm{P}}\left(T^2\right).$$

Statement (ii) for $k = 1$, follows from statement (i) and Lemma 6.

Now let us use mathematical induction to prove the theorem for $k > 1$. Suppose that the theorem holds for $k < m$. Consider a representation $\hat{F}_m = \sum_{q=1}^{T-1} \alpha_q v_q$. Since $\hat{F}_m' \hat{F}_j = 0$ for all $j < m$, and since $\left|\hat{F}_j' v_j\right| = 1 + o_{\mathrm{P}}(1)$ by the induction hypothesis, we must have $\alpha_j = o_{\mathrm{P}}(1)$ for all $j < m$. In particular,

$$\hat{F}_m' S \hat{F}_m = \sum_{q,r=m}^{T-1} \alpha_q \alpha_r \sigma_q \sigma_r u_q' \varepsilon' \varepsilon u_r / n + o_{\mathrm{P}}\left(T^2\right). \tag{29}$$

Indeed, to see that (29) holds, it is sufficient to establish equalities $\alpha_j v_j' S \sum_{r=m}^{T-1} \alpha_r v_r = o_{\mathrm{P}}\left(T^2\right)$ for any $j < m$, and equalities $\alpha_j \alpha_r v_j' S v_r = o_{\mathrm{P}}\left(T^2\right)$ for any $j, r < m$.

Such equalities easily follow from the facts that $\alpha_j = o_{\mathrm{P}}(1)$ for all $j < m$ and $\|S\| = \lambda_1 = \gamma_n T^2/\pi^2 + o_{\mathrm{P}}(T^2)$.

In addition to (29), we must have

$$\sum_{i=1}^{m-1} \lambda_i + \hat{F}_m' S \hat{F}_m \geq \sum_{i=1}^{m} v_i' M U' \varepsilon' \varepsilon U M v_i / n = \sum_{i=1}^{m} \frac{\gamma_n}{4 \sin^2(\omega_i/4)} + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right) + O(T),$$

where the latter equality is obtained similarly to (28). Combining the above two displays, and using the induction hypothesis, this time regarding the validity of the identities

$$\lambda_i = \frac{\gamma_n}{4 \sin^2(\omega_i/4)} + o_{\mathrm{P}}\left(T^2\right)$$

for all $i < m$, we obtain

$$\sum_{q,r=m}^{T-1} \alpha_q \alpha_r \sigma_q \sigma_r u_q' \varepsilon' \varepsilon u_r / n \geq \frac{\gamma_n}{4 \sin^2(\omega_m/4)} + o_{\mathrm{P}}\left(T^2\right). \tag{30}$$

The theorem for $k = m$ now follows by arguments that are very similar to those used above for the case $k = 1$.

That is, we represent the sum on the left hand side of (30) in the form $A'A$, where $A = \sum_{r=m}^{T-1} \alpha_r \sigma_r \varepsilon u_r / \sqrt{n}$. Then we split A into parts corresponding to $r \leq R$ and $r > R$, and proceed along the lines of the above proof to obtain an upper bound on $A'A$, similar to the right hand side of (27). Then, combining this upper bound with the lower bound (30), we prove the convergence $\alpha_m^2 \xrightarrow{\mathrm{P}} 1$. Finally, we proceed to establishing parts (i) and (ii) using part (iii). We omit details to save space.

5.3 Proof of Lemma 5

Consider the finite Fourier transform of $\varepsilon_{j\cdot}$ (e.g. Brillinger (2001, ch. 3.1), which will be referred to in what follows as B01)

$$d(\omega) = \sum_{t=1}^{T} \varepsilon_{j,t} e^{-\mathrm{i}(t-1)\omega}, \omega \in [0, 2\pi].$$

Let us denote $d(\omega_r/2)$ as d_r and $d(-\omega_r/2)$ as d_{-r}. By definition (see Lemma 4), the t-th entry of u_r for $r = 1, ..., T-1$ equals

$$u_{rt} = \sqrt{2/T}\left(e^{\mathrm{i}(t-1)\omega_r/2} - e^{-\mathrm{i}(t-1)\omega_r/2}\right) / (2\mathrm{i}).$$

Therefore, $\varepsilon_{j\cdot}u_r = \sqrt{2/T}\left(d_{-r} - d_r\right)/\left(2\mathrm{i}\right)$ and thus,

$$E\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r\right) = -\frac{1}{2T}E\left[\left(d_{-q} - d_q\right)\left(d_{-r} - d_r\right)\right],$$

and

$$Cov\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r, u_p'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_l\right) = \frac{1}{4T^2}Cov\left(\left(d_{-q} - d_q\right)\left(d_{-r} - d_r\right), \left(d_{-p} - d_p\right)\left(d_{-l} - d_l\right)\right).$$

To evaluate the latter expectation and covariance, we use Theorem 4.3.2 of B01, which describes joint cumulants of finite Fourier transforms. Therefore, we need to represent the expectation and covariance in terms of the joint cumulants. By their definition, and by Theorem 2.3.1 (B01, p.19),

$$E\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r\right) = \frac{1}{2T}\sum\nolimits_{s_1,s_2\in\{-1,+1\}} s_1 s_2 cum\left(d_{s_1 q}, d_{s_2 r}\right). \tag{31}$$

Similarly, $Cov\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r, u_p'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_l\right)$ equals

$$\frac{1}{4T^2}\sum\nolimits_{s_1,s_2,s_3,s_4\in\{-1,+1\}} s_1 s_2 s_3 s_4 cum\left(d_{s_1 q}d_{s_2 r}, d_{s_3 p}d_{s_4 l}\right).$$

By Theorem 2.3.2 of B01, the joint cumulant of the two products of d, as in the latter display, can be represented in the form of a sum of the products of the cumulants of order two and the fourth-order cumulant. Precisely, we have

$$\begin{aligned} & Cov\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r, u_p'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_l\right) = \frac{1}{4T^2}\sum\nolimits_{s_1,s_2,s_3,s_4\in\{-1,+1\}} s_1 s_2 s_3 s_4 \\ & \times\left\{cum\left(d_{s_1 q}, d_{s_3 p}\right)cum\left(d_{s_2 r}, d_{s_4 l}\right) + cum\left(d_{s_1 q}, d_{s_4 l}\right)cum\left(d_{s_2 r}, d_{s_3 p}\right)\right. \\ & \left. + cum\left(d_{s_1 q}, d_{s_2 r}, d_{s_3 p}, d_{s_4 l}\right)\right\}. \end{aligned} \tag{32}$$

Lemma 7 *Under assumptions A1, A2, there exists an absolute constant C such that, for any $q, r, p, l = 1, ..., T-1$, and any $s_1, s_2, s_3, s_4 \in \{-1, +1\}$,*

$$\left|cum\left(d_{s_1 q}, d_{s_2 r}\right) - 2\pi H_{s_1 q, s_2 r} f_j\left(\omega_q/2\right)\right| \leq CB^2, \tag{33}$$

where $H_{s_1 q, s_2 r} = \sum\nolimits_{t=0}^{T-1} e^{-\mathrm{i}t(s_1\omega_q + s_2\omega_r)/2}$, and

$$\left|cum\left(d_{s_1 q}, d_{s_2 r}, d_{s_3 p}, d_{s_4 l}\right) - (2\pi)^3 H_{s_1 q, s_2 r, s_3 p, s_4 l} f_{j4}\right| \leq C\kappa_4 B^4, \tag{34}$$

where $H_{s_1 q, s_2 r, s_3 p, s_4 l} = \sum\nolimits_{t=0}^{T-1} e^{-\mathrm{i}t(s_1\omega_q + s_2\omega_r + s_3\omega_p + s_4\omega_l)/2}$, and f_{j4} is the 4-th order *cumulant spectrum of the series* ε_{jt} *at frequencies* $s_1\omega_q/2,\ s_2\omega_r/2,\ s_3\omega_p/2.$

Proof: The proof of Theorem 4.3.2 in B01 implies that the left hand side of (33) can be bounded by $C\sum_{k=0}^{\infty}(1+k)\left|\Gamma_j(k)\right|,$ where C is an absolute constant and

$$\Gamma_j(k)=E\varepsilon_{js}\varepsilon_{j,s-k}=\sum_{t=-\infty}^{\infty}\theta_{jt}\theta_{j,t-k}.$$

Here θ_{jt} with negative t are defined as zero. On the other hand,

$$\begin{aligned}\sum_{k=0}^{\infty}(1+k)\left|\Gamma_j(k)\right| &\leq \sum_{k=0}^{\infty}(1+k)\sum_{t=-\infty}^{\infty}\left|\theta_{jt}\right|\left|\theta_{j,t-k}\right| \\ &\leq \sum_{k=0}^{\infty}\sum_{t=-\infty}^{\infty}(1+|t-k|)\left|\theta_{jt}\right|\left|\theta_{j,t-k}\right|+\sum_{k=0}^{\infty}\sum_{t=-\infty}^{\infty}(1+|t|)\left|\theta_{jt}\right|\left|\theta_{j,t-k}\right|\leq 2B^2,\end{aligned}\tag{35}$$

where the last inequality follows from assumption A2. This yields (33).

Similarly, from the proof of Theorem 4.3.2 in B01, we know that the left hand side of (34) can be bounded by

$$C\sum_{k_1,k_2,k_3=-\infty}^{\infty}(1+|k_1|+|k_2|+|k_3|)\left|c_{j4}(k_1,k_2,k_3)\right|,\tag{36}$$

where C is an absolute constant and $c_{j4}(k_1,k_2,k_3)$ is the joint 4-th order cumulant of ε_{js}, $\varepsilon_{j,s-k_1}$, $\varepsilon_{j,s-k_2}$, and $\varepsilon_{j,s-k_3}.$ By Theorem 2.3.1 of B01, this cumulant equals

$$\begin{aligned}&\sum_{t_1,t_2,t_3,t_4=-\infty}^{\infty}\theta_{j,t_1-k_1}\theta_{j,t_2-k_2}\theta_{j,t_3-k_3}\theta_{j,t_4}cum\left(\xi_{j,-t_1},\xi_{j,-t_2},\xi_{j,-t_3},\xi_{j,-t_4}\right)\\ =\ &\sum_{t=-\infty}^{\infty}\theta_{j,t_1-k_1}\theta_{j,t_2-k_2}\theta_{j,t_3-k_3}\theta_{j,t_4}\left(E\xi_{jt}^4-3\right)\\ \leq\ &\sum_{t=-\infty}^{\infty}\left|\theta_{j,t_1-k_1}\theta_{j,t_2-k_2}\theta_{j,t_3-k_3}\theta_{j,t_4}\right|\kappa_4,\end{aligned}$$

where the last line follows from assumption A1. By an argument similar to (35), expression (36) can be bounded by $C\kappa_4B^4$, where C is an absolute constant. This yields (34). □

Returning to the proof of Lemma 5, consider (31). Inequality (33) implies that

$$\left|E\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r\right)-\frac{1}{2T}\sum_{s_1,s_2\in\{-1,+1\}}s_1s_2H_{s_1q,s_2r}2\pi f_j(\omega_q/2)\right|\leq\frac{2KB^2}{T}.\tag{37}$$

Further, for $q=r$,

$$\sum_{s_1,s_2\in\{-1,+1\}}s_1s_2H_{s_1q,s_2r}=\sum_{s_1,s_2\in\{-1,+1\}}s_1s_2\sum_{t=0}^{T-1}e^{-\mathrm{i}t(s_1+s_2)\pi q/T}=2T.\tag{38}$$

For $q \neq r$ and such that $s_1q + s_2r$ is even for all $s_1, s_2 \in \{-1, +1\}$,

$$\sum\nolimits_{s_1,s_2\in\{-1,+1\}} s_1s_2H_{s_1q,s_2r} = \sum\nolimits_{s_1,s_2\in\{-1,+1\}} s_1s_2 \sum\nolimits_{t=0}^{T-1} e^{-\mathrm{i}t(s_1q+s_2r)\pi/T} = 0. \quad (39)$$

Here, the latter equality holds because $s_1q + s_2r$ is an even nonzero integer, such that $|s_1q + s_2r| < 2T$ (recall that $1 \leq q, r \leq T - 1$). For $q \neq r$ and such that $s_1q + s_2r$ is odd for all $s_1, s_2 \in \{-1, +1\}$, we have

$$\sum\nolimits_{t=0}^{T-1} e^{-\mathrm{i}t(s_1q+s_2r)\pi/T} = \frac{-2}{e^{-\mathrm{i}(s_1q+s_2r)\pi/T} - 1}.$$

Nevertheless, $\sum\nolimits_{s_1,s_2\in\{-1,+1\}} s_1s_2H_{s_1q,s_2r}$ still equals zero because

$$\frac{-2}{e^{-\mathrm{i}(q+r)\pi/T} - 1} + \frac{-2}{e^{\mathrm{i}(q+r)\pi/T} - 1} + \frac{2}{e^{-\mathrm{i}(q-r)\pi/T} - 1} + \frac{2}{e^{\mathrm{i}(q-r)\pi/T} - 1} = 2 - 2.$$

Therefore, (39) still holds. Using identities (38) and (39) in (37), we obtain statement (i) of Lemma 5.

Next, consider (32). By (33) and (34), the difference between

$$\begin{aligned}&\left(\frac{\pi}{T}\right)^2 \sum\nolimits_{s_1,s_2,s_3,s_4\in\{-1,+1\}} s_1s_2s_3s_4 \{2\pi H_{s_1q,s_2r,s_3p,s_4l}f_{j4} \\ &+ (H_{s_1q,s_3p}H_{s_2r,s_4l} + H_{s_1q,s_4l}H_{s_2r,s_3p}) f_j(\omega_q/2) f_j(\omega_r/2)\}\end{aligned}$$

and $Cov\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r, u_p'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_l\right)$ is no larger by absolute value than

$$\begin{aligned}\frac{1}{4T^2} \quad &\sum\nolimits_{s_1,s_2,s_3,s_4\in\{-1,+1\}} \{C\kappa_4B^4 + 2C^2B^4 + 2\pi CB^2 (|H_{s_1q,s_3p}f_j(\omega_q/2)| \\ &+ |H_{s_2r,s_4l}f_j(\omega_r/2)| + |H_{s_1q,s_4l}f_j(\omega_q/2)| + |H_{s_2r,s_3p}f_j(\omega_r/2)|)\},\end{aligned}$$

which, in its turn, is bounded from above by $C(1 + \kappa_4) B^4/T$, where C is an absolute constant (throughout the paper, the value of the absolute constant C may change from one appearance to another). Indeed, such a bound follows from (6) and the fact that $|H_{a,b}| \leq T$. Further, from the above analysis of $E\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r\right)$,

$$\sum\nolimits_{s_1,s_2,s_3,s_4\in\{-1,+1\}} s_1s_2s_3s_4 (H_{s_1q,s_3p}H_{s_2r,s_4l} + H_{s_1q,s_4l}H_{s_2r,s_3p}) = 4T^2 (\delta_{qp}\delta_{rl} + \delta_{ql}\delta_{rp}).$$

Therefore, from (6),

$$\left|\sum\nolimits_{s_1,s_2,s_3,s_4\in\{-1,+1\}} s_1s_2s_3s_4 (H_{s_1q,s_3p}H_{s_2r,s_4l} + H_{s_1q,s_4l}H_{s_2r,s_3p}) f_j(\omega_q/2) f_j(\omega_r/2)\right|$$

is no larger than $4T^2B^4\left(\delta_{qp}\delta_{rl}+\delta_{ql}\delta_{rp}\right)/\left(2\pi\right)^2$. Next, by Theorem 2.8.1 of B01,

$$f_{j4}\left(\mu_1,\mu_2,\mu_3\right)=\Theta\left(\mu_1\right)\Theta\left(\mu_2\right)\Theta\left(\mu_3\right)\Theta\left(-\mu_1-\mu_2-\mu_3\right)\frac{E\xi_{jt}^4-3}{\left(2\pi\right)^3},$$

where $\Theta\left(\mu\right)=\sum_{k=0}^{\infty}\theta_k e^{-\mathrm{i}k\mu}$, and since $\left|H_{a,b,c,d}\right|\leq T$,

$$\left|\sum_{s_1,s_2,s_3,s_4\in\{-1,+1\}}s_1s_2s_3s_4 2\pi H_{s_1q,s_2r,s_3p,s_4l}f_{j4}\right|\leq\frac{TB^4\kappa_4}{\left(2\pi\right)^2}.$$

Overall, we conclude that $\left|Cov\left(u_q'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_r,u_p'\varepsilon_{j\cdot}'\varepsilon_{j\cdot}u_l\right)\right|$ is no larger than

$$\left(\frac{\pi}{T}\right)^2\left(\frac{TB^4\kappa_4}{\left(2\pi\right)^2}+\frac{4T^2B^4\left(\delta_{qp}\delta_{rl}+\delta_{ql}\delta_{rp}\right)}{\left(2\pi\right)^2}\right)+\frac{C\left(1+\kappa_4\right)B^4}{T},$$

which yields statement (ii) of Lemma 5.

5.4 Proof of Lemma 6

Let R be a fixed positive integer. Consider the representation

$$\mathrm{tr}\,S=\sum_{r=1}^{T-1}v_r'Sv_r=S_1+S_2,$$

where

$$S_1=\sum_{r=1}^{R}\frac{1}{n}\sum_{j=1}^{n}\left(\sigma_r\varepsilon_{j\cdot}u_r\right)^2,\ S_2=\sum_{r=R+1}^{T-1}\frac{1}{n}\sum_{j=1}^{n}\left(\sigma_r\varepsilon_{j\cdot}u_r\right)^2.$$

Lemma 5 (i) together with inequality (7) and Lemma 4 yield

$$\begin{aligned}ES_1&=\sum_{r=1}^{R}\sigma_r^2\frac{1}{n}\sum_{j=1}^{n}2\pi f_j\left(\omega_r/2\right)+O(T)=\sum_{r=1}^{R}\sigma_r^2\gamma_n+O(T)\\&=\sum_{r=1}^{R}\frac{\gamma_nT^2}{\pi^2r^2}+O(T).\end{aligned}$$

On the other hand, by Lemma 5 (ii) and Lemma 4,

$$VarS_1=\frac{1}{n^2}\sum_{j=1}^{n}Var\left(\sum_{r=1}^{R}\left(\sigma_r\varepsilon_{j\cdot}u_r\right)^2\right)=O\left(T^4/n\right).$$

Therefore, by Chebyshev's inequality

$$S_1 = \sum_{r=1}^{R} \frac{\gamma_n T^2}{\pi^2 r^2} + O(T) + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right). \tag{40}$$

Similarly, Lemma 5 (i) and inequality (6) yield

$$ES_2 \leq B^2\left(1 + C/T\right) \sum_{r=R+1}^{T-1} \sigma_r^2,$$

where C is an absolute constant. Further, by Lemma 5 (ii) and Lemma 4,

$$VarS_2 = \frac{1}{n^2} \sum_{j=1}^{n} Var\left(\sum_{r=R+1}^{T-1} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2\right) = O\left(T^4/n\right),$$

and hence, by Chebyshev's inequality

$$S_2 \leq B^2\left(1 + C/T\right) \sum_{r=R+1}^{T-1} \sigma_r^2 + O_{\mathrm{P}}\left(T^2/\sqrt{n}\right). \tag{41}$$

The statement of the lemma follows from (40), (41), the fact that $\sum_{r=R+1}^{T-1} \sigma_r^2/T^2$ can be made arbitrarily small by choosing R sufficiently large, and the Euler formula $\sum_{r=1}^{\infty} r^{-2} = \pi^2/6$.

5.5 Proof of Proposition 2

Validity of part (i) has been shown in Section 3.1. To establish parts (ii) and (iii) we need the following lemma.

Lemma 8 *Under Assumptions A1 and A2, for* $k_{\max} = [\delta K_{nT}]$, *we have* $\hat{\sigma}^2 = O_{\mathrm{P}}\left(T/K_{nT}\right)$.

Proof: Recall that $\hat{\sigma}^2 = V(k_{\max})$. We have

$$TV(k_{\max}) = \operatorname{tr} S - \sum_{j=1}^{k_{\max}} \lambda_j \leq \sum_{r=k_{\max}+1}^{T-1} v_r' S v_r = \sum_{r=k_{\max}+1}^{T-1} \frac{1}{n} \sum_{j=1}^{n} \left(\sigma_r \varepsilon_{j\cdot} u_r\right)^2.$$

Denote this upper bound on $TV(k_{\max})$ as $T\bar{V}$. Lemma 5 (i), inequality (6), and the fact that

$$\sigma_r = \left(2\sin\left(\omega_r/4\right)\right)^{-1} \leq T/2r \tag{42}$$

yield

$$\begin{aligned} E\left(T\bar{V}\right) &\leq B^2\left(1+C/T\right)\sum_{r=k_{\max}+1}^{T-1}\sigma_r^2 \leq B^2\left(1+C/T\right)\frac{T^2}{4}\sum_{r=k_{\max}+1}^{T-1}\frac{1}{r^2} \\ &< \frac{B^2\left(1+C/T\right)T^2}{4k_{\max}} = \frac{B^2\left(1+C/T\right)T^2}{4\left[\delta K_{nT}\right]}. \end{aligned}$$

Further, by Lemma 5 (ii),

$$\begin{aligned} Var\left(T\bar{V}\right) &= \frac{1}{n^2}\sum_{j=1}^{n} Var\left(\sum_{r=k_{\max}+1}^{T-1}\left(\sigma_r\varepsilon_{j.}u_r\right)^2\right) \\ &= \frac{O(1)}{n}\sum_{r=k_{\max}+1}^{T-1}\sum_{q=k_{\max}+1}^{T-1}\left(\delta_{rq}+1/T\right)\sigma_r^2\sigma_q^2. \end{aligned}$$

From this and (42), we obtain

$$\begin{aligned} Var\left(T\bar{V}\right) &\leq \frac{O(1)}{n}\left(\sum_{r=k_{\max}+1}^{T-1}\frac{T^4}{16r^4}+\frac{1}{T}\sum_{r=k_{\max}+1}^{T-1}\sum_{q=k_{\max}+1}^{T-1}\frac{T^4}{16r^2q^2}\right) \\ &\leq \frac{O(1)}{n}\left(\frac{T^4}{k_{\max}^3}+\frac{T^3}{k_{\max}^2}\right) = O\left(\frac{T^4}{n\left[\delta K_{nT}\right]^3}\right). \end{aligned}$$

Therefore, by Chebyshev's inequality,

$$\hat{\sigma}^2 = V(k_{\max}) \leq \frac{B^2\left(1+C/T\right)T}{4\left[\delta K_{nT}\right]} + O_{\mathrm{P}}\left(\frac{T}{n^{1/2}\left[\delta K_{nT}\right]^{3/2}}\right) = O_{\mathrm{P}}\left(\frac{T}{K_{nT}}\right). \ \square$$

For n growing at least as fast as T, Lemma 8 implies that the penalties for IPC_1 and IPC_2 are asymptotically dominated by $V(k) = O_P(T)$ for any fixed k. Therefore, $\hat{k}_j \xrightarrow{\mathrm{P}} \infty$ for $j = 1, 2$. If n grows slower than T, then $V(k)$ does not asymptotically dominate the penalties for IPC_1 and IPC_2 only as long as

$$\hat{\sigma}^2 T \log n / \log\log T \geq O_{\mathrm{P}}(T).$$

By Lemma 8 this can hold only if $n^2/\log n = O(T/\log\log T)$, which completes the proof of (ii).

Similarly, for n growing at the same rate as T, Lemma 8 implies that the penalty for IPC_3 is asymptotically dominated by $V(k) = O_P(T)$ for any fixed k. Hence, $\hat{k}_3 \xrightarrow{\mathrm{P}} \infty$. For n growing faster than T, $V(k)$ does not asymptotically dominate the penalty only as long as

$$\hat{\sigma}^2 \log(nT) / \log\log T \geq O_{\mathrm{P}}(T).$$

By Lemma 8 this can hold only if $T \log\log T = O(\log n)$. Finally, for n growing slower than T, $V(k)$ does not asymptotically dominate the penalty only as long as

$$\hat{\sigma}^2 T \log(nT) / (n \log\log T) \geq O_{\mathrm{P}}(T).$$

By Lemma 8 this can hold only if $n^2 = O(T \log T / \log\log T)$. □

5.6 Proof of Proposition 3

Consider a decomposition

$$\left(\tilde{F}_1' \check{F}_1\right)^2 = 2\left(\left(\tilde{F}_1 - \hat{F}_1\right)' \check{F}_1\right)^2 + 2\left(\hat{F}_1' \check{F}_1\right)^2, \tag{43}$$

where $\hat{F}_j$ is the principal eigenvector of $S = \frac{1}{n} MU'\varepsilon'\varepsilon UM$. We will show that, for all sufficiently large n and T, both terms on the right hand side of (43) are small with a positive probability.

For the first term, we have

$$2\left(\left(\tilde{F}_1 - \hat{F}_1\right)' \check{F}_1\right)^2 \leq 2\left\|\tilde{F}_1 - \hat{F}_1\right\|^2 = 4\left(1 - \left(\tilde{F}_1' \hat{F}_1\right)^2\right). \tag{44}$$

Denote $\tilde{F}_1' \hat{F}_1$ as α and consider the decomposition $\tilde{F}_1 = \alpha \hat{F}_1 + \sqrt{1-\alpha^2}\hat{F}_{-1}$, where $\hat{F}_{-1}$ is a unit-length vector orthogonal to $\hat{F}_1$. Let

$$S_1 = S_Y - S = \frac{1}{n}\left(MU'\varepsilon'\Lambda\check{F}_1' + \check{F}_1\Lambda'\varepsilon UM\right)\|MF_1\| + \frac{1}{n}\check{F}_1\Lambda'\Lambda\check{F}_1'\|MF_1\|^2. \tag{45}$$

Since $\hat{F}_1$ is the principal eigenvector of S and $\hat{F}_{-1}$ is orthogonal to it, we have

$$\tilde{F}_1' S_Y \tilde{F}_1 \leq \tilde{F}_1' S \tilde{F}_1 + \|S_1\| \leq \alpha^2 \lambda_1 + \left(1-\alpha^2\right)\lambda_2 + \|S_1\|,$$

where λ_j denotes the j-th largest eigenvalue of S. On the other hand, since $\tilde{F}_1$ is the principal eigenvector of S_Y,

$$\lambda_1 = \hat{F}_1' S \hat{F}_1 = \hat{F}_1' (S_Y - S_1) \hat{F}_1 \leq \tilde{F}_1' S_Y \tilde{F}_1 + \|S_1\|.$$

The latter two displays yield

$$\lambda_1 \leq \alpha^2 \lambda_1 + \left(1-\alpha^2\right)\lambda_2 + 2\left\|S_1\right\|.$$

Recalling that $\alpha = \tilde{F}_1' \hat{F}_1$ and using (44), we obtain

$$2\left(\left(\tilde{F}_1 - \hat{F}_1\right)' \check{F}_1\right)^2 \leq 8\left\|S_1\right\| / \left(\lambda_1 - \lambda_2\right). \tag{46}$$

From (45), by the triangle and Cauchy-Schwarz inequalities,

$$\begin{aligned}\frac{8\left\|S_1\right\|}{\lambda_1-\lambda_2} &\leq \frac{8\left(2\lambda_1^{1/2}\left(\Lambda'\Lambda/n\right)^{1/2}\left\|MF_1\right\| + \left(\Lambda'\Lambda/n\right)\left\|MF_1\right\|^2\right)}{\lambda_1-\lambda_2} \\ &= \left(16\left(\gamma_n \xi_1\xi_2\xi_3\right)^{1/2} + 8\xi_1\xi_2\right) / \left(\gamma_n \xi_4\right),\end{aligned}$$

where

$$\xi_1 = \left\|MF_1/T\right\|^2,\ \xi_2 = \Lambda'\Lambda/n,\ \xi_3 = \lambda_1/\gamma_n T^2,\ \text{and } \xi_4 = \left(\lambda_1-\lambda_2\right)/\left(\gamma_n T^2\right).$$

By (6) and (8), $b^2 \leq \gamma_n \leq B^2$. Therefore,

$$8\left\|S_1\right\| / \left(\lambda_1 - \lambda_2\right) \leq \bar{S} \equiv \left(16B\left(\xi_1\xi_2\xi_3\right)^{1/2} + 8\xi_1\xi_2\right) / \left(b^2\xi_4\right) \tag{47}$$

Standard arguments (e.g. Stock (1994, p. 2754)) yield

$$\xi_1 \xrightarrow{\mathrm{d}} 2\pi f_F(0) \int_0^1 \left(\tilde{W}(x)\right)^2 \mathrm{d}x,$$

where $f_F(0)$ is the spectral density of the process ΔF_{1t} at zero, and $\tilde{W}(x)$ is the demeaned Brownian motion. By adjusting the factor loadings Λ, we can normalize $2\pi f_F(0)$ to 1 without loss of generality. Further, by assumption, $\xi_2 \to L$, and by Theorem 1 (i), $\xi_3 \xrightarrow{\mathrm{P}} 1/\pi^2$ and $\xi_4 \xrightarrow{\mathrm{P}} 3/\left(2\pi\right)^2$. Therefore,

$$\bar{S} \xrightarrow{\mathrm{d}} w_1 \left(\int_0^1 \left(\tilde{W}(x)\right)^2 \mathrm{d}x\right)^{1/2} + w_2 \int_0^1 \left(\tilde{W}(x)\right)^2 \mathrm{d}x. \tag{48}$$

with

$$w_1 = 64\pi B L^{1/2} / \left(3b^2\right) \text{ and } w_2 = 32\pi^2 L / \left(3b^2\right).$$

Now, let us analyze the second term on the right hand side of (43). We have

$$2\left(\hat{F}_1'\check{F}_1\right)^2 = 2\left(\left(\hat{F}_1 - v_1\right)'\check{F}_1 + v_1'\check{F}_1\right)^2,$$

where v_1 is the vector described in Theorem 1 (iii). Since by that theorem, $\left(\hat{F}_1'v_1\right)^2 \xrightarrow{\text{P}} 1$, we have

$$\left(\hat{F}_1 - v_1\right)'\check{F}_1 \xrightarrow{\text{P}} 0.$$

On the other hand,

$$v_1'\check{F}_1 \xrightarrow{\text{d}} -\sqrt{2}\int_0^1 \cos(\pi x)\,\tilde{W}(x)\mathrm{d}x / \left(\int_0^1 \left(\tilde{W}(x)\right)^2 \mathrm{d}x\right)^{1/2}.$$

Hence,

$$2\left(\hat{F}_1'\check{F}_1\right)^2 \xrightarrow{\text{d}} 4\left(\int_0^1 \cos(\pi x)\,\tilde{W}(x)\mathrm{d}x\right)^2 / \left(\int_0^1 \left(\tilde{W}(x)\right)^2 \mathrm{d}x\right). \qquad (49)$$

Note that the latter convergence is joint with (48). Using the Karhunen-Loève expansion (9), we have

$$\int_0^1 \left(\tilde{W}(x)\right)^2 \mathrm{d}x = \sum_{k=1}^{\infty}\left(\frac{z_k}{\pi k}\right)^2 \text{ and } 4\left(\int_0^1 \cos(\pi x)\,\tilde{W}(x)\mathrm{d}x\right)^2 = 2\left(\frac{z_1}{\pi}\right)^2,$$

where z_k with $k = 1, 2, ...$ are i.i.d. standard normal random variables.

Finally let $\mathcal{A}$ be the event that, simultaneously,

$$Q/2 \le \sum_{k=2}^{\infty}\left(\frac{z_k}{\pi k}\right)^2 \le 11Q/12 \text{ and } \left(\frac{z_1}{\pi}\right)^2 \le Q/12,$$

where $Q = \min\left\{1/\left(16w_1^2\right), 1/\left(4w_2\right)\right\}$. Let ϵ be the probability of $\mathcal{A}$. Clearly, $\epsilon > 0$. On the other hand, $\mathcal{A}$ implies that the right hand side of (48) is no larger than 1/2, whereas the right hand side of (49) is no larger than 1/3. Therefore, the probability of the event that, simultaneously,

$$2\left(\left(\tilde{F}_1 - \hat{F}_1\right)'\check{F}_1\right)^2 \le 1/2 \text{ and } 2\left(\hat{F}_1'\check{F}_1\right)^2 \le 1/3$$

must be no smaller than ϵ for all sufficiently large n and T. This fact together with

(43) yield the statement of the proposition.

References

[1] Bai, J. (2004) "Estimating cross-section common stochastic trends in nonstationary panel data," *Journal of Econometrics* 122, 137–183.

[2] Bai, J. and S. Ng (2004) "A PANIC attack on unit roots and cointegration," *Econometrica* 72, 1127–1177.

[3] Banerjee, A., Marcellino, M., and Masten, I. (2017) "Structural FECM: Cointegration in large-scale structural FAVAR models," *Journal of Applied Econometrics* 32, 1069–1086.

[4] Barigozzi, M., M. Lippi, and M. Luciani (2018) "Non-Stationary Dynamic Factor Models for Large Datasets," *arXiv 1602.02398v3*

[5] Boivin, J., Giannoni, M. P., and Mojon, B. (2009) "How Has the Euro Changed the Monetary Transmission Mechanism?" in Acemoglu, D., Rogoff, K., and Woodford, M. (eds.) *NBER Macroeconomic Annual 2008*, Volume 23.

[6] Brillinger, D. R. (2001) *Time Series: Data Analysis and Theory*. SIAM.

[7] Bryzgalova, S. (2018) "Spurious Factors in Linear Asset Pricing Models," *working paper*, Stanford Graduate School of Business.

[8] Engel, C., N.C. Mark, and K.D. West (2015) "Factor Model Forecasts of Exchange Rates," *Econometric Reviews* 34, 32-55.

[9] Grenander, U. and Szegö, G. (1958) *Toeplitz Forms and Their Applications.* University of California Press.

[10] Hotelling, H. (1936) "Relations Between Two Sets of Variates," *Biometrika* 28, 321–377.

[11] Koltchinskii, V. and Lounichi, K. (2016) "Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance," *Annales de l'Institut Henri Poincaré* 52, 1976–2013.

[12] Koltchinskii, V. and Lounichi, K. (2017) "Normal Approximation and Concentration of Spectral Projectors of Sample Covariance," *Annals of Statistics* 45, 121–157.

[13] McCracken, M. W. and Ng, S. (2015) "FRED-MD: A monthly database for macroeconomic research," *Working Papers 2015-12*, St. Louis, MO, Federal Reserve Bank of St. Louis.

[14] Müller, U. K. and Watson, M. W. (2008) "Testing Models of Low-Frequency Variability," *Econometrica* 76, 979–1016.

[15] Müller, U. K. and Watson, M. W. (2016) "Measuring Uncertainty About Long-Run Predictions," *Review of Economic Studies* 83, 1711–1740.

[16] Müller, U. K. and Watson, M. W. (2017) "Low-Frequency Econometrics," in *Advances in Economics: Eleventh World Congress of the Econometric Society*, ed. by B. Honoré and L. Samuelson. Cambridge: Cambridge University Press.

[17] Müller, U. K. and Watson, M. W. (2018) "Long-Run Covariability," *Econometrica* 86, 775–804.

[18] Phillips, P.C.B. (1986) "Understanding Spurious Regression in Econometrics," *Journal of Econometrics* 33, 311–340.

[19] Phillips, P.C.B. (1998) "New tools for understanding spurious regression," *Econometrica* 66, 1299-1325.

[20] Sargan, J. D., and Bhargava, A. (1983) "Testing residuals from least squares regression for being generated by the Gaussian random walk," *Econometrica* 51, 153–174.

[21] Shorack, G. R., and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics.* New York: Wiley.

[22] Stock, J. H. (1994) " Unit Roots, Structural Breaks and Trends," in Engle, R. F. and McFadden, D. L. (eds.) *Handbook of Econometrics*, Vol. 4, Ch. 46, 2739–2841.

[23] Stock, J.H., and Watson, M.W. (2016) "Factor Models and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, Vol2A, John B. Taylor and Harald Uhlig (eds), Chapter 8, 415–526.

[24] Uhlig, H. (2009) "Comment on "How Has the Euro Changed the Monetary Transmission Mechanism?"," in Acemoglu, D., Rogoff, K., and Woodford, M. (eds) *NBER Macroeconomic Annual 2008*, Volume 23.