Econometric Theory, **0**, 2024, 1–38.
doi:10.1017/S0266466624000094

SPURIOUS FACTORS IN DATA WITH LOCAL-TO-UNIT ROOTS

Alexei Onatski
University of Cambridge

Chen Wang
University of Hong Kong

This paper extends the spurious factor analysis of Onatski and Wang (2021, Spurious factor analysis. *Econometrica*, 89(2), 591–614.) to high-dimensional data with heterogeneous local-to-unit roots. We find a spurious factor phenomenon similar to that observed in the data with unit roots. Namely, the "factors" estimated by the principal components analysis converge to principal eigenfunctions of a weighted average of the covariance kernels of the demeaned Ornstein–Uhlenbeck processes with different decay rates. Thus, such "factors" reflect the structure of the strong temporal correlation of the data and do not correspond to any cross-sectional commonalities, that genuine factors are usually associated with. Furthermore, the principal eigenvalues of the sample covariance matrix are very large relative to the other eigenvalues, creating an illusion of the "factors"capturing much of the data's common variation. We conjecture that the spurious factor phenomenon holds, more generally, for data obtained from high frequency sampling of heterogeneous continuous time (or spacial) processes, and provide an illustration.

1. INTRODUCTION

In a recent paper, Onatski and Wang (2021) (OW) describe a spurious factor phenomenon observed in high-dimensional data *with unit roots*. When principal components analysis (PCA) is applied to such data, the extracted "common factors" apparently explain very large portions of variation even in situations where the data are cross-sectionally independent. The time series plots of such estimated "factors" resemble cosine waves corresponding to the principal eigenfunctions of the covariance operator of a demeaned Brownian motion. The danger for an empirical researcher is to take the high explanatory power of these waves for evidence of common economic factors, and read too much into the associated cyclical pattern.

The authors would like to thank Peter C.B. Phillips, Yixiao Sun, and three anonymous referees for valuable suggestions that deepen our understanding of the spurious factor phenomenon. Chen Wang's research is supported by the Research Grants Council of Hong Kong (Grant No. ECS 27308219) and the National Natural Science Foundation of China (Grant No. 72033002). Address correspondence to Chen Wang, Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam, Hong Kong, China, stacw@hku.hk

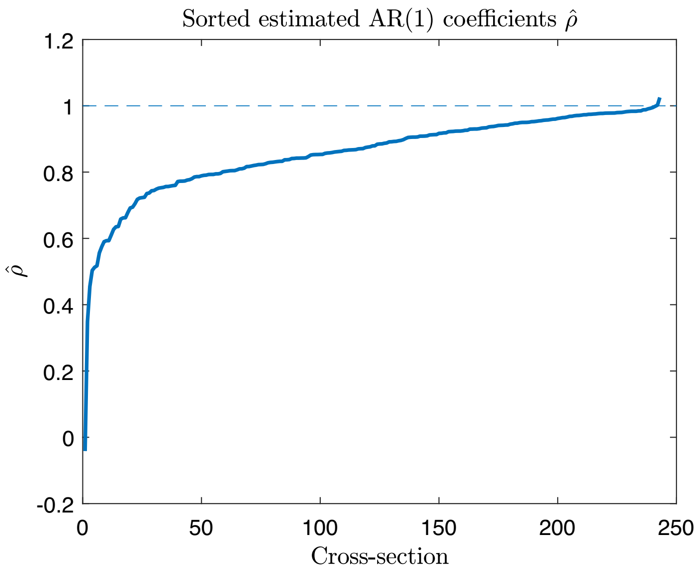

FIGURE 1. Distribution of the coefficients of AR(1) fitted series-by-series to the BGM data.

An early warning came from Uhlig's (2009) discussion of Boivin, Giannoni, and Mojon (2009) (BGM), who study a factor augmented VAR with factors extracted from 245 quarterly macroeconomic series, for the period from 1980:1 to 2007:3. As illustrated by Figure 1 (which reproduces Uhlig's Figure 3), the data are rather persistent. Uhlig (2009) generated *cross-sectionally independent* data with similar high degree of persistence and compared the explanatory power of the factors extracted from the original data with that of "factors" extracted from the so generated factorless data. He found a similarity so striking (see Figure 2) that he initially attributed the finding to a coding error.

OW provides a theoretical explanation of Uhlig's finding under the assumption that the data have unit roots. However, a cursory look at Figure 1 suggests that the assumption of unit roots in all of the persistent macroeconomic series is extreme. In fact, the null of the unit root is not rejected by the augmented Dickey–Fuller test at 5% significance level only for about half of the 245 series used in BGM. Furthermore, about a third of these non-rejections correspond to "marginal" p-values in between 0.05 and 0.15. This supports an obvious idea that it may be more appropriate to model typical macroeconomic data as a combination of stationary, unit root, and *local-to-unit root* series with potentially *different localizing parameters*. This paper extends OW's analysis to such data.

Precisely, we consider N-dimensional data X_t, $t = 1, \dots, T$, satisfying a near integrated system

$$X_t - \mu_X = \rho\,(X_{t-1} - \mu_X) + \Psi\,(L)\,\varepsilon_t, \tag{1}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

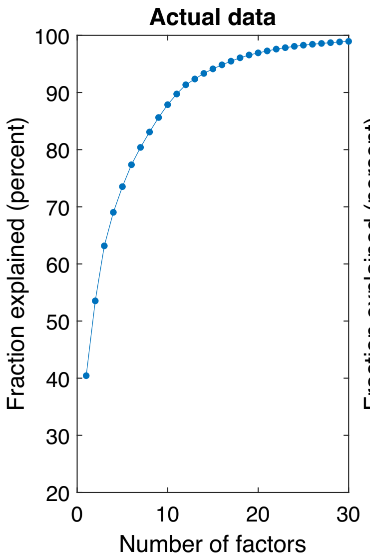

FIGURE 2. Factor contribution to the overall variance. Left panel: Actual BMG data. Right panel: Factorless simulated data with similar autocorrelation properties.

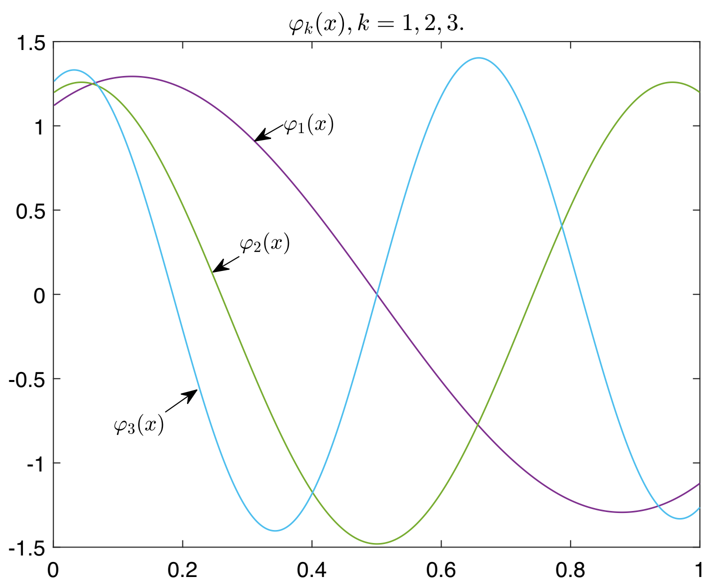

FIGURE 3. The probability "limits" of the first three spurious factor estimates. Local-to-unity parameter uniformly distributed on $[0, 10]$.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

where $\rho = \text{diag}\{\rho_1, \ldots, \rho_N\}$ with $\rho_j = \exp\left\{-\phi_j/T\right\}$ and $\phi_j \geq 0$, μ_X is an arbitrary deterministic vector, and ε_t are N-dimensional innovation vectors. The data may have unit roots ($\phi_j = 0$) and local-to-unit roots with different localizing parameters ($\phi_j > 0$). Furthermore, we allow the $N \times N$ matrix lag polynomial $\Psi(L)$ to be such that some of these unit or local-to-unit roots "cancel out" so that a portion of the data may be not persistent, and cointegration is allowed.

Literature on near integrated systems (e.g., Phillips, 1988; Elliott, 1998) usually considers a triangular form of the system, where the data generating process for near integrated stochastic trends and the "cointegrating" relationships are modeled explicitly as two sub-systems. We work with (1) because this form naturally generalizes OW, which considers a special case of such a system with ρ equal to the identity matrix I_N.

Our focus is on the asymptotic properties of the principal components extracted from data generated by (1) as both N and T go to infinity with unconstrained relative rate of the divergence. In modern economic studies that deal with high-dimensional data, PCA is often used as a dimension reduction technique, usually justified by an assumption that the data have a common factor structure (see, e.g., Stock and Watson, 2016).

The PCA estimate, $\hat{F}_k$, of the kth factor is simply the normalized eigenvector of the matrix

$$\hat{\Sigma} := (X - \bar{X})'(X - \bar{X})/N,$$

which corresponds to its kth largest eigenvalue $\hat{\lambda}_k$. Here, $X = [X_1, \ldots, X_T]$, and $\bar{X}$ is an $N \times T$ matrix with all columns equal to $\frac{1}{T}\sum_{t=1}^{T} X_t$. The eigenvalue $\hat{\lambda}_k$ measures the explanatory power of the kth factor, with $\left(\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma}\right) \times 100\%$ expressing this power in percentage terms.

Of course, system (1) does not necessarily generate data with common factors. In fact, if $\Psi(L)$ is diagonal, then the data are cross-sectionally independent. However, an empirical researcher would not know the data generating process and may want to reduce the data dimensionality using PCA. The main point of this paper is to show that, even in the absence of any common factors, such PCA would reveal disproportionally large explanatory power of a few extracted "factors." The time series plots of these "factors" would reflect the form of the principal eigenfunctions of the covariance operators associated with the series, and would have nothing to do with a history of common shocks driving the data's dynamics. We call this phenomenon the spurious factor analysis.

To develop intuition, consider first the unit root case, studied in OW. Then rows of matrix X can be viewed as discrete approximations to Brownian motions. Hence, the sample covariance matrix $\hat{\Sigma}$ can be interpreted as a discrete analog of the covariance kernel (aka covariance function) of a demeaned Brownian motion. Therefore, we would expect the eigenstructure of $\hat{\Sigma}$ to resemble that of the covariance kernel. The latter is characterized by quickly decaying eigenvalues $1/\pi^2, 1/(2\pi)^2, 1/(3\pi)^2, \ldots$, so that the principal ones appear to be disproportionally large, and cosine-wave eigenfunctions $\sqrt{2}\cos(\pi x), \sqrt{2}\cos(2\pi x), \sqrt{2}\cos(3\pi x), \ldots$ with $x \in [0,1]$.

For the local-to-unit root case studied here, the rows of X can be viewed as discrete approximations to OU processes with decay rates ϕ_j (e.g., Phillips, 1988; Stock, 1994). Had these decay rates been the same for all j, the situation would have been analogous to the unit root case with the Brownian motion replaced by the Ornstein–Uhlenbeck (OU) process. Then we would have expected that the PCA "factors" correspond to the eigenfunctions of the covariance kernel of the demeaned OU process. As we show below, when ϕ_j's are different, the spurious factors correspond to eigenfunctions of a weighted average of the covariance kernels of the demeaned OU processes with different decay rates. Such weighted averages still have very large principal eigenvalues and wave-like eigenfunctions, even though explicit formulas describing these eigenvalues and eigenvectors do not exist.

Our strategy for the asymptotic analysis of the eigenstructure of $\hat{\Sigma}$ in the heterogeneous local-to-unit root case can be outlined as follows. First, we use an extension of the Beveridge–Nelson decomposition to nearly integrated series (1) to extract the "long-run component" of X_t, and define an analog of $\hat{\Sigma}$, denoted as $\tilde{\Sigma}$, that substitutes this long-run component for X_t. We show that the asymptotics of the principal eigenvalues and eigenvectors of $\hat{\Sigma}$ and $\tilde{\Sigma}$ are the same so that it is sufficient to analyze the latter.

Then we split $\tilde{\Sigma}$ into deterministic and random parts: $\mathbb{E}\tilde{\Sigma}$ and $\tilde{\Sigma} - \mathbb{E}\tilde{\Sigma}$, respectively. We show that the asymptotics of the principal eigenstructure of $\tilde{\Sigma}$ and $\mathbb{E}\tilde{\Sigma}$ coincide. Finally, we show that there exist "approximating integral operators" $K_{N,T}$ acting on the space of continuous functions on $[0,1]$, such that, on one hand, their nonzero eigenvalues coincide with those of $\mathbb{E}\tilde{\Sigma}/T^2$ and the corresponding eigenfunctions, evaluated on the grid $1/T, 2/T, \ldots, T/T$, are eigenvectors of $\mathbb{E}\tilde{\Sigma}/T^2$. On the other hand, the principal eigenvalues and eigenfunctions of $K_{N,T}$ converge to those of a weighted average of covariance kernels of OU processes. This yields our main results (see Theorem 1).

Anselone (1967) traces the technique of approximating integral operators by matrices back to Fredholm, and the idea of mapping matrices to operators with essentially same spectral properties to Nystrom. Our proof of the convergence of $K_{N,T}$ (in Section 4) is based on the ideas of Anselone (1967). The key facts to establish are: the pointwise convergence $K_{N,T}$ and the so-called collective compactness of the sequence of operators $\{K_{N,T} : N, T = 1, 2, \ldots\}$. We find these techniques and ideas are very powerful and hope that they may be useful for econometricians working with high-dimensional asymptotics beyond the analysis of near integrated systems.

We conjecture that the spurious factor phenomenon holds, more generally, for data obtained from high frequency sampling of heterogeneous continuous time (or spacial) processes.[1] In Section 5, we illustrate this conjecture by considering PCA analysis of data that comes from high frequency sampling of continuous time

[1] We are grateful to Yixiao Sun for his stimulating questions regarding such a possibility.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

Markov chains with only two states, and low probabilities of switching between the states. We show that the PCA analysis of such data delivers very similar results to the PCA analysis of stationary data with local-to-unit roots. In addition, Section 5 discusses the spurious factor phenomenon in data with nearly explosive roots, and data that may contain a growing number of genuine factors.

Before turning to the detailed analysis, let us briefly discuss some related previous literature. Phillips (1998) provides an analytic foundation and intuitive explanation for results of the type we find, in the context of spurious regression. In particular, that paper explains that prototypical spurious regressions reproduce Karhunen–Loève representation (based on the spectral decomposition of the covariance kernels) of the relevant stochastic processes, such as Brownian motion and diffusions. Gonzalo and Pitarakis (2021) study the asymptotic behavior of the eigenvalues of the sample covariance matrix of vector auto-regression series with strong or mild persistence, including the local-to-unit root case. They show that the largest eigenvalues are disproportionally large even in the absence of any factors in the series. They consider a special case of (1) with homogeneous localizing parameter ϕ and $\Psi(L) = I$. Zhang, Gao, and Pan (2020) find the probability limits of the largest eigenvalues of the sample covariance matrix of high-dimensional near unit root time series. They show that the asymptotic fluctuations around these probability limits are Gaussian. Their assumptions about the data generating process are different from ours. In particular, their Assumption 3 restricts relative rates of divergence of the cross-sectional and temporal dimensions of the data whereas we leave these rates unrestricted. Furthermore, their Assumption 7 essentially assumes away the heterogeneity in the local-to-unit root localizing parameters, asymptotically. Neither of the latter two papers discusses the behavior of the eigenvectors.

Similar to PCA, cointegration analysis could be viewed as another dimension reduction device, where a large system of persistent time series is dominated by a smaller number of common components. Accordingly, we do observe the counterpart of spurious factors, that is, spurious cointegration in large vector auto-regressions. The possibility of spurious cointegration was noted in many previous works. For recent exciting advances in this area, we refer the reader to Bykhovskaya and Gorin (2022a, 2022b).

The remaining part of the paper is organized as follows: Section 2 describes our setup and discusses assumptions that we make. Section 3 formulates and discusses the main result, Theorem 1. Section 4 contains the proof of the theorem. Section 5 considers various extensions and provides a more general discussion of the spurious factors phenomenon. Section 6 concludes. Proofs of auxiliary results are given in the Supplementary Material.

2. SETUP

Recall system (1):

$$X_t - \mu_X = \rho\,(X_{t-1} - \mu_X) + \Psi\,(L)\,\varepsilon_t, \qquad \textbf{(1)}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

where $\rho = \rho(N,T) = \mathrm{diag}\{\rho_1,\ldots,\rho_N\}$ with $\rho_j = \exp\{-\phi_j/T\}$ and $\phi_j \geq 0$. We suppress the dependence of ρ on N and T to make our notations simpler. Similarly, the dependence of $\Psi(L)$ on N and T is tacitly assumed.

System (1) generalizes the setting in OW, which considers $\rho = I_N$. Here, we do allow some of $\rho_j = 1$ (hence, $\phi_j = 0$). Without loss of generality, we may assume that $\phi_j = 0$ for $j \leq N_1$ and $\phi_j > 0$ for $j > N_1$. That is, the first $N_1 \leq N$ components of X_t are unit root processes. Let us denote the subvector of X_t that consists of these components as $X_t^{(1)}$ and the complementary subvector as $X_t^{(2)}$. Conformably to this partition, let us partition μ_X into $\mu_X^{(1)}$ and $\mu_X^{(2)}$. We impose no constraints on the initial values $X_0^{(1)}$ of the unit root components, and set $X_0^{(2)}$ so that the process $X_t^{(2)}$ is stationary (albeit with local-to-unity roots). Precisely,

$$X_0^{(2)} - \mu_X^{(2)} = \sum\nolimits_{s=0}^{\infty} \left(\rho^{(2)}\right)^s \Psi^{(2)}(L)\,\varepsilon_{-s},$$

where $\rho^{(2)} = \mathrm{diag}\left\{\rho_{N_1+1},\ldots,\rho_N\right\}$ and $\Psi^{(2)}(L)$ is the matrix lag polynomial that consists of the last $N-N_1$ rows of $\Psi(L)$. A similar assumption on the initial values of local-to-unity processes is made in Elliott (1999) (for a discussion of other initial conditions, see Section 5).

Denote the ith component of the vector ε_t as ε_{it}, and let $N \times N$ matrices Ψ_k be the coefficients of the matrix lag polynomial $\Psi(L) = \sum_{k=0}^{\infty}\Psi_k L^k$. We make the following assumptions.

Assumption A1. *Random variables ε_{it} with $i \in \mathbb{N}$ and $t \in \mathbb{Z}$ are independent and such that $\mathbb{E}\varepsilon_{it} = 0$, $\mathbb{E}\varepsilon_{it}^2 = 1$, and $\varkappa_4 = \sup_{i\in\mathbb{N},\,t\in\mathbb{Z}} \mathbb{E}\varepsilon_{it}^4 < \infty$.*

Note that ε_{it} may have different distributions, although they have to be independent. Further, the normalization $\mathbb{E}\varepsilon_{it}^2 = 1$ is not restrictive as it may be accommodated by the matrix lag polynomial $\Psi(L)$.

Assumption A2. *As $N,T \to \infty$ at arbitrary relative rates, $\sum_{k=0}^{\infty}(1+k)\|\Psi_k\| = O(1)$, where $\|\cdot\|$ denotes the spectral norm of a matrix.*

Assumption A2 is implied by uniform (over N and T) one-summability of filter $\Psi(L)$ (see Brillinger, 1981, Sect. 2.7). Such an assumption mildly restricts temporal and cross-sectional dependence of $\Psi(L)\varepsilon_t$, and is often used in the analysis of finite-dimensional integrated series (e.g., Watson, 1994, Sect. 2.3).

Assumption A3. *The effective rank of the long-run covariance matrix $\Omega = \Psi(1)\Psi(1)'$, defined as $\mathrm{tr}\,\Omega/\|\Omega\|$, diverges to infinity as $N,T \to \infty$ at arbitrary relative rates.*

The divergence of the effective rank implies the divergence of the usual concept of rank. Hence, Assumption A3 implies that the number of stochastic trends in the data has to diverge, possibly slowly, as $N,T \to \infty$.

Further, Assumption A3 allows some of the local-to-unit roots in (1), or even most of them, to cancel out. For example, suppose $\Psi(L)$ is diagonal with $\Psi(L)_{ii} = 1 - \mathbf{1}\{i > n\}\rho_i L$, where $n \to \infty$, possibly slowly, as $N,T \to \infty$. Then, for any $i > n$,

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

the local-to-unit roots cancel out, and we have $X_{it} = \mu_{Xi} + \varepsilon_{it}$ which are serially uncorrelated processes. In this example, $\|\Omega\| = 1$ and $\operatorname{tr}\Omega > n$ so that Assumption A3 holds.

Our last assumption involves the covariance kernels of demeaned OU processes. As mentioned in the introduction, the jth row of X can be viewed as a discrete time approximation to an OU process with decay rate ϕ_j (unless a cancellation of roots takes place). Such a process, let us call it $x_{\phi_j}(s)$, $s \in [0,1]$, is generated by a stochastic differential equation

$$\mathrm{d}x_{\phi_j}(s) = -\phi_j x_{\phi_j}(s)\mathrm{d}s + \Omega_{jj}^{1/2}\mathrm{d}W(s),$$

with the standard Wiener process $W(s)$. The scaling $\Omega_{jj}^{1/2}$ is the long-run standard deviation of the jth component of $\Psi(L)\varepsilon_t$. The initial observation $x_{\phi_j}(0)$ is drawn from the unconditional distribution of $x_{\phi_j}(s)$.

As is well-known (e.g., Karatzas and Shreve, 1991, p. 358), the covariance kernel of $x_{\phi_j}(s)$ is given by $\Omega_{jj}R_{\phi_j}(s,t)$, where $R_{\phi_j}(s,t) = e^{-\phi_j|t-s|}/\left(2\phi_j\right)$. It is straightforward to verify that the covariance kernel of the *demeaned* OU process equals $\Omega_{jj}k_{\phi_j}(s,t)$ with

$$k_{\phi_j}(s,t) = R_{\phi_j}(s,t) - \int_0^1 R_{\phi_j}(s,t)\,\mathrm{d}s - \int_0^1 R_{\phi_j}(s,t)\,\mathrm{d}t + \int_0^1\int_0^1 R_{\phi_j}(s,t)\,\mathrm{d}s\mathrm{d}t. \quad \textbf{(2)}$$

Define the weighted average kernel as

$$k_{\mathcal{F}}(s,t) = \int\int \omega k_{\phi}(s,t)\,\mathcal{F}(\mathrm{d}\omega,\mathrm{d}\phi),$$

where $\mathcal{F}$ is a probability distribution on $[0,\infty)^2$. We will interpret $\mathcal{F}$ as a weak limit of $\mathcal{F}_N$—the empirical joint distribution of Ω_{jj} and $\phi_j, j = 1,\ldots,N$. Let $K_{\mathcal{F}}$ be the integral operator, acting in the space $C[0,1]$ of continuous functions on $[0,1]$, with kernel $k_{\mathcal{F}}(s,t)$.

Assumption A4. *$\mathcal{F}_N$ weakly converges to $\mathcal{F}$ as $N,T \to \infty$ at arbitrary relative rates. The supports of $\mathcal{F}_N$ and $\mathcal{F}$ belong to $[0,\bar{\omega}] \times \left[0,\bar{\phi}\right]$ for some $0 < \bar{\omega},\bar{\phi} < \infty$. The eigenvalues $\mu_1 > \mu_2 > \ldots$ of $K_{\mathcal{F}}$ are simple.*

The weak convergence of $\mathcal{F}_N$ to $\mathcal{F}$ would happen almost surely if pairs $\left(\Omega_{jj},\phi_j\right)$ were drawn at random from the distribution $\mathcal{F}$. However, such a random sampling is not necessary for the convergence, and we leave its underlying mechanism unspecified. The assumption of simple eigenvalues sharpens our results and makes them easier to interpret. Furthermore, cases of multiple eigenvalues are not stable under perturbations. Therefore, the potential loss of generality due to the exclusion of such cases seems relatively minor to us.

The restriction on the supports of $\mathcal{F}_N$ and $\mathcal{F}$ implies that $\Omega_{jj} \le \bar{\omega}$ for all j. Note that $\Omega_{jj}/(2\pi)$ equals the spectral density at frequency zero of the quasi-difference $X_{jt} - \rho_j X_{j,t-1}$. Hence, Assumption A4 requires that such spectral densities are bounded. Furthermore, the assumption $\mu_1 > \mu_2 > \ldots$ implies that the distribution $\mathcal{F}$ cannot be concentrated at $\omega = 0$. In other words, a nontrivial fraction of the

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

series have spectral densities at frequency zero that are bounded away from zero, and hence, $\operatorname{tr}\Omega$ diverges to infinity at the same rate as N.

3. MAIN RESULT

Recall that $\hat{\lambda}_k$ and $\hat{F}_k$ denote the kth principal eigenvalue and eigenvector of matrix $\hat{\Sigma}$, respectively. Under necessary identifying restrictions (e.g., Bai and Ng, 2013), one would normally interpret these quantities as a measure of the strength of the kth factor in the data and an estimate of the kth factor itself. The following theorem shows that, if the data are persistent, such an interpretation may be deceiving.

THEOREM 1. *Let $N,T\to\infty$ at arbitrary relative rates. Then under Assumptions A1–A4, for any fixed positive integer k,*

(i) $\left|\hat{F}_k' d_k\right| \xrightarrow{P} 1$, *where* $d_k = (\varphi_k(1/T),\ldots,\varphi_k(T/T))/\sqrt{T}$ *and* $\varphi_k(s)$ *is the kth principal eigenfunction of* $K_{\mathcal{F}}$.
(ii) $\hat{\lambda}_k/T^2 \xrightarrow{P} \mu_k$, *where* μ_k *is the kth principal eigenvalue of* $K_{\mathcal{F}}$.
(iii) $\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma} \xrightarrow{P} \mu_k/\sum_{j=1}^{\infty}\mu_j$.

Although, in general, Theorem 1 does not give us closed form expressions for the limits of the normalized principal eigenvalues and eigenvectors of $\hat{\Sigma}$, its message is similar to that of theorems in OW. First, the PCA may be spurious in the sense that the estimated factors do not reflect cross-sectional linkages in the data. Second, the principal eigenvalues of the sample covariance matrix decay fast (μ_k, $k=1,2,\ldots$, being summable (see, e.g., Gohberg and Goldberg, 1981, Sect. IV.4) and thus, fast decreasing), creating an impression of high "explanatory" content of the "factors."

In the special case where there is no heterogeneity in local-to-unit roots, so that $\phi_i=\phi>0$ for all $i=1,\ldots,N$, the principal eigenvalues μ_k and eigenfunctions $\varphi_k(x)$ of $K_{\mathcal{F}}$ admit explicit expressions (see Section A1 of the Supplementary Material for a derivation)

$$\mu_k=\frac{1}{\phi^2+\omega^2},\quad \varphi_k(x)=h_k\left(a_k\cos(\omega_k x)+b_k\sin(\omega_k x)+c_k\right), \tag{3}$$

where the normalizing constant $h_k>0$ is chosen so that $\int_0^1\varphi_k^2(x)=1$,

$$a_k=\frac{1}{\omega_k}+\frac{\phi\omega_k}{\phi^2+\omega_k^2}-\frac{\cos\omega_k}{\omega_k},$$
$$b_k=\frac{\phi^2}{\phi^2+\omega_k^2}-\frac{\sin\omega_k}{\omega_k},$$
$$c_k=\frac{\phi^2}{\phi^2+\omega_k^2}\left(\frac{\cos\omega_k-1}{\omega_k}-\frac{\sin\omega_k}{\phi}\right),$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

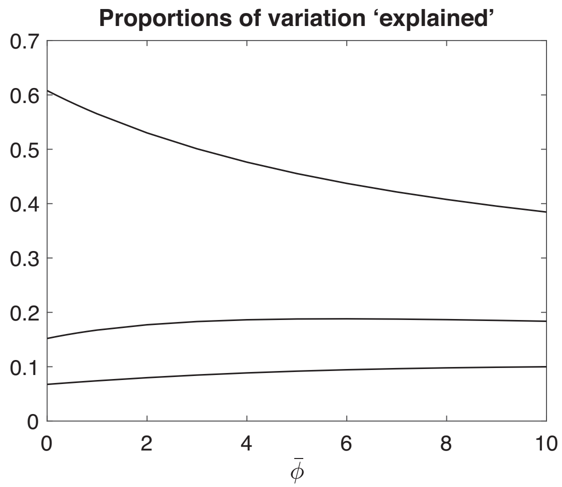

FIGURE 4. Proportions of the data variation "explained" by the first three spurious factors as functions of $\bar{\phi}$.

and ω_k is the kth smallest positive root of the equation

$$-2(\omega^2+\phi^2)+\omega(\phi^2-2\phi-\omega^2-2\omega^2/\phi)\sin\omega+2(\omega^2+\phi^2+\phi\omega^2)\cos\omega=0.$$

For example, if $\phi=1$, then the smallest three solutions of the latter equation are

$$\omega_1\approx 3.68,\quad \omega_2\approx 6.39,\quad \text{and}\quad \omega_3\approx 9.53.$$

For comparison, in the pure unit root case ($\phi=0$), these roots degenerate to π, 2π, and 3π.

In general, the local-to-unit roots may all be different, and the eigenfunctions of the operators corresponding to the individual-specific kernels $k_{\phi_i}(s,t)$ will not match. Then the eigenfunctions of the operator $K_{\mathcal{F}}$ corresponding to the weighted average of the individual-specific kernels will not admit explicit expressions. However, intuitively, since all individual-specific kernels have similar eigenstructures, we would still expect wave-like graphs of the eigenfunctions of $K_{\mathcal{F}}$ with the waves' frequency increasing as the corresponding eigenvalues decrease.

To illustrate Theorem 1 in such a general case, consider a simple scenario where $\Psi(L)=I_N$ so that $\mathcal{F}$ is concentrated at $\omega=1$, and where $\mathcal{F}$ is uniform on $\left[0,\bar{\phi}\right]$ with respect to the local-to-unity parameter ϕ. Figure 3 plots the principal eigenfunctions φ_1,φ_2, and φ_3, which we compute numerically for the case, where $\bar{\phi}=10$. These eigenfunctions are similar to the principal eigenfunctions $\sqrt{2}\cos(\pi x)$, $\sqrt{2}\cos(2\pi x)$, and $\sqrt{2}\cos(3\pi x)$ of the covariance kernel of the demeaned Brownian motion.

Figure 4 shows the proportions of variation "explained" by the first three spurious factors as functions of $0\le\bar{\phi}\le 10$. For $\bar{\phi}=0$ (the unit root case),

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

TABLE 1. Monte Carlo analysis of finite sample approximations based on Theorem 1

		$\sqrt{\mathbb{E}_{MC}(1-\lvert\hat{F}_k' d_k\rvert)^2}$			$\frac{1}{\mu_k}\sqrt{\mathbb{E}_{MC}\left(\frac{\hat{\lambda}_k}{T^2}-\mu_k\right)^2}$			$\frac{\sum\mu_i}{\mu_k}\sqrt{\mathbb{E}_{MC}\left(\frac{\hat{\lambda}_k}{\operatorname{tr}\hat{\Sigma}}-\frac{\mu_k}{\sum\mu_i}\right)^2}$		
T	N	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
20	20	0.228	0.379	0.501	0.568	0.440	0.453	0.219	0.240	0.289
20	50	0.110	0.181	0.298	0.373	0.354	0.371	0.146	0.180	0.209
20	100	0.039	0.090	0.185	0.289	0.306	0.331	0.111	0.140	0.162
20	1,000	0.007	0.019	0.039	0.186	0.236	0.278	0.045	0.046	0.073
50	20	0.217	0.354	0.463	0.475	0.344	0.294	0.220	0.234	0.249
50	50	0.110	0.174	0.263	0.296	0.247	0.224	0.149	0.179	0.189
50	100	0.032	0.065	0.132	0.218	0.190	0.179	0.111	0.131	0.141
50	1,000	0.003	0.007	0.013	0.093	0.100	0.106	0.040	0.048	0.049
100	20	0.209	0.349	0.464	0.516	0.321	0.278	0.239	0.238	0.258
100	50	0.090	0.160	0.264	0.295	0.222	0.207	0.154	0.178	0.189
100	100	0.036	0.068	0.128	0.211	0.165	0.152	0.116	0.136	0.140
100	1,000	0.003	0.006	0.010	0.072	0.066	0.066	0.038	0.046	0.048
1,000	20	0.219	0.363	0.470	0.451	0.290	0.245	0.239	0.243	0.256
1,000	50	0.094	0.163	0.260	0.282	0.210	0.172	0.162	0.188	0.179
1,000	100	0.036	0.068	0.124	0.193	0.149	0.137	0.117	0.140	0.142
1,000	1,000	0.003	0.005	0.008	0.061	0.051	0.046	0.037	0.046	0.045

Note: All reported numbers are square roots of the Monte Carlo means (denoted as $\mathbb{E}_{MC}$) of squared deviations from the probability limits, divided by these limits.

these proportions equal $6/(k\pi)^2$, $k=1,2,3,$ as in Theorem 1(iii) in OW. As $\bar{\phi}$ increases so that the local-to-unity roots may deviate from the unity further, the proportion of variation "explained" by the first factor decreases. For $\bar{\phi}=10,$ it equals 38%, which brings it closer to the explanatory power of Uhlig's first "factor" (see Figure 2) extracted from factorless persistent, but stationary, data.

To see how well the probability limits described by Theorem 1 approximate the corresponding finite sample quantities, we simulate $N\times T$ data from model (1) with $N=20,50,100,1{,}000$ and $T=20,50,100,1{,}000$. We set $\mu_X=0$, $\rho=\operatorname{diag}\{e^{-\phi_1/T},\ldots,e^{-\phi_N/T}\}$ with i.i.d. $\phi_i\sim U[0,10]$, $\Psi(L)=I_N$, i.i.d. $\varepsilon_{it}\sim\mathcal{N}(0,1)$, and draw the initial values from the stationary distribution. We repeat the Monte Carlo (MC) simulations 1,000 times.

Table 1 reports root mean squared deviations of $\lvert\hat{F}_k' d_k\rvert$, $\hat{\lambda}_k/T^2$, and $\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma}$ from their probability limits, 1, μ_k, and $\mu_k/\sum\mu_i$, respectively. These root mean squared deviations are divided by the corresponding probability limits, so they are expressed in relative terms. We see that the deviations quickly decrease when N increases. They are much less sensitive to the magnitude of T. This is consistent with Zhang, Pan, and Gao (2018) and Zhang, Gao, and Pan (2020) who established (under different assumptions than ours) that the rate of convergence of $\hat{\lambda}_k/T^2$ to μ_k is $N^{-1/2}$.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

Overall, the MC results suggest that the probability limits described in Theorem 1(i) and (iii) are reasonably well approximating the corresponding finite sample quantities when $N \geq 50$. The probability limit described in Theorem 1(ii) needs larger N to provide finite sample approximations of similar quality. We leave a formal analysis of the rates of convergence in our setting to future research.

4. PROOF OF THE MAIN RESULT

As mentioned in the introduction, our proof proceeds in several steps. First, we introduce matrix $\tilde{\Sigma}$ by replacing X in the definition of $\hat{\Sigma}$ with its long-run component. Second, we derive a result analogous to Theorem 1 for eigenvalues and eigenvectors of the deterministic part of $\tilde{\Sigma}$, $\mathbb{E}\tilde{\Sigma}$. Next, we show that adding the stochastic part of $\tilde{\Sigma}$ does not change the result, so that the theorem holds for $\tilde{\Sigma}$. Finally, we prove that the result does not change when $\tilde{\Sigma}$ is replaced by $\hat{\Sigma}$.

4.1. Step 1: Introducing $\tilde{\Sigma}$

Consider the following extension of the Beveridge–Nelson (BN) decomposition to nearly integrated series (1),

$$X_t = Z_t + \Psi^{**}(L)\varepsilon_t, \tag{4}$$

where

$$Z_t - \mu_X = \rho\,(Z_{t-1} - \mu_X) + \Psi\,(1)\,\varepsilon_t \tag{5}$$

with

$$Z_0 = X_0 - \Psi^{**}(L)\varepsilon_0, \tag{6}$$

and $\Psi^{**}(L) = \sum_{k=0}^{\infty} \Psi_k^{**} L^k$ with

$$\Psi_k^{**} = \sum_{j=1}^{k} \left(\rho^{k-j} - \rho^k\right)\Psi_j - \rho^k \sum_{j=k+1}^{\infty} \Psi_j.$$

The series Z_t can be interpreted as the “long-run component” of X_t. When $\rho = I_N$, the decomposition reduces to the standard BN one.

To see the validity of (4), use a standard recursive substitution in (1) and (5) to obtain

$$X_t - \mu_X = \sum\nolimits_{j=0}^{t-1} \rho^j \Psi(L)\varepsilon_{t-j} + \rho^t\,(X_0 - \mu_X) \text{ and} \tag{7}$$

$$Z_t - \mu_X = \sum\nolimits_{j=0}^{t-1} \rho^j \Psi(1)\varepsilon_{t-j} + \rho^t\,(Z_0 - \mu_X)\,. \tag{8}$$

Subtract (8) from (7), substitute $\rho^t\,(X_0 - Z_0)$ by $\rho^t \Psi^{**}(L)\varepsilon_0$, and verify that the right-hand side of the so obtained equality has form $\Psi^{**}(L)\varepsilon_t$ by matching the coefficients on different lags of ε_t.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

We will show that the first-order asymptotic behavior of principal eigenvalues and eigenvectors of $\hat{\Sigma}$ is not affected when X is replaced by its long-run component $Z = [Z_1, \ldots, Z_T]$. Intuitively, most variation in the data with near unit roots comes from the long-run component. Hence, one would expect an asymptotically negligible effect on $\hat{\Sigma}$ of such a replacement. This is similar to the unit root case, studied in OW.

By definition, $\hat{\Sigma} = MX'XM/N$, where M is the projection matrix on the space orthogonal to the T-dimensional vector of ones. In contrast to the unit root case, $MZ'ZM/N$ is not invariant with respect to the initial values Z_0, which are not eliminated by time averaging $Z \longmapsto ZM$.

To handle the effect of the initial values, we will treat components of Z_t having unit root (first N_1 components) and local-to-unity roots with positive local parameters (last $N - N_1$ components) separately. Denote the jth rows of $\Psi(L)$ and Z as $\Psi_{j\cdot}(L)$ and $Z_{j\cdot}$, respectively. By assumption, for any $j > N_1$, we have $X_{j0} - \mu_{Xj} = \sum_{i=0}^{\infty} \rho_j^i \Psi_{j\cdot}(L)\, \varepsilon_{-i}$. Using this in (6) yields $Z_{j0} - \mu_{Xj} = \sum_{i=0}^{\infty} \rho_j^i \Psi_{j\cdot}(1)\, \varepsilon_{-i}$. Combining this with (8), we see that for any $j > N_1$, Z_{jt} is a stationary process with the initial value Z_{j0} distributed according to its unconditional distribution.

The recursive substitution in the equation $Z_{jt} - \mu_{Xj} = \rho_j \left(Z_{j,t-1} - \mu_{Xj}\right) + \Psi_{j\cdot}(1)\, \varepsilon_t$ yields

$$Z_{jt} - \mu_{Xj} = \Psi_{j\cdot}(1) \sum_{i=0}^{\tau+t-1} \rho_j^i \varepsilon_{t-i} + \rho_j^{\tau+t} \left(Z_{j,-\tau} - \mu_{Xj}\right) \quad \textbf{(9)}$$

for any $\tau \geq 0$ and $j > N_1$. For $\tau \geq 0$ and $j \leq N_1$, let us define $Z_{j,-\tau}$ as $Z_{j0} - \Psi_{j\cdot}(1) \sum_{i=0}^{\tau-1} \varepsilon_{-i}$. With this definition, representation (9) holds for any $\tau \geq 0$ and all $j = 1, \ldots, N$, not only for $j > N_1$.

Let us set $\tau = T^3$, and let $\xi = \left[\varepsilon_{1-T^3}, \varepsilon_{2-T^3}, \ldots, \varepsilon_T\right]$. Finally, let U_j be a $T\left(T^2+1\right) \times T$ matrix such that

$$U_j' = \begin{pmatrix} \rho_j^{T^3} & \ldots & \rho_j^1 & \rho_j^0 & 0 & \ldots & 0 \\ \rho_j^{T^3+1} & \ldots & \rho_j^2 & \rho_j^1 & \rho_j^0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T^3+T-1} & \ldots & \rho_j^T & \rho_j^{T-1} & \rho_j^{T-2} & \ldots & \rho_j^0 \end{pmatrix}.$$

With this notation, we have

$$Z_{j\cdot} = \Psi_{j\cdot}(1)\, \xi U_j + \rho_j^{T^3} \left[\rho_j^1, \rho_j^2, \ldots, \rho_j^T\right] \left(Z_{j,-T^3} - \mu_{Xj}\right) + \mu_{Xj} \ell_T'$$

for all $j = 1, \ldots, N$, where ℓ_T is a T-dimensional vector of ones. Using this representation together with (4), we obtain

$$XM = \begin{bmatrix} \Psi_{1\cdot}(1)\, \xi U_1 \\ \vdots \\ \Psi_{N\cdot}(1)\, \xi U_N \end{bmatrix} M + X_{\text{ini}} M + \Psi^{**}(L)\, \varepsilon M, \quad \textbf{(10)}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

where

$X_{\text{ini}} = \rho^{T^3}\left[\rho^1\left(Z_{-T^3}-\mu_X\right),\ldots,\rho^T\left(Z_{-T^3}-\mu_X\right)\right]+\mu_X\ell_T'.$

We will show that, under the assumptions of Theorem 1, the behavior of a few of the largest eigenvalues and the corresponding eigenvectors of $\hat{\Sigma}$ is asymptotically equivalent to that of a few of the largest eigenvalues and corresponding eigenvectors of

$$\tilde{\Sigma} = \frac{1}{N}M\begin{bmatrix}\Psi_{1\cdot}(1)\,\xi U_1\\ \vdots \\ \Psi_{N\cdot}(1)\,\xi U_N\end{bmatrix}'\begin{bmatrix}\Psi_{1\cdot}(1)\,\xi U_1\\ \vdots \\ \Psi_{N\cdot}(1)\,\xi U_N\end{bmatrix}M. \tag{11}$$

Therefore, our proof strategy is as follows. First, establish statements (i)–(iii) of Theorem 1 for $\tilde{\lambda}_k, \tilde{F}_k, \tilde{\Sigma}$ instead of $\hat{\lambda}_k, \hat{F}_k, \hat{\Sigma}$ and then, prove that replacing "tildes" by "hats" does not affect the theorem's validity. Here, $\tilde{\lambda}_k$ and $\tilde{F}_k$ denote the kth principal eigenvalue and eigenvector of $\tilde{\Sigma}$ defined by (11).

4.2. Step 2: Analyzing $\mathbb{E}\tilde{\Sigma}$

As we will see below, the asymptotic behavior of the principal eigenstructure of $\tilde{\Sigma}$ is the same as that of its expected value, $\mathbb{E}\tilde{\Sigma}$. In this subsection, we analyze this expected value.

Write $\tilde{\Sigma}$ in the following form:

$$\tilde{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} MU_i'\xi'\Psi_{i\cdot}'(1)\,\Psi_{i\cdot}(1)\,\xi U_i M. \tag{12}$$

Taking expectation of the left- and right-hand sides yields

$$\mathbb{E}\tilde{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}\Psi_{i\cdot}(1)\,\Psi_{i\cdot}'(1)\,MU_i'U_iM = \frac{1}{N}\sum_{i=1}^{N}\Omega_{ii}MU_i'U_iM.$$

Let us denote the kth principal eigenvalue and eigenvector of $\mathbb{E}\tilde{\Sigma}$ as $\tilde{\mu}_k$ and $\tilde{\varphi}_k$, respectively. Further, denote the kth principal eigenvalue and eigenfunction of the integral operator $K_{\mathcal{F}}$ as μ_k and φ_k, and let $d_k = (\varphi_k(1/T),\ldots,\varphi_k(T/T))/\sqrt{T}$. We are going to show that, under Assumptions A1–A4, for any fixed positive integer k, as $N,T\to\infty$ at arbitrary relative rates,

$$\tilde{\mu}_k/T^2 \to \mu_k, \tag{13}$$

$$\left|\tilde{\varphi}_k'd_k\right| \to 1, \text{ and} \tag{14}$$

$$\tilde{\mu}_k/\operatorname{tr}\mathbb{E}\tilde{\Sigma} \to \mu_k/\sum\nolimits_{j=1}^{\infty}\mu_j. \tag{15}$$

To establish (13–15), we will prove that there exist approximating integral operators $K_{N,T}$ acting on the space of continuous functions on $[0,1]$ equipped with the supremum norm, $\|\cdot\|_{\sup}$, such that, on one hand, their principal eigenvalues

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

and eigenfunctions converge to those of $K_{\mathcal{F}}$, and on the other hand, the nonzero eigenvalues of $K_{N,T}$ coincide with those of $\mathbb{E}\tilde{\Sigma}/T^2$, and the corresponding eigenfunctions evaluated on the grid $1/T, 2/T, \ldots, T/T$ are eigenvectors of $\mathbb{E}\tilde{\Sigma}/T^2$. Convergences (13 and 14) immediately follow from the existence of such approximating operators. Convergence (15) would follow from such an existence and an auxiliary result (Lemma 4).

Let us now establish the existence of $K_{N,T}$ with the above described properties. Consider the stationary (unscaled) OU process $z_\phi(s)$, generated by stochastic differential equation

$$\mathrm{d}z_\phi(s) = -\phi z_\phi(s)\mathrm{d}s + \mathrm{d}W(s),$$

with the standard Wiener process $W(s)$ and $\phi > 0$. The initial observation $z_\phi(0)$ is drawn from the unconditional distribution of $z_\phi(s)$. As follows from a discussion in Section 2, the covariance kernel of demeaned $z_\phi(s)$ is $k_\phi(s,t)$.

Using (2), it is straightforward to verify that

$$k_\phi(s,t) = a_\phi(s,t) - b_\phi(t) - b_\phi(s) + c_\phi,$$

where

$$a_\phi(s,t) = \begin{cases} \left(e^{-\phi|t-s|} - 1\right)/(2\phi), & \text{if } \phi > 0, \\ -|t-s|/2, & \text{if } \phi = 0, \end{cases}$$

$$b_\phi(t) = \begin{cases} \left(2 - \phi - e^{-\phi t} - e^{-\phi(1-t)}\right)/\left(2\phi^2\right), & \text{if } \phi > 0 \\ -t^2/2 + t/2 - 1/4, & \text{if } \phi = 0, \end{cases}$$

$$c_\phi = \begin{cases} (e^{-\phi} - 1 + \phi - \phi^2/2)/\phi^3, & \text{if } \phi > 0, \\ -1/6, & \text{if } \phi = 0. \end{cases}$$

Let

$$U_\phi = \begin{pmatrix} e^{-T^2\phi} & e^{-\left(T^2+1/T\right)\phi} & \ldots & e^{-\left(T^2+1-1/T\right)\phi} \\ \vdots & \vdots & & \vdots \\ e^{-\phi/T} & e^{-2\phi/T} & \ldots & e^{-\phi} \\ 1 & e^{-\phi/T} & \ldots & e^{-\phi(T-1)/T} \\ 0 & 1 & \ldots & e^{-\phi(T-2)/T} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Note that $\mathbb{E}\tilde{\Sigma} = \frac{1}{N}\sum_{k=1}^N \Omega_{kk} M U_k' U_k M$, where $U_k = U_{\phi_k}$. The following elementary lemma is established in Section A2 of the Supplementary Material.

LEMMA 2. *For any $\phi \geq 0$,*

$$\left(MU_\phi' U_\phi M\right)_{ij}/T = k_{\phi,T}\left(s_i, t_j\right) \quad \textbf{(16)}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

with $s_i = i/T$, $t_j = j/T$, *and*

$$k_{\phi,T}(s,t) = \omega_{\phi 1,T} a_\phi(s,t) - \omega_{\phi 2,T}\left(b_\phi(t) + b_\phi(s)\right) + d_{\phi,T} - e_{\phi,T}(s,t),$$

where a_ϕ *and* b_ϕ *are as defined above, whereas* $\omega_{\phi 1,T}$, $\omega_{\phi 2,T}$, $d_{\phi,T}$, *and* $e_{\phi,T}(s,t)$ *are as follows. For* $\phi > 0$,

$$\omega_{\phi 1,T} = \frac{2\phi}{T\left(1-e^{-2\phi/T}\right)}, \qquad \omega_{\phi 2,T} = \frac{2\phi^2}{T^2\left(1-e^{-2\phi/T}\right)\left(e^{\phi/T}-1\right)},$$

$$d_{\phi,T} = \frac{2e^{-\phi/T}\left(e^{-\phi}-1\right) + T\left(1-e^{-2\phi/T}\right) - T^2\left(1-e^{-\phi/T}\right)^2}{T^3\left(1-e^{-2\phi/T}\right)\left(1-e^{-\phi/T}\right)^2},$$

and

$$\begin{aligned} e_{\phi,T}(s,t) = & \frac{2-e^{-\phi s}-e^{-\phi t}}{T^2\left(1-e^{-2\phi/T}\right)} - \frac{2(e^{\phi/T}-1-\phi/T)}{T(1-e^{-2\phi/T})(e^{\phi/T}-1)} + e^{-2\phi T^2}\left(\frac{e^{-\phi(t+s)}}{T\left(1-e^{-2\phi/T}\right)}\right. \\ & \left. - \frac{\left(e^{-\phi s}+e^{-\phi t}\right)\left(1-e^{-\phi}\right)}{T^2\left(1-e^{-2\phi/T}\right)\left(e^{\phi/T}-1\right)} + \frac{\left(1-e^{-\phi}\right)^2}{T^3\left(1-e^{-2\phi/T}\right)\left(e^{\phi/T}-1\right)^2}\right). \end{aligned}$$

For $\phi = 0$,

$$\begin{aligned} \omega_{01,T} &= \omega_{02,T} = 1, \\ d_{0,T} &= (T+1)(2T+1)/\left(6T^2\right) - 1/2, \text{ and} \\ e_{0,T}(s,t) &= (s+t)/(2T). \end{aligned}$$

Now consider integrated kernels

$$\begin{aligned} k_{N,T}(s,t) &= \int \omega k_{\phi,T}(s,t)\,\mathrm{d}\mathcal{F}_N(\omega,\phi) \text{ and} \\ k_{\mathcal{F}}(s,t) &= \int \omega k_\phi(s,t)\,\mathrm{d}\mathcal{F}(\omega,\phi), \end{aligned}$$

where $\mathcal{F}_N(\omega,\phi)$ is the empirical distribution function of the pairs (Ω_{ii},ϕ_i), $i = 1,\ldots,N$, and $\mathcal{F}(\omega,\phi)$ is its weak limit as $N\to\infty$. By definition, $k_{\mathcal{F}}(s,t)$ is the kernel of the operator $K_{\mathcal{F}}$.

Let $K_{N,T}$ be approximating operators, acting on $x \in C[0,1]$ as follows:

$$\begin{aligned} \left(K_{N,T}x\right)(s) &= \frac{1}{T}\sum\nolimits_{j=1}^{T} k_{N,T}\left(s,t_j\right)x(t_j) \\ &= \frac{1}{T}\sum\nolimits_{j=1}^{T}\int \omega k_{\phi,T}\left(s,t_j\right)x(t_j)\mathrm{d}\mathcal{F}_N(\omega,\phi). \end{aligned}$$

Identity (16) implies that the eigenvalues of $\mathbb{E}\tilde{\Sigma}/T^2$ are also eigenvalues of $K_{N,T}$. Moreover, if $x(t)$ is an eigenfunction of $K_{N,T}$, then $(x(t_1),\ldots,x(t_T))'$ is an eigenvector of $\mathbb{E}\tilde{\Sigma}/T^2$. Vice versa, if $(x_1,\ldots,x_T)'$ is an eigenvector of $\mathbb{E}\tilde{\Sigma}/T^2$, then there exists $x \in C[0,1]$ with $x\left(t_j\right) = x_j$ such that x is an eigenfunction of $K_{N,T}$. In

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

other words, the spectral properties of $K_{N,T}$ and $\mathbb{E}\tilde{\Sigma}/T^2$ are essentially the same, even though the first is an operator in $C[0,1]$ while the second is a $T \times T$ matrix.

It remains to prove that the principal eigenvalues and eigenfunctions of $K_{N,T}$ converge to those of $K_{\mathcal{F}}$. As discussed in the introduction, we need to establish the pointwise convergence $K_{N,T} \to K_{\mathcal{F}}$ and the collective compactness of the sequence of operators $\{K_{N,T} : N,T = 1,2,\dots\}$ (see Anselone, 1967 and the discussion below for the definition of collective compactness). After establishing these facts, we show how they imply the convergence of the principal eigenvalues and eigenfunctions.

4.2.1. *Pointwise Convergence.* Let x be an arbitrary function from $C[0,1]$. In this subsection, we show that $\|K_{N,T}x - K_{\mathcal{F}}x\|_{\sup} \to 0$ as $N,T \to \infty$ at arbitrary relative rates. In other words, $\forall \epsilon > 0$ $\exists N_0, T_0$ s.t. $\forall N > N_0$ and $T > T_0$, $\|K_{N,T}x - K_{\mathcal{F}}x\|_{\sup} < \epsilon$. Without loss of generality, we assume that $\|x\|_{\sup} \le 1$.

Let $\phi_\epsilon > 0$ and $N_2 > 0$ be such that

$$\int \mathbf{1}\{\phi \ge \phi_\epsilon\} \mathrm{d}\mathcal{F}(z,\phi) < \epsilon/(3\bar{\omega}) \text{ and } \int \mathbf{1}\{\phi \ge \phi_\epsilon\} \mathrm{d}\mathcal{F}_N(z,\phi) < \epsilon/(24\bar{\omega})$$

for all $N > N_2$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function. For any $\epsilon > 0$, the displayed inequalities can be satisfied by choosing ϕ_ϵ sufficiently large because $\mathcal{F}(z,\phi)$ is a cumulative distribution function of a proper probability distribution and $\mathcal{F}_N$ weakly converges to $\mathcal{F}$ as $N \to \infty$. In fact, by Assumption A4, any ϕ from the supports of $\mathcal{F}(z,\phi)$ and $\mathcal{F}_N(z,\phi)$ satisfies $\phi \le \bar{\phi}$. In particular, we can set $\phi_\epsilon = \bar{\phi}$. However, in this subsection, we do not need to (and will not) assume the boundedness of the supports of $\mathcal{F}(z,\phi)$ and $\mathcal{F}_N(z,\phi)$ with respect to ϕ.

Let $f_\epsilon(\phi)$ be a continuously differentiable function of $\phi \ge 0$, such that $|f_\epsilon(\phi)| \le 1$, $f_\epsilon(\phi) = 1$ for $\phi \le \phi_\epsilon$, and $f_\epsilon(\phi) = 0$ for $\phi \ge 2\phi_\epsilon$. We split the difference $K_{N,T}x - K_{\mathcal{F}}x$ into three parts, $P_1 + P_2 + P_3$, where

$$P_1 = -\int_0^1 \int \omega(1 - f_\epsilon(\phi)) k_\phi(s,t) x(t) \mathrm{d}\mathcal{F}(\omega,\phi)\, \mathrm{d}t,$$

$$P_2 = \frac{1}{T}\sum_{j=1}^T \int \omega(1 - f_\epsilon(\phi)) k_{\phi,T}(s,t_j) x(t_j) \mathrm{d}\mathcal{F}_N(\omega,\phi),$$

and $P_3 = K_{N,T}x - K_{\mathcal{F}}x - P_1 - P_2$ is the remainder. To analyze P_1 and P_2, we need the following lemma. Its proof is in Section A6 of the Supplementary Material.

LEMMA 3. *Kernels $k_\phi(s,t)$ and $k_{\phi,T}(s,t)$ are bounded by absolute value uniformly in $\phi \ge 0$. Specifically,*

$$\sup_{\phi \ge 0} \max_{s,t \in [0,1]^2} |k_\phi(s,t)| \le 1 \textit{ and } \sup_{\phi \ge 0} \sup_{T \ge 1} \max_{s,t \in [0,1]^2} |k_{\phi,T}(s,t)| \le 8.$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

Lemma 3 implies that, for all $N > N_2$,

$$|P_1| \leq \int_0^1 \int \left|\omega\left(1-f_\epsilon(\phi)\right)k_\phi(s,t)x(t)\right| \mathrm{d}\mathcal{F}(\omega,\phi)\,\mathrm{d}t$$
$$\leq \int \bar{\omega}\mathbf{1}\{\phi \geq \phi_\epsilon\}\,\mathrm{d}\mathcal{F}(\omega,\phi) < \bar{\omega}\epsilon/(3\bar{\omega}) = \epsilon/3. \quad \mathbf{(17)}$$

Similarly, for all $N > N_2$,

$$|P_2| \leq \frac{1}{T}\sum\nolimits_{j=1}^{T}\int \left|\omega\left(1-f_\epsilon(\phi)\right)k_{\phi,T}\left(s,t_j\right)x(t_j)\right| \mathrm{d}\mathcal{F}_N(\omega,\phi)$$
$$\leq \int 8\bar{\omega}\mathbf{1}\{\phi \geq \phi_\epsilon\}\,\mathrm{d}\mathcal{F}_N(\omega,\phi) < 8\bar{\omega}\epsilon/(24\bar{\omega}) = \epsilon/3. \quad \mathbf{(18)}$$

To establish the pointwise convergence of $K_{N,T}$ to $K_{\mathcal{F}}$, it remains to prove that $|P_3| < \epsilon/3$ for all sufficiently large T and N.

Consider the following decomposition:

$$P_3 = y_1(s) + y_2(s) + y_3(s),$$

where

$$y_1(s) = \int_0^1 \int \omega f_\epsilon(\phi)\,k_\phi(s,t)\,x(t)\mathrm{d}\left(\mathcal{F}_N(\omega,\phi) - \mathcal{F}(\omega,\phi)\right)\mathrm{d}t,$$

$$y_2(s) = \int_0^1 \int \omega f_\epsilon(\phi)\left(k_{\phi,T}(s,t) - k_\phi(s,t)\right)x(t)\mathrm{d}\mathcal{F}_N(\omega,\phi)\,\mathrm{d}t,$$

and

$$y_3(s) = \frac{1}{T}\sum\nolimits_{j=1}^{T}\int \omega f_\epsilon(\phi)\,k_{\phi,T}\left(s,t_j\right)x(t_j)\mathrm{d}\mathcal{F}_N(\omega,\phi)$$
$$- \int_0^1 \int \omega f_\epsilon(\phi)\,k_{\phi,T}(s,t)\,x(t)\mathrm{d}\mathcal{F}_N(\omega,\phi)\,\mathrm{d}t.$$

Note that $f_\epsilon(\phi)\,k_\phi(s,t)$, viewed as a function of s is Lipschitz with the Lipschitz constant that depends on ϵ, but not on ϕ and t. Therefore, function $y_1(s)$ is Lipschitz on $s \in [0,1]$ with the Lipschitz constant that does not depend on N. Furthermore, for each fixed $s \in [0,1]$, it converges to 0 as $N \to \infty$ because $\mathcal{F}_N$ weakly converges to $\mathcal{F}$ and $\int_0^1 \omega f_\epsilon(\phi)\,k_\phi(s,t)\,x(t)\mathrm{d}t$ is a bounded continuous function on $(\omega,\phi) \in [0,\bar{\omega}] \times [0,\infty)$. Therefore, $y_1(s)$ converges to zero uniformly on $[0,1]$.

Next, the uniform convergence of $y_2(s)$ to zero would follow from the convergence

$$\sup_{\phi\geq 0}\sup_{s,t\in[0,1]^2}\left|f_\epsilon(\phi)\left(k_\phi(s,t) - k_{\phi,T}(s,t)\right)\right| \to 0 \quad \mathbf{(19)}$$

as $T \to \infty$. To see that (19) holds, consider the decomposition

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

$$\begin{aligned}&f_\epsilon(\phi)\left(k_\phi(s,t)-k_{\phi,T}(s,t)\right)\\&=f_\epsilon(\phi)\left(1-\omega_{\phi 1,T}\right)a_\phi(s,t)-f_\epsilon(\phi)\left(1-\omega_{\phi 2,T}\right)\left(b_\phi(t)+b_\phi(s)\right)\\&\quad+f_\epsilon(\phi)\left(c_\phi-d_{\phi,T}\right)+f_\epsilon(\phi)e_{\phi,T}(s,t).\end{aligned}$$

As follows from the proof of Lemma 3, $\left|a_\phi(s,t)\right|$ and $\left|b_\phi(t)+b_\phi(s)\right|$ are bounded uniformly in $\phi\geq 0$. On the other hand, $1-\omega_{\phi 1,T}\to 0$, $1-\omega_{\phi 2,T}\to 0$, and $c_\phi-d_{\phi,T}\to 0$ uniformly on $\phi\in[0,2\phi_\epsilon]$ (the support of f_ϵ). Hence, the first three terms on the right-hand side of the above display converge to zero uniformly in ϕ,s, and t.

For the last term, we have (see a derivation in Section A5 of the Supplementary Material)

$$\left|f_\epsilon(\phi)e_{\phi,T}(s,t)\right|\leq\frac{4}{T}. \tag{20}$$

Hence, $\left|f_\epsilon(\phi)e_{\phi,T}(s,t)\right|\to 0$ uniformly over $s,t\in[0,1]^2$ and $\phi\geq 0$.

Turning to the analysis of $y_3(s)$, let us define, similar to Anselone (1967, p. 9), bounded linear functionals

$$\psi x=\int_0^1 x(t)\,\mathrm{d}t \text{ and } \psi_T x=\frac{1}{T}\sum\nolimits_{j=1}^{T}x(j/T).$$

Functionals ψ_T converge to ψ uniformly on totally bounded subsets of $C[0,1]$. We have

$$y_3(s)=-\int\omega\left((\psi-\psi_T)g_{s\phi}\right)\mathrm{d}\mathcal{F}_N(\omega,\phi),$$

where

$$g_{s\phi}(t)=f_\epsilon(\phi)k_{\phi,T}(s,t)x(t).$$

The family of functions $\left\{g_{s\phi}(t):s\in[0,1],\phi\geq 0\right\}$ is bounded and equicontinuous. Hence, by Arzela–Ascoli lemma, this family forms a totally bounded set in $C[0,1]$. Therefore, $(\psi-\psi_T)g_{s\phi}$ converges to zero uniformly over $(s,\phi)\in[0,1]\times[0,\infty)$. This yields the uniform convergence of $y_3(s)$ to zero.

To summarize, functions y_1,y_2,y_3 converge to zero as $N,T\to\infty$ at arbitrary relative rates. Hence, there exist N_3,T_0 such that for all $N>N_3$ and $T>T_0$, $\|P_3\|_{\sup}<\epsilon/3$. Combining this with (17) and (18), and setting $N_0=\max\{N_2,N_3\}$, we see that, for all $N>N_0$ and $T>T_0$, $\left\|K_{T,N}x-K_{\mathcal{F}}x\right\|_{\sup}<\epsilon$, which finishes the proof of the pointwise convergence $K_{T,N}\to K_{\mathcal{F}}$.

4.2.2. *Collective Compactness.* The set of operators $\left\{K_{N,T}:N,T=1,2,\dots\right\}$ is called collectively compact if the subset $\left\{K_{N,T}x:N,T=1,2,\dots,\|x\|_{\sup}\leq 1\right\}$ of $C[0,1]$ is totally bounded. Recall that a set S is totally bounded if and only if for any $\epsilon>0$, there exists a finite set $\{x_1,\dots,x_m\}$, such that for any $x\in S$, $\min_{1\leq i\leq m}\|x-x_i\|<\epsilon$.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

We have $K_{N,T}x = K_{N,T}^{(1)}x + P_2,$ where

$$\left(K_{N,T}^{(1)}x\right)(s) = \frac{1}{T}\sum\nolimits_{j=1}^{T}\int \omega f_{\epsilon}(\phi)\, k_{\phi,T}\left(s,t_j\right)x(t_j)\mathrm{d}\mathcal{F}_N(\omega,\phi),$$

with $f_{\epsilon}(\phi)$ and P_2 defined in the previous subsection. As we have seen above, $\|P_2\|_{\sup} < \epsilon/3$. Therefore, to establish the collective compactness of $K_{N,T}$, it is sufficient to show that $\forall\epsilon$, the set $\left\{K_{N,T}^{(1)}x : N,T = 1,2,\ldots,\ \|x\|_{\sup} \leq 1\right\}$ is totally bounded. But such total boundedness follows from the Arzela–Ascoli lemma and the fact that functions $g_{\phi,t}(s) = f_{\epsilon}(\phi)\, k_{\phi,T}(s,t)\, x(t)$ are bounded and equicontinuous for $\phi \geq 0$ and $t \in [0,1]$.

4.2.3. *Convergence of the Principal Eigenvalues and Eigenfunctions.* Recall that we denote the eigenvalues of $K_{\mathcal{F}}$ as $\mu_1, \mu_2, \ldots$ and corresponding eigenfunctions as $\varphi_1, \varphi_2, \ldots$. By Assumption A4, these eigenvalues are simple so that $\mu_1 > \mu_2 > \ldots$. Denote the eigenvalues of $K_{N,T}$ as $\mu_{1,NT} \geq \mu_{2,NT} \geq \ldots$ and corresponding eigenfunctions as $\varphi_{1,NT}, \varphi_{2,NT}, \ldots$. Let us show that, for any fixed k, $\mu_{k,NT} \to \mu_k$ and $\varphi_{k,NT} \to \varphi_k$, the latter convergence being in $C[0,1]$.[2]

Take $k = 1$. Since $\mu_{1,NT}$, $N,T = 1,2,3,\ldots$ forms a bounded sequence, there exists a converging sub-sequence $\mu_{1,N_jT_j} \to m_1$. By Lemmas 2.5 and 2.6 of Anselone (1967), $\varphi_{1,N_jT_j} \to y_1$ and m_1, y_1 is an eigenvalue–eigenfunction pair for $K_{\mathcal{F}}$. On the other hand, it must be the case that $m_1 = \mu_1$. Indeed, if $m_1 < \mu_1$, then by Theorem 2.2 of Anselone (1967), μ_1 must belong to the resolvent set of $K_{\mathcal{F}}$, which is not true. Hence, any convergent sub-sequence of $\mu_{1,NT}, N,T = 1,2,3,\ldots$ converges to μ_1 and the sub-sequence of corresponding eigenfunctions converges to φ_1. Therefore, $\mu_{1,NT} \to \mu_1$ and $\varphi_{1,NT} \to \varphi_1$. Similar convergences for any positive integer k follow by mathematical induction.

This finishes our proof of equations (13) and (14). To establish (15), we need the following lemma. Its proof is in Section A6 of the Supplementary Material.

LEMMA 4. *For any fixed positive integer J,*

$$\sum\nolimits_{j=J+1}^{T}\tilde{\mu}_j \leq T^2\,\mathrm{tr}\,\Omega/(9JN)$$

for all sufficiently large T. Furthermore, for any fixed positive integer k, there exists a constant $C_k > 0$ such that

$$\tilde{\mu}_k \geq C_k T^2\,\mathrm{tr}\,\Omega/N \tag{21}$$

for all sufficiently large T.

Convergence (15) now follows from (13), Lemma 4, and the fact that, by Assumption A4, $\mathrm{tr}\,\Omega/N \leq \bar{\omega} < \infty$. Indeed, in light of (13), it is sufficient to prove that $\mathrm{tr}\,\mathbb{E}\tilde{\Sigma}/T^2 \to \sum_{j=1}^{\infty}\mu_j$. Consider an arbitrary $\epsilon > 0$, and let $J_\epsilon \in \mathbb{Z}_+$ be such that

[2]The eigenfunctions are defined up to sign, and we assume that it is chosen so that $\int \varphi_{k,NT}(s)\varphi_k(s)\mathrm{d}s > 0$.

$J_\epsilon > \bar{\omega}/(3\epsilon)$ and $\sum_{j=J_\epsilon}^{\infty} \mu_j < \epsilon/3.$

Then, by Lemma 4 and the fact that $\operatorname{tr}\Omega/N \le \bar{\omega} < \infty$, we have for all sufficiently large T

$$\sum_{j=J_\epsilon+1}^{T} \tilde{\mu}_j/T^2 < \epsilon/3.$$

Therefore, for such T, we have

$$\left|\operatorname{tr}\mathbb{E}\tilde{\Sigma}/T^2 - \sum_{j=1}^{\infty} \mu_j\right| \le \left|\sum_{j=1}^{J_\epsilon} \tilde{\mu}_j/T^2 - \sum_{j=1}^{J_\epsilon} \mu_j\right| + 2\epsilon/3.$$

On the other hand, by Lemma 4, the first term on the right-hand side is smaller than $\epsilon/3$ for all sufficiently large T. Hence, we have proven the convergence $\operatorname{tr}\mathbb{E}\tilde{\Sigma}/T^2 \to \sum_{j=1}^{\infty} \mu_j$, and thus, have established (15). Note that this proof does not rely on (21), which will be used later.

4.3. Step 3: Proof of Theorem 1 for $\tilde{\lambda}_k, \tilde{F}_k, \tilde{\Sigma}$

Given equations 13–15, Theorem 1 with $\hat{\lambda}_k$, $\hat{F}_k$, and $\hat{\Sigma}$ replaced by $\tilde{\lambda}_k$, $\tilde{F}_k$, and $\tilde{\Sigma}$ would follow from

$$\left|\tilde{F}_k'\tilde{\varphi}_k\right| \xrightarrow{P} 1,\ \tilde{\lambda}_k/\tilde{\mu}_k - 1 \xrightarrow{P} 0, \text{ and } \tilde{\lambda}_k/\operatorname{tr}\tilde{\Sigma} - \tilde{\mu}_k/\operatorname{tr}\mathbb{E}\tilde{\Sigma} \xrightarrow{P} 0. \quad \textbf{(22)}$$

Let us now prove the convergencies in (22).

We start from the case $k = 1$, and then extend the analysis to $k > 1$ using mathematical induction. Let us represent $\tilde{F}_1$ in the form

$$\tilde{F}_1 = \sum_{q=1}^{T} \alpha_q \tilde{\varphi}_q = \sum_{q=1}^{T-1} \alpha_q \tilde{\varphi}_q,$$

where the latter equality holds because $\tilde{F}_1$ must be orthogonal to $l_T/\sqrt{T} = \tilde{\varphi}_T$, which is an eigenvector of $\tilde{\Sigma}$ and of $\mathbb{E}\tilde{\Sigma}$ corresponding to the zero eigenvalue (we remind the reader that l_T denotes the T-dimensional vector of ones). The above representation and the definition (12) of $\tilde{\Sigma}$ yield

$$\tilde{\lambda}_1 = \sum_{r,q=1}^{T-1} \alpha_r \alpha_q \frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \quad \textbf{(23)}$$

where

$$A^{(i)} = M U_i' \xi' \Psi_{i\cdot}' (1) \Psi_{i\cdot} (1) \xi U_i M.$$

Let K be a fixed positive integer. Represent $\tilde{\lambda}_1$ in the form $\tilde{\lambda}_{11} + \tilde{\lambda}_{12} + \tilde{\lambda}_{13}$, where

$$\tilde{\lambda}_{11} = \sum_{r,q=1}^{K} \alpha_r \alpha_q \frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \quad \textbf{(24)}$$

$$\tilde{\lambda}_{12} = \sum_{r,q=K+1}^{T-1} \alpha_r \alpha_q \frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \quad \textbf{(25)}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

and

$$\tilde{\lambda}_{13} = 2\sum_{r=1}^{K}\sum_{q=K+1}^{T-1}\alpha_r\alpha_q\frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q. \tag{26}$$

Note that

$$\tilde{\lambda}_1 \le \left(\tilde{\lambda}_{11}^{1/2} + \tilde{\lambda}_{12}^{1/2}\right)^2. \tag{27}$$

To analyze $\tilde{\lambda}_{1i}$, we need the following elementary lemma, established in Section A7 of the Supplementary Material.

LEMMA 5. *Suppose Assumption A1 holds. Let a,b,c,d and A,B be any deterministic T-dimensional vectors and $N\times N$ matrices, respectively. Then*

$$\mathbb{E}\left(a'\varepsilon' A\varepsilon b\right) = a'b\,\mathrm{tr}A, \text{ and} \tag{28}$$

$$\begin{aligned}&\left|Cov\left(a'\varepsilon' A\varepsilon b, c'\varepsilon' B\varepsilon d\right) - \left(a'c\right)\left(b'd\right)\mathrm{tr}\left(A'B\right) - \left(a'd\right)\left(b'c\right)\mathrm{tr}\left(AB\right)\right| \\ &\qquad \le 2\varkappa_4\sum_{i=1}^{N_\varepsilon}\sum_{t=1}^{T}\left|A_{ii}B_{ii}a_tb_tc_td_t\right|,\end{aligned} \tag{29}$$

where $a_t, b_t, c_t,$ and d_t are the tth components of vectors $a, b, c,$ and d.

Consider the inner sum in the expression (24) for $\tilde{\lambda}_{11}$. Equation (28) of Lemma 5 yields

$$\begin{aligned}\mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q &= \frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}_r' MU_i'U_iM\tilde{\varphi}_q\,\mathrm{tr}\left(\Psi_{i\cdot}'(1)\Psi_{i\cdot}(1)\right) \\ &= \tilde{\varphi}_r'\left(\frac{1}{N}\sum_{i=1}^{N}\Omega_{ii}MU_i'U_iM\right)\tilde{\varphi}_q \\ &= \tilde{\varphi}_r'\mathbb{E}\tilde{\Sigma}\tilde{\varphi}_q = \tilde{\mu}_r\delta_{rq}.\end{aligned}$$

Further,

$$Var\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov\left(\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q, \tilde{\varphi}_r' A^{(j)}\tilde{\varphi}_q\right).$$

Equation (29) of Lemma 5 yields

$$\begin{aligned}&Cov\left(\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q, \tilde{\varphi}_r' A^{(j)}\tilde{\varphi}_q\right) \\ &\le \tilde{\varphi}_r' MU_i'U_jM\tilde{\varphi}_r\tilde{\varphi}_q' MU_i'U_jM\tilde{\varphi}_q\,\mathrm{tr}\left(\Psi_{i\cdot}'(1)\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\Psi_{j\cdot}(1)\right) \\ &\quad + \tilde{\varphi}_r' MU_i'U_jM\tilde{\varphi}_q\tilde{\varphi}_q' MU_i'U_jM\tilde{\varphi}_r\,\mathrm{tr}\left(\Psi_{i\cdot}'(1)\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\Psi_{j\cdot}(1)\right) \\ &\quad + 2\varkappa_4\left\|U_iM\tilde{\varphi}_r\right\|\left\|U_iM\tilde{\varphi}_q\right\|\left\|U_jM\tilde{\varphi}_r\right\|\left\|U_jM\tilde{\varphi}_q\right\|\sum_{s=1}^{N}\left(\Psi_{is}(1)\Psi_{js}(1)\right)^2.\end{aligned}$$

Section A8 of the Supplementary Material proves the following inequality:

$$\sup_{\rho_i\in[0,1]}\left\|U_iM\right\|^2 \le \sup_{\rho_i\in[0,1]}\mathrm{tr}\,MU_i'U_iM \le 2T^2. \tag{30}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

This inequality and the above bound for $Cov\left(\tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \tilde{\varphi}_r' A^{(j)} \tilde{\varphi}_q\right)$ yield

$$Cov\left(\tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \tilde{\varphi}_r' A^{(j)} \tilde{\varphi}_q\right) \leq 8T^4 \left(\left(\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\right)^2 + \varkappa_4 \sum_{s=1}^{N_\varepsilon} \left(\Psi_{is}(1)\Psi_{js}(1)\right)^2\right)$$

and

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q\right) \leq \frac{8T^4}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\left(\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\right)^2 + \varkappa_4 \sum_{s=1}^{N}\left(\Psi_{is}(1)\Psi_{js}(1)\right)^2\right].$$

We have

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{N_\varepsilon}\left(\Psi_{is}(1)\Psi_{js}(1)\right)^2 = \sum_{s=1}^{N_\varepsilon}\sum_{i=1}^{N}\left(\Psi_{is}(1)\right)^2\sum_{j=1}^{N}\left(\Psi_{js}(1)\right)^2$$
$$= \sum_{s=1}^{N}\left(\left(\Psi'(1)\Psi(1)\right)_{ss}\right)^2 \leq \operatorname{tr}\left[\left(\Psi'(1)\Psi(1)\right)^2\right] = \operatorname{tr}\left[\left(\Psi(1)\Psi'(1)\right)^2\right].$$

Therefore,

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q\right) \leq \frac{8T^4}{N^2}(1+\varkappa_4)\operatorname{tr}\left[\left(\Psi(1)\Psi'(1)\right)^2\right] \tag{31}$$

$$= \frac{8T^4}{N^2}(1+\varkappa_4)\operatorname{tr}\left[\Omega^2\right] = o(1)\frac{T^4}{N^2}(\operatorname{tr}\Omega)^2,$$

where the last equality follows from Assumption A3 because $\operatorname{tr}\Omega^2 \leq \|\Omega\| \operatorname{tr}\Omega$. By Chebyshev's inequality,

$$\tilde{\lambda}_{11} = \sum_{r=1}^{K} \alpha_r^2 \tilde{\mu}_r + o_{\mathrm{P}}(1) T^2 \operatorname{tr}\Omega/N. \tag{32}$$

Next, consider $\tilde{\lambda}_{12}$. The definition of $A^{(i)}$ yields

$$\tilde{\lambda}_{12} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{r=K+1}^{T-1}\alpha_r \Psi_{i\cdot}(1)\,\xi\, U_i M \tilde{\varphi}_r\right)^2.$$

By the Cauchy–Schwarz inequality,

$$\tilde{\lambda}_{12} \leq \frac{1}{N}\sum_{i=1}^{N}\sum_{r=K+1}^{T-1}\alpha_r^2\sum_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\,\xi\, U_i M\tilde{\varphi}_r\right)^2$$
$$\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\,\xi\, U_i M\tilde{\varphi}_r\right)^2.$$

Lemma 5 yields

$$\mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\sum_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\,\xi\, U_i M\tilde{\varphi}_r\right)^2 = \sum_{r=K+1}^{T-1}\tilde{\mu}_r \tag{33}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

and

$$Var\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\xi U_i M\tilde{\varphi}_r\right)^2\right)$$
$$=\frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{r,q=K+1}^{T-1}Cov\left(\left(\Psi_{i\cdot}(1)\xi U_i M\tilde{\varphi}_r\right)^2,\left(\Psi_{j\cdot}(1)\xi U_j M\tilde{\varphi}_q\right)^2\right)$$
$$\leq\frac{2}{N^2}\sum_{i,j=1}^{N}\sum_{r,q=K+1}^{T-1}\|U_i M\tilde{\varphi}_r\|^2\|U_j M\tilde{\varphi}_q\|^2\left[\left(\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\right)^2+\varkappa_4\sum_{s=1}^{T}\Psi_{is}^2(1)\Psi_{js}^2(1)\right].$$

Note that

$$\sum_{r,q=K+1}^{T-1}\|U_i M\tilde{\varphi}_r\|^2\left\|U_j M\tilde{\varphi}_q\right\|^2\leq \operatorname{tr}\left(MU_i'U_iM\right)\operatorname{tr}\left(MU_j'U_jM\right)\leq 4T^4,$$

where the latter inequality follows from (30).

Therefore,

$$Var\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\xi U_i M\tilde{\varphi}_r\right)^2\right)$$
$$\leq\frac{8T^4}{N^2}\sum_{i,j=1}^{N}\left[\left(\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\right)^2+\varkappa_4\sum_{s=1}^{T}\Psi_{is}^2(1)\Psi_{js}^2(1)\right].$$

Following the steps of the above analysis leading to (31), we obtain

$$Var\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\xi U_i M\tilde{\varphi}_r\right)^2\right)\leq o(1)\frac{T^4}{N^2}(\operatorname{tr}\Omega)^2. \tag{34}$$

Chebyshev's inequality together with (33) and (34) yields

$$\tilde{\lambda}_{12}\leq\sum_{r=K+1}^{T-1}\tilde{\mu}_r+o_{\mathrm{P}}(1)T^2\operatorname{tr}\Omega/N. \tag{35}$$

Using the first inequality of Lemma 4 in (35), we obtain

$$\tilde{\lambda}_{12}\leq(1+o_{\mathrm{P}}(1))\frac{T^2}{9K}\operatorname{tr}\Omega/N. \tag{36}$$

Now, use (36) and (32) in (27), noting the following two facts. First, as implied by the two inequalities of Lemma 4, $\sum_{r=1}^{K}\tilde{\mu}_r/(\operatorname{tr}\Omega/N)$ is of order T^2 for large T. Second, $1/K$ in (36) can be chosen arbitrarily close to zero. Hence, (27) yields

$$\begin{aligned}\tilde{\lambda}_1 &\leq\sum_{r=1}^{T}\alpha_r^2\tilde{\mu}_r+o_{\mathrm{P}}(1)T^2\operatorname{tr}\Omega/N\\ &\leq\alpha_1^2\tilde{\mu}_1+(1-\alpha_1^2)\tilde{\mu}_2+o_{\mathrm{P}}(1)T^2\operatorname{tr}\Omega/N.\end{aligned} \tag{37}$$

On the other hand, $\tilde{\lambda}_1$ must be no smaller than $\tilde{\varphi}_1'\tilde{\Sigma}\tilde{\varphi}_1$. Since

$$\mathbb{E}\tilde{\varphi}_1'\tilde{\Sigma}\tilde{\varphi}_1=\tilde{\varphi}_1'\left(\mathbb{E}\tilde{\Sigma}\right)\tilde{\varphi}_1=\tilde{\mu}_1$$

and, by (31),

$$Var(\tilde{\varphi}_1'\tilde{\Sigma}\tilde{\varphi}_1)=o(1)\frac{T^4}{N^2}(\operatorname{tr}\Omega)^2,$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

we have by Chebyshev's inequality

$$\tilde{\varphi}_1' \tilde{\Sigma} \tilde{\varphi}_1 = \tilde{\mu}_1 + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N. \tag{38}$$

Therefore,

$$\tilde{\lambda}_1 \geq \tilde{\mu}_1 + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N. \tag{39}$$

Combining this with (37), we obtain

$$\tilde{\mu}_1 + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N \leq \alpha_1^2 \tilde{\mu}_1 + (1 - \alpha_1^2) \tilde{\mu}_2 + o_{\mathrm{P}}(T^2) \operatorname{tr} \Omega / N,$$

which implies

$$1 - \alpha_1^2 \leq o_{\mathrm{P}}(1) \frac{T^2}{\tilde{\mu}_1 - \tilde{\mu}_2}.$$

But by (13), $\tilde{\mu}_1 / T^2 \to \mu_1$ and $\tilde{\mu}_2 / T^2 \to \mu_2$. Since by Assumption A4, $\mu_1 > \mu_2$, we have

$$\left(\tilde{F}_1' \tilde{\varphi}_1 \right)^2 = \alpha_1^2 \xrightarrow{\mathrm{P}} 1. \tag{40}$$

This establishes the first convergence in (22) for $k = 1$.

Next, inequalities (37) and (39) yield

$$\left| \tilde{\lambda}_1 - \tilde{\mu}_1 \right| \leq \left| 1 - \alpha_1^2 \right| (\tilde{\mu}_1 + \tilde{\mu}_2) + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N.$$

Combining this with the facts that $\alpha_1^2 = 1 + o_{\mathrm{P}}(1)$ and, by Lemma 4, $\tilde{\mu}_1 \geq CT^2 \operatorname{tr} \Omega / N$ for some $C > 0$, we obtain

$$\tilde{\lambda}_1 = \tilde{\mu}_1 (1 + o_{\mathrm{P}}(1)), \tag{41}$$

which gives us the second convergence in (22) for $k = 1$.

Further,

$$\operatorname{tr} \tilde{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{T} \tilde{\varphi}_j' A^{(i)} \tilde{\varphi}_j.$$

Hence,

$$\mathbb{E} \operatorname{tr} \tilde{\Sigma} = \operatorname{tr} \left(\mathbb{E} \tilde{\Sigma} \right) = \sum_{j=1}^{T} \tilde{\mu}_j$$

and, by (34) which holds for all fixed K, including $K = 0$,

$$Var \left(\operatorname{tr} \tilde{\Sigma} \right) = o(1) \frac{T^4}{N^2} (\operatorname{tr} \Omega)^2.$$

Hence, by Chebyshev's inequality

$$\operatorname{tr} \tilde{\Sigma} = \sum_{j=1}^{T} \tilde{\mu}_j + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

and

$$\frac{\tilde{\lambda}_1}{\operatorname{tr}\tilde{\Sigma}} = \frac{\tilde{\mu}_1\,(1+o_{\mathrm{P}}(1))}{\sum_{j=1}^{T}\tilde{\mu}_j + o_{\mathrm{P}}(1)T^2\operatorname{tr}\Omega/N} = \frac{\tilde{\mu}_1}{\sum_{j=1}^{T}\tilde{\mu}_j} + o_{\mathrm{P}}(1),$$

where the latter equality is a consequence of Lemma 4. Thus,

$$\tilde{\lambda}_1/\operatorname{tr}\tilde{\Sigma} - \tilde{\mu}_1/\operatorname{tr}\mathbb{E}\tilde{\Sigma} \xrightarrow{\mathrm{P}} 0,$$

which establishes the last convergence in (22) for $k=1$. Note that, by Lemma 4, $\tilde{\mu}_1/\operatorname{tr}\mathbb{E}\tilde{\Sigma}$ remains bounded away from zero as $N,T\to\infty$.

For $k=m>1$, the statements of (22) follow by mathematical induction. Indeed, suppose they hold for $k<m$. Consider a representation $\tilde{F}_m = \sum_{q=1}^{T-1}\alpha_q\tilde{\varphi}_q$. Since $\tilde{F}_m'\tilde{F}_j = 0$ for all $j<m$, and since $\left|\tilde{F}_j'\tilde{\varphi}_j\right| = 1+o_{\mathrm{P}}(1)$ by the induction hypothesis, we must have $\alpha_j = o_{\mathrm{P}}(1)$ for all $j<m$. In particular,

$$\tilde{F}_m'\tilde{\Sigma}\tilde{F}_m = \sum_{q,r=m}^{T-1}\alpha_q\alpha_r\frac{1}{N}\sum\nolimits_{i=1}^{N}\tilde{\varphi}_r'A^{(i)}\tilde{\varphi}_q + o_{\mathrm{P}}(1)\,T^2\operatorname{tr}\Omega/N. \tag{42}$$

Indeed, to see that (42) holds, it is sufficient to establish equalities

$$\alpha_j\tilde{\varphi}_j'\tilde{\Sigma}\sum_{r=m}^{T-1}\alpha_r\tilde{\varphi}_r = o_{\mathrm{P}}(1)\,T^2\operatorname{tr}\Omega/N$$

for any $j<m$, and equalities

$$\alpha_j\alpha_r\tilde{\varphi}_j'\tilde{\Sigma}\tilde{\varphi}_r = o_{\mathrm{P}}(1)\,T^2\operatorname{tr}\Omega/N$$

for any $j,r<m$. Such equalities easily follow from the facts that $\alpha_j = o_{\mathrm{P}}(1)$ for all $j<m$ and $\left\|\tilde{\Sigma}\right\| = \tilde{\lambda}_1 = O_{\mathrm{P}}(1)\,T^2\operatorname{tr}\Omega/N$. In addition to (42), we must have

$$\sum_{i=1}^{m-1}\tilde{\lambda}_i + \tilde{F}_m'\tilde{\Sigma}\tilde{F}_m \geq \sum_{j=1}^{m}\frac{1}{N}\sum\nolimits_{i=1}^{N}\tilde{\varphi}_j'A^{(i)}\tilde{\varphi}_j = \sum_{i=1}^{m}\tilde{\mu}_i + o_{\mathrm{P}}(1)\,T^2\operatorname{tr}\Omega/N,$$

where the latter equality is obtained similarly to (38). Combining the above two displays, and using the induction hypothesis, this time regarding the validity of the identities

$$\tilde{\lambda}_i/\tilde{\mu}_i - 1 = o_{\mathrm{P}}(1)$$

for all $i<m$, we obtain

$$\sum_{q,r=m}^{T-1}\alpha_q\alpha_r\frac{1}{N}\sum\nolimits_{i=1}^{N}\tilde{\varphi}_r'A^{(i)}\tilde{\varphi}_q \geq \tilde{\mu}_m + o_{\mathrm{P}}(1)\,T^2\operatorname{tr}\Omega/N. \tag{43}$$

Statements of (22) for $k=m$ now follow by arguments that are very similar to those used above for the case $k=1$.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

That is, we represent the sum on the left-hand side of (43) in the form $\tilde{\lambda}_{m1} + \tilde{\lambda}_{m2} + \tilde{\lambda}_{m3}$, defined similarly to (24–26). Then proceed along the lines of the above proof to obtain an upper bound on $\tilde{\lambda}_{m1} + \tilde{\lambda}_{m2} + \tilde{\lambda}_{m3}$, similar to the right-hand side of (37). Then, combining this upper bound with the lower bound (43), we prove the convergence $\alpha_m^2 \xrightarrow{\mathrm{P}} 1$. Finally, we proceed to establishing the other statements of (22) using this convergence.

4.4. Step 4: Proof of Theorem 1 for $\hat{\lambda}_k, \hat{F}_k, \hat{\Sigma}$

We need to show that the theorem's validity for $\tilde{F}_k$, $\tilde{\lambda}_k$, and $\tilde{\Sigma}$ implies its validity for $\hat{F}_k$, $\hat{\lambda}_k$, and $\hat{\Sigma}$. By standard perturbation theory (e.g., Kato, 1980, Chap. 2), such an implication for statements (i) and (ii) would follow if we are able to show that $\left\| \hat{\Sigma} - \tilde{\Sigma} \right\| = \frac{T^2}{N} \operatorname{tr} \Omega \cdot o_{\mathrm{P}}(1)$. Equation (10) implies that it is sufficient to establish two facts. First, $\|X_{\mathrm{ini}} M\|^2 = T^2 \operatorname{tr} \Omega \cdot o_{\mathrm{P}}(1)$, and second, $\|\Psi^{**}(L) \varepsilon M\|^2 = T^2 \operatorname{tr} \Omega \cdot o_{\mathrm{P}}(1)$.

We have $\|X_{\mathrm{ini}} M\|^2 \le \|X_{\mathrm{ini}} M\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm. A direct calculation yields

$$\|X_{\mathrm{ini}} M\|_F^2 = \sum\nolimits_{i=N_1+1}^{N} \rho_i^{2T^3+2} \left(\frac{1-\rho_i^{2T}}{1-\rho_i^2} - \frac{1}{T} \left(\frac{1-\rho_i^T}{1-\rho_i} \right)^2 \right) \left(Z_{i,-T^3} - \mu_{Xi} \right)^2,$$

and

$$\begin{aligned} \mathbb{E}\,\|X_{\mathrm{ini}} M\|_F^2 &= \sum\nolimits_{i=N_1+1}^{N} \frac{\rho_i^{2T^3+2}}{1-\rho_i^2} \left(\frac{1-\rho_i^{2T}}{1-\rho_i^2} - \frac{1}{T} \left(\frac{1-\rho_i^T}{1-\rho_i} \right)^2 \right) \Psi_{i\cdot}(1)\, \Psi_{i\cdot}'(1) \\ &= \sum\nolimits_{i=N_1+1}^{N} \frac{\rho_i^{2T^3+2}}{1-\rho_i^2} \frac{1-\rho_i^T}{1-\rho_i} \left(\frac{1+\rho_i^T}{1+\rho_i} - \frac{1}{T} \frac{1-\rho_i^T}{1-\rho_i} \right) \Psi_{i\cdot}(1)\, \Psi_{i\cdot}'(1) \\ &\le T \sum\nolimits_{i=N_1+1}^{N} \frac{\rho_i^{2T^3+2}}{1-\rho_i^2} \left(\frac{1+\rho_i^T}{1+\rho_i} - \frac{1}{T} \frac{1-\rho_i^T}{1-\rho_i} \right) \Psi_{i\cdot}(1)\, \Psi_{i\cdot}'(1). \end{aligned}$$

Let us define

$$h(\rho) = \begin{cases} \frac{1}{1-\rho^2} \left(\frac{1+\rho^T}{1+\rho} - \frac{1}{T} \frac{1-\rho^T}{1-\rho} \right), & \rho \in [0,1), \\ 0, & \rho = 0. \end{cases}$$

Then

$$\mathbb{E}\,\|X_{\mathrm{ini}} M\|_F^2 = T^2 \sum\nolimits_{i=N_1+1}^{N} \rho_i^{2T^3+2} h(\rho_i)\, \Psi_{i\cdot}(1)\, \Psi_{i\cdot}'(1).$$

As shown in Section A8 of the Supplementary Material, $h(\rho_i)$ is nonnegative, continuous, $|h(\rho_i)| \le 1$ for all T, and $h(1 - x/T) \le x/4$ for $x \in [0,1)$. This implies that

$$\max_{\rho_i \in [0,1-1/T]} \rho_i^{2T^3+2} h(\rho_i) \le (1-1/T)^{2T^3+2} \le e^{-2T^2},$$

and

$$\max_{\rho_i\in[1-1/T,1]}\rho_i^{2T^3+2}h(\rho_i)\le\max_{\rho_i\in[1-1/T,1]}\rho_i^{2T^3+2}\frac{T(1-\rho_i)}{4}$$
$$=\left(1-\frac{1}{2T^3+3}\right)^{2T^3+2}\frac{T}{4\left(2T^3+3\right)}\le\frac{1}{8T^2}.$$

Since $e^{-2T^2}\le 1/\left(2T^2\right)$, we have overall, $\max_{\rho_i\in[0,1]}\rho_i^{2T^3+2}h(\rho_i)\le 1/(2T^2)$ and

$$\mathbb{E}\left\|X_{\text{ini}}M\right\|_F^2\le\frac{1}{2}\sum\nolimits_{i=N_1+1}^{N}\Psi_{i\cdot}(1)\Psi_{i\cdot}'(1)\le\frac{1}{2}\operatorname{tr}\Omega. \tag{44}$$

By Markov's inequality, $\|X_{\text{ini}}M\|_F^2=\operatorname{tr}\Omega\cdot O_{\mathrm{P}}(1)$, so that

$$\|X_{\text{ini}}M\|^2=\operatorname{tr}\Omega\cdot O_{\mathrm{P}}(1)=T^2\operatorname{tr}\Omega\cdot o_{\mathrm{P}}(1), \tag{45}$$

as required.

It remains to show that $\|\Psi^{**}(L)\varepsilon M\|^2=T^2\operatorname{tr}\Omega\cdot o_{\mathrm{P}}(1)$. To establish this equality, we need the following lemma. The lemma may have an independent interest because it describes a probabilistic bound on the norm of a large random matrix, and such bounds are often useful, for example, in factor analysis.

LEMMA 6. *Suppose that Assumption A1 holds, and that $N,T\to\infty$ at arbitrary relative rates. Let $Z_t=\Pi(L)\varepsilon_t$ and $Z=[Z_1,\ldots,Z_T]$, where $\Pi(L)=\sum_{k=0}^{\infty}\Pi_kL^k$ is an $N\times N$ matrix lag polynomial that may depend on N and T. If $\sum_{k=0}^{T}\|\Pi_k\|=O(1)$ and $T\sum_{k=T+1}^{\infty}\|\Pi_k\|_F^2=O(N)$, where $\|\cdot\|_F$ denotes the Frobenius norm, then*

$$\|Z\|=O_{\mathrm{P}}\left(T^{1/2}+N^{1/2}\right). \tag{46}$$

Proof. This is a modification of Proposition 6 from Onatski (2015), where a proportional asymptotic regime with N/T converging to a nonzero constant is considered. The triangle inequality yields

$$\|Z\|\le\sum\nolimits_{k=0}^{T}\|\Pi_k\|\,\|\varepsilon_{-k}\|+\|r_T\|,$$

where $\varepsilon_{-k}=[\varepsilon_{1-k},\ldots,\varepsilon_{T-k}]$ and $r_T=\sum_{k=T+1}^{\infty}\Pi_k\varepsilon_{-k}$. Obviously, for any $k=0,\ldots,T$, $\|\varepsilon_{-k}\|\le\|\varepsilon_+\|$, where $\varepsilon_+=[\varepsilon_{1-T},\ldots,\varepsilon_T]$. Latała's (2004, Thm. 2) inequality implies that $\|\varepsilon_+\|=O_{\mathrm{P}}\left(T^{1/2}+N^{1/2}\right)$. Therefore,

$$\|Z\|\le O_{\mathrm{P}}\left(T^{1/2}+N^{1/2}\right)\sum\nolimits_{k=0}^{T}\|\Pi_k\|+\|r_T\|=O_{\mathrm{P}}\left(T^{1/2}+N^{1/2}\right)+\|r_T\|. \tag{47}$$

On the other hand,

$$\mathbb{E}\|r_T\|^2\le\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\mathbb{E}\left[(r_T)_{it}^2\right]=\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\mathbb{E}\left[\sum\nolimits_{k=T+1}^{\infty}\sum\nolimits_{s=1}^{N}(\Pi_k)_{is}\varepsilon_{s,t-k}\right]^2$$
$$\le T\sum\nolimits_{k=T+1}^{\infty}\|\Pi_k\|_F^2=O(N).$$

Hence, $\|r_T\|=O_{\mathrm{P}}\left(N^{1/2}\right)$. Combining this with (47) yields (46). □

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

Remark 7. The lemma holds under following simple but stronger assumptions: $\sum_{k=0}^{\infty}\|\Pi_k\| = O(1)$ and $\sum_{k=0}^{\infty} k\|\Pi_k\|^2 = O(1)$. This follows from the inequalities $\|\Pi_k\|_F^2 \leq N\|\Pi_k\|^2$ and $T\sum_{k=T+1}^{\infty}\|\Pi_k\|^2 \leq \sum_{k=0}^{\infty} k\|\Pi_k\|^2$.

Note that

$$\left\|\Psi^{**}(L)\varepsilon M\right\|^2 \leq \left\|\Psi^{**}(L)\varepsilon\right\|^2 \leq 2\left\|\Theta^{**}(L)\varepsilon\right\|^2 + 2\left\|\Pi^{**}(L)\varepsilon\right\|^2,$$

where $\Theta^{**}(L) = \sum_{k=1}^{\infty}\Theta_k^{**}L^k$ and $\Pi^{**}(L) = \sum_{k=0}^{\infty}\Pi_k^{**}L^k$ with

$$\Theta_k^{**} = \sum_{j=1}^{k}\left(\rho^{k-j} - \rho^k\right)\Psi_j \text{ and } \Pi_k^{**} = -\rho^k\sum_{j=k+1}^{\infty}\Psi_j.$$

We have

$$\sum_{k=0}^{\infty}\left\|\Pi_k^{**}\right\| = \sum_{k=0}^{\infty}\left\|\rho^k\sum_{j=k+1}^{\infty}\Psi_j\right\| \leq \sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty}\left\|\Psi_j\right\| \leq \sum_{j=1}^{\infty} j\left\|\Psi_j\right\| = O(1).$$

Further,

$$\left\|\Pi_k^{**}\right\| \leq \sum_{j=k+1}^{\infty}\left\|\Psi_j\right\| \leq \frac{1}{k+1}\sum_{j=k+1}^{\infty} j\left\|\Psi_j\right\| = \frac{1}{k+1}O(1).$$

Combining the latter two displays, we obtain

$$\sum_{k=0}^{\infty} k\left\|\Pi_k^{**}\right\|^2 \leq O(1)\sum_{k=0}^{\infty}\left\|\Pi_k^{**}\right\| = O(1).$$

Hence, by Lemma 6 and Remark 7,

$$\left\|\Pi^{**}(L)\varepsilon\right\|^2 = O_{\mathrm{P}}(T+N). \tag{48}$$

This equality and the fact that, under Assumption A4, $N/\operatorname{tr}\Omega = O(1)$, yield

$$\left\|\Pi^{**}(L)\varepsilon\right\|^2 = O_{\mathrm{P}}(T+N)\frac{\operatorname{tr}\Omega}{N}\frac{N}{\operatorname{tr}\Omega} = T\operatorname{tr}\Omega\cdot o_{\mathrm{P}}(1)$$

for $N,T\to\infty$ at arbitrary relative rates.

Next, recall that $\Theta_k^{**} = \sum_{j=1}^{k}\left(\rho^{k-j} - \rho^k\right)\Psi_j$. For any $k \geq 1$,

$$\begin{aligned}\left\|\frac{\rho^{k-j} - \rho^k}{j}\right\| &= \left\|(I-\rho)\frac{\rho^{k-j} + \cdots + \rho^{k-1}}{j}\right\| \\ &\leq \left\|(I-\rho)\frac{I + \cdots + \rho^{k-1}}{k}\right\| = \left\|\frac{I-\rho^k}{k}\right\|.\end{aligned}$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

Therefore,

$$k\left\|\Theta_k^{**}\right\|_F \leq k\sum_{j=1}^{k}\left\|\frac{\rho^{k-j}-\rho^k}{j}\right\| j\left\|\Psi_j\right\|_F \leq \left\|I-\rho^k\right\|\sum_{j=1}^{k} j\left\|\Psi_j\right\|_F$$
$$\leq N^{1/2}\left\|I-\rho^k\right\|\sum_{j=1}^{k} j\left\|\Psi_j\right\| = O\left(N^{1/2}\right)$$

uniformly in k, where the last equality follows by Assumption A2. Therefore,

$$T\sum\nolimits_{k=T+1}^{\infty}\left\|\Theta_k^{**}\right\|_F^2 \leq O(N)T\sum\nolimits_{k=T+1}^{\infty}\frac{1}{k^2} = O(N).$$

Further, $\Theta_0^{**}=0$ and

$$\sum\nolimits_{k=1}^{T}\left\|\Theta_k^{**}\right\| \leq \sum\nolimits_{k=1}^{T}\sum_{j=1}^{k}\left\|\rho^{k-j}-\rho^k\right\|\left\|\Psi_j\right\| \tag{49}$$
$$=\sum_{j=1}^{T}\left\|\Psi_j\right\|\sum\nolimits_{k=j}^{T}\left\|\rho^{k-j}-\rho^k\right\|.$$

By Assumption A4, there exists $\bar{\phi}>0$ such that $\phi_j \leq \bar{\phi}$ for all $j\in\mathbb{N}$. Note that the maximum of $r^{k-j}-r^k$ on $r\in[0,1]$ is achieved at $r=(1-j/k)^{1/j}$. On the other hand, the smallest possible diagonal element of ρ equals $e^{-\bar{\phi}/T}$, and

$$e^{-\bar{\phi}/T} \geq (1-j/k)^{1/j}$$

for $k \leq T/\bar{\phi}$. Therefore, for such k,

$$\left\|\rho^{k-j}-\rho^k\right\| \leq e^{-k\bar{\phi}/T}\left(e^{j\bar{\phi}/T}-1\right)$$

and

$$\sum\nolimits_{k=j}^{[T/\bar{\phi}]}\left\|\rho^{k-j}-\rho^k\right\| \leq \frac{\left(e^{j\bar{\phi}/T}-1\right)\left(e^{-j\bar{\phi}/T}-e^{-1}\right)}{1-e^{-\bar{\phi}/T}}$$
$$\leq \frac{\left(e^{j\bar{\phi}/T}-1\right)\left(1-e^{-1}\right)}{1-e^{-\bar{\phi}/T}}.$$

But for $x\in[0,1]$, $e^x-1 \leq (e-1)x$ and $1-e^{-x} > \left(1-e^{-1}\right)x$. Therefore, for all $j=1,\ldots,\left[T/\bar{\phi}\right]$,

$$e^{j\bar{\phi}/T}-1 \leq (e-1)\left(j\bar{\phi}/T\right)$$

and for all sufficiently large T,

$$1-e^{-\bar{\phi}/T} \geq \left(1-e^{-1}\right)\left(\bar{\phi}/T\right).$$

Hence,

$$\sum\nolimits_{k=j}^{[T/\bar{\phi}]} \left\| \rho^{k-j} - \rho^{k} \right\| \le \frac{(e-1)\left(j\bar{\phi}/T\right)\left(1-e^{-1}\right)}{\left(1-e^{-1}\right)\left(\bar{\phi}/T\right)} \le 2j.$$

Next, for $k > T/\bar{\phi}$, we have

$$\left\| \rho^{k-j} - \rho^{k} \right\| \le \left(\frac{k-j}{k}\right)^{k/j-1} \frac{j}{k} \le \frac{j}{k} \tag{50}$$

and

$$\sum\nolimits_{k=[T/\bar{\phi}]+1}^{T} \left\| \rho^{k-j} - \rho^{k} \right\| \le j \sum\nolimits_{k=[T/\bar{\phi}]+1}^{T} \frac{1}{k} \le j\left(\ln T - \ln\left(T/2\bar{\phi}\right)\right)$$
$$= j \ln\left(2\bar{\phi}\right).$$

Hence, overall,

$$\sum\nolimits_{k=j}^{T} \left\| \rho^{k-j} - \rho^{k} \right\| \le j\left(\ln\left(2\bar{\phi}\right) + 2\right)$$

and thus,

$$\sum\nolimits_{k=1}^{T} \left\| \Theta_k^{**} \right\| \le \sum_{j=1}^{T} \left\| \Psi_j \right\| j \left(\ln\left(2\bar{\phi}\right) + 2\right) = O(1).$$

In particular, the assumptions of Lemma 6 are satisfied and

$$\left\| \Theta^{**}(L)\varepsilon \right\|^2 = O_{\mathrm{P}}(T+N) = O_{\mathrm{P}}(T+N)\frac{\operatorname{tr}\Omega}{N}\frac{N}{\operatorname{tr}\Omega} = T\operatorname{tr}\Omega \cdot o_{\mathrm{P}}(1), \tag{51}$$

which concludes our proof of parts (i) and (ii) of the theorem.

Part (iii) of the theorem can be established similarly to part (iii) of Theorem 1 in OW, using the fact that, by Lemma 4, there exist positive constants C_1 and C_2 such that

$$C_1 \frac{T^2}{N} \operatorname{tr}\Omega \le \operatorname{tr}\tilde{\Sigma} \le C_2 \frac{T^2}{N} \operatorname{tr}\Omega. \tag{52}$$

Specifically, we need to show that $\left|\operatorname{tr}\hat{\Sigma} - \operatorname{tr}\tilde{\Sigma}\right|$ is asymptotically dominated by $\operatorname{tr}\tilde{\Sigma}$. The above inequalities and the fact that $N/\operatorname{tr}\Omega = O(1)$ imply that it is sufficient to establish the asymptotic dominance of $\left|\operatorname{tr}\hat{\Sigma} - \operatorname{tr}\tilde{\Sigma}\right|$ by T^2.

From (10),

$$\left|\hat{\lambda}_i^{1/2} - \tilde{\lambda}_i^{1/2}\right| \le \left\| \Psi^{**}(L)\varepsilon M \right\| / \sqrt{N} + \left\| X_{\mathrm{ini}} M \right\| / \sqrt{N}$$

and $\hat{\lambda}_i = \tilde{\lambda}_i = 0$ for $i > \min\{N, T\}$. Therefore, by Minkowski's inequality,

$$\left|\left(\operatorname{tr}\hat{\Sigma}\right)^{1/2} - \left(\operatorname{tr}\tilde{\Sigma}\right)^{1/2}\right| \le \left(\left\| \Psi^{**}(L)\varepsilon M \right\| + \left\| X_{\mathrm{ini}} M \right\|\right) \min\left\{1, \sqrt{T/N}\right\},$$

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

and

$$\left|\operatorname{tr}\hat{\Sigma}-\operatorname{tr}\tilde{\Sigma}\right| \leq 2\left(\left\|\Psi^{**}(L)\,\varepsilon M\right\|+\|X_{\text{ini}}M\|\right)\min\left\{1,\sqrt{T/N}\right\}\left(\operatorname{tr}\tilde{\Sigma}\right)^{1/2} + 2\left\|\Psi^{**}(L)\,\varepsilon M\right\|^{2}\min\{1,T/N\} + 2\|X_{\text{ini}}M\|^{2}\min\{1,T/N\}.$$

By (52), $\left(\operatorname{tr}\tilde{\Sigma}\right)^{1/2} = O_{\mathrm{P}}(T)$. Therefore, to establish the asymptotic dominance of $\left|\operatorname{tr}\hat{\Sigma}-\operatorname{tr}\tilde{\Sigma}\right|$ by T^2 it is sufficient to show that

$$\left\|\Psi^{**}(L)\,\varepsilon M\right\|^{2}\min\{1,T/N\} = o_{\mathrm{P}}\left(T^{2}\right) \text{ and} \tag{53}$$

$$\|X_{\text{ini}}M\|^{2}\min\{1,T/N\} = o_{\mathrm{P}}\left(T^{2}\right). \tag{54}$$

Since $\|\Psi^{**}(L)\,\varepsilon M\| \leq \|\Theta^{**}(L)\,\varepsilon\| + \|\Pi^{**}(L)\,\varepsilon\|$, equality (48) and the first equality in (51) yield

$$\left\|\Psi^{**}(L)\,\varepsilon M\right\| = \left(T^{1/2}+N^{1/2}\right)O_{\mathrm{P}}(1).$$

Hence,

$$\left\|\Psi^{**}(L)\,\varepsilon M\right\|^{2}\min\{1,T/N\} = \min\{T+N, T^{2}/N+T\}O_{P}(1) = o_{\mathrm{P}}(T^{2})$$

as $N,T\to\infty$ at arbitrary relative rates, and therefore (53) holds.

Finally, by (45),

$$\|X_{\text{ini}}M\|^{2} = \operatorname{tr}\Omega\cdot O_{\mathrm{P}}(1) = N\cdot O_{\mathrm{P}}(1).$$

Therefore,

$$\|X_{\text{ini}}M\|^{2}\min\{1,T/N\} = \min\{N,T\}O_{\mathrm{P}}(1) = o_{\mathrm{P}}(T^{2})$$

as $N,T\to\infty$ at arbitrary relative rates, and (54) holds too. This concludes our proof of Theorem 1.

5. EXTENSIONS AND DISCUSSION

In general, the covariance kernel of an OU process with parameter ϕ_i has form (see Karatzas and Shreve, 1991, p. 358)

$$R_{\phi_i}(s,t) + (V_i(0) - (2\phi_i)^{-1})e^{-\phi_i(t+s)}, \tag{55}$$

where $R_{\phi_i}(s,t) = e^{-\phi_i|t-s|}/(2\phi_i)$ and $V_i(0)$ is the variance of the initial value. So far, we have assumed that the initial values X_{i0} for $i > N_1$ were distributed according to their unconditional distribution, so that $V_i(0) = (2\phi_i)^{-1}$ and the second term in the above display disappears. This simplification yielded the covariance kernel of the demeaned OU process given by $k_{\phi_i}(s,t)$, described just above Lemma 2.

Had we initialized X_{i0} differently, say making them zero, the second term in (55) would have been equal $-e^{-\phi_i(t+s)}/(2\phi_i)$, and the covariance kernel of the

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

demeaned OU processes would have changed as follows:

$$k_{\phi_i}(s,t) \mapsto k_{\phi_i}(s,t) - \frac{(\phi_i e^{\phi_i t} + e^{-\phi_i} - 1)(\phi_i e^{\phi_i s} + e^{-\phi_i} - 1)}{2\phi_i^3}. \quad \textbf{(56)}$$

In such a case, the scaled sample covariance matrix of the data would have converged to the weighted average of such a new $k_{\phi_i}(s,t)$, and the results of Theorem 1 would have needed a corresponding adjustment.

However, the main message of Theorem 1 would have remained the same: even in the absence of any common factors in the data, PCA would find a few relatively large eigenvalues of the sample covariance matrix, creating a misleading impression of the existence of a few factors driving the dynamics of the entire data. Indeed, the weighted average of the adjusted covariance kernels will still be a continuous function on $(s,t) \in [0,1]^2$. Hence, the eigenvalues of the corresponding integral operator will be absolutely summable and thus, quickly decaying.

Note that the effect of the initial values on our PCA analysis has been cushioned by the demeaning. For example, demeaning pure random walks completely removes any effect of the initial values on the analysis based on second moments of the data. Furthermore, demeaning local-to-unit root processes initialized from their stationary distributions "regularizes" the behavior of the corresponding covariance kernel as a function of ϕ. Indeed, it transforms $R_\phi(s,t) = e^{-\phi|t-s|}/(2\phi)$, which has a singularity at zero, to $k_\phi(s,t)$, which is a continuous function of ϕ at $\phi = 0$. Although our main reason for the demeaning was that it is a standard feature of practical PCA, we also used it to discipline the effect of the initial values.

Setting initial values at zero would have a similar regularizing effect on the covariance kernels. Indeed, from (55), the covariance kernel of an OU process initialized at zero (and not demeaned) is

$$R_\phi(s,t) - (2\phi)^{-1} e^{-\phi(t+s)} = \frac{e^{-\phi|t-s|} - e^{-\phi(t+s)}}{2\phi},$$

which is, obviously, a continuous function of ϕ at $\phi = 0$. This would allow one to obtain an analog of Theorem 1 for the eigenvalues and eigenvectors of matrix $X'X/N$ as opposed to $MX'XM/N$. In such an analog, the covariance kernel $k_{\mathcal{F}}(s,t) = \int\int \omega k_\phi(s,t)\mathcal{F}(\mathrm{d}\omega, \mathrm{d}\phi)$ of the integral operator $K_{\mathcal{F}}$ must be changed by replacing $k_\phi(s,t)$ with the right-hand side of the latter display.

Initializing at zero would also allow one to consider nearly explosive processes with negative ϕ_i along with the processes with positive ϕ_i. Clearly, the weighted average of the covariance kernels of not necessarily stationary OU processes with zero initial values (demeaned or not demeaned) would remain a continuous function on $(s,t) \in [0,1]^2$ (as long as $|\phi_i|$ is bounded). Hence, asymptotically, a few of the eigenvalues of the sample covariance matrix of the data will still be comparable with the sum of all the remaining eigenvalues, creating an impression of a few sources of the data variation. Of course, the form of the principal eigenfunctions $\varphi_k(s)$, $k = 1, 2, \ldots$, would depend on the distribution of all parameters ϕ_i, positive, zero and negative.

Next, we would like to point out that the spurious factor phenomenon is not tied up to the unit root and local-to-unit root cases, and to the corresponding Brownian motion or Ornstein–Uhlenbeck limits. For example, we may still expect finding spurious factors in the data with X_{it}, $t = 1,2,\ldots,T$, coming from a high frequency sampling of continuous time processes with covariance kernels $k_i(s,t)$ as long as weighted averages of such kernels converge to a continuous limit $k_{\mathcal{F}}(s,t)$.

To give a concrete example, suppose that $X_{it} = Z_i(t/T)$, $t = 1,2,\ldots,T$, where $Z_i(s), s \in [0,1]$ are independent (across $i = 1,\ldots,N$) continuous time stationary Markov chains with only two states, 0 and 1. Suppose that, for some $\phi_i > 0$, any $s_2 \geq s_1$ such that $s_2, s_1 \in [0,1]$, and any $k,j \in \{0,1\}$,

$$\Pr(Z_i(s_2) = k | Z_i(s_1) = j) = \begin{cases} \frac{1+e^{-\phi_i(s_2-s_1)}}{2}, & \text{if } k = j, \\ \frac{1-e^{-\phi_i(s_2-s_1)}}{2}, & \text{if } k \neq j, \end{cases} \quad \textbf{(57)}$$

and that the chain is initialized from its stationary distribution (0 or 1 with equal probabilities). It is straightforward to verify that the covariance kernel of $Z_i(s) - \int_0^1 Z_i(s)\mathrm{d}s$ equals $\phi_i k_{\phi_i}(s,t)/2$. That is, it is proportional with the coefficient of proportionality $\phi_i/2$ to the covariance kernel of the demeaned OU process with parameter ϕ_i, initialized at its stationary distribution.

Of course, sample paths of $Z_i(s)$ would be very different from those of the OU process. They will be step functions, jumping from zero to one and back to zero with frequency depending on ϕ_i. However, the PCA analysis of X_{it} would be similar to that of the data with local-to-unit roots. The only difference will be that kernels with relatively small ϕ_i will be downweighted (because of the coefficient of proportionality $\phi_i/2$) in the weighted average kernel. Intuitively, it is because sample paths of $Z_i(s)$ corresponding to small ϕ_i will be just constants with relatively high probability, so they will not substantially contribute to the PCA analysis of the data variation.

In the special case of no heterogeneity in ϕ_i, say all $\phi_i = 1$, the asymptotics of the eigenvectors of the sample covariance of the data with local-to-unit roots and the Markov chain data will be identical. Figure 5 illustrates this. The left and right panels correspond to the local-to-unit root and the Markov chain data, respectively. They show the first three normalized eigenvectors of the corresponding sample covariance matrix (rougher lines) superimposed with the plots of the three principle eigenfunctions of the integral operator with kernel $k_\phi(s,t), \phi = 1$ (smoother lines). The explicit form of these eigenfunctions is given in (3) for general values of ϕ. The eigenvectors are computed using single realizations of data with $N = T = 1{,}000$. As expected, we see very similar behavior of the eigenvectors in the local-to-unity and the Markov chain cases. This is yet another manifestation of the old observation (see, e.g., Perron, 1989) that it is easy to misinterpret data with structural breaks as having unit roots. A detailed analysis of the Markov chain example can be found in Section A10 of the Supplementary Material.

It may be worthwhile to note that, in general, the time series plots of the spurious factors in the data obtained by high frequency sampling of continuous processes

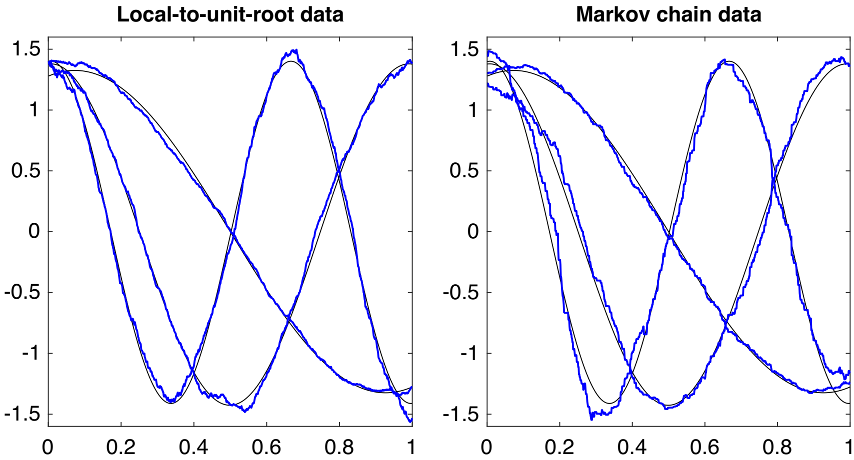

FIGURE 5. Three principal eigenvectors (rough blue lines) of the sample covariance matrix and eigenfunctions (smooth black lines) of the limiting integral operator. Left panel: $N \times T$ data with homogeneous local-to-unit roots $e^{-1/T}$. Right panel: $N \times T$ Markov chain data. In both cases, $T = N = 1,000$, and the lines corresponding to the eigenvectors are piece-wise linear functions connecting the consecutive 1,000 components of the vectors.

do not need to resemble cosine waves. In fact, one may pre-specify virtually any probability limits for such spurious factors by constructing a covariance kernel with corresponding principal eigenfunctions and sampling from continuous time process having that covariance kernel. Hence, although observing the cosine-type time series plots of the principal eigenvectors in PCA analysis should be taken as a warning sign of possibly spurious factors, not observing such cosine-type fluctuations does not necessarily exclude the possibility of spurious factors.

Another observation that we would like to make here concerns interchanging the roles of the temporal and cross-sectional dimensions. One could imagine data that come from "high frequency" sampling of a continuous spacial processes, with such a sampling repeated over time. We would expect similar results for PCA analysis of such data. For a recent discussion of the failure of PCA in such a situation in the context of the analysis of the term structure of interest rates, see Crump and Gospodinov (2022).

In the remaining part of this section, we would like to consider situations where data do have some genuine common factors, but the PCA is still spurious. Suppose that the econometrician observes N-dimensional vectors Y_t, $t = 1, \ldots, T$, satisfying

$$Y_t = \Lambda F_t + X_t,$$

where X_t are as in the previous sections, Λ is an $N \times r$ matrix of factor loadings, and F_t, $t = 1, \ldots, T$, are r-dimensional vectors of latent factors at time t. For now, we assume that the number of factors, r, is fixed. We will briefly comment on the case when r is growing later. Whether results of PCA applied to such data will still be spurious or reflect some information about true factors F_t would depend on the nature of these factors and the magnitude of their loadings.

A standard assumption on factor loadings (see, e.g., Bai and Ng, 2023) is that

$$\Lambda'\Lambda/N^{\alpha} \to \Sigma_{\Lambda} \quad \text{as} \quad N \to \infty, \tag{58}$$

where Σ_{Λ} is a fixed positive-definite $r \times r$ matrix. Strong (or pervasive) factors correspond to $\alpha = 1$, whereas weak factors have $\alpha < 1$. Similarly, for factors, typical assumptions yield

$$F'F/T^{\beta} \xrightarrow{\mathrm{P}} \Sigma_F \quad \text{as} \quad T \to \infty, \tag{59}$$

where $F = [F_1, \ldots, F_T]'$ is a $T \times r$ matrix of factors, and Σ_F is a (possibly random) positive-definite $r \times r$ matrix. Stationary factors correspond to $\beta = 1$ and deterministic Σ_F, whereas factors with local-to-unit roots correspond to $\beta = 2$ and random Σ_F.

Let $Y = [Y_1, \ldots, Y_T]$ be the $N \times T$ matrix of data, M be the $T \times T$ projector $I_T - \ell_T \ell_T'/T$, as above, and $\hat{\lambda}_{Y,k}$ denote the kth largest eigenvalue of $MY'YM/N$ (similar to the kth largest eigenvalue, $\hat{\lambda}_k$, of $MX'XM/N$). Then, by Weyl's inequalities for singular values (e.g., Horn and Johnson, 1985, p. 423),

$$\hat{\lambda}_{r+j}^{1/2} \le \hat{\lambda}_{Y,j}^{1/2} \le \|\Lambda F'M\|/\sqrt{N} + \hat{\lambda}_j^{1/2}.$$

These inequalities together with the standard perturbation theory imply that Theorem 1 continues to hold with $\hat{\lambda}_k$ replaced by $\hat{\lambda}_{Y,k}$ and $\hat{F}_k$ replaced by the kth principal eigenvectors of $MY'YM/N$, as long as $\|\Lambda F'M\|/\sqrt{N} = o_{\mathrm{P}}(T)$.

Now, by (58) and (59), $\|\Lambda F'M\|/\sqrt{N} = O_{\mathrm{P}}(N^{(\alpha-1)/2}T^{\beta/2})$. Therefore, the spurious factor phenomenon, described in Theorem 1, still holds for data with genuine factors as long as $N^{\alpha-1} = o(T^{2-\beta})$. The latter equality holds for both stationary and nonstationary factors ($\beta = 1$ or $\beta = 2$), which are weak ($\alpha < 1$), and for stationary strong ($\alpha = 1$) factors. However, the existence of a fixed number r of strong nonstationary factors in the data will break Theorem 1.

Interestingly, the spurious factor phenomenon will reappear if r is growing, even very slowly. Indeed, consider for simplicity the case with $\Lambda'\Lambda/N = I_r$ and F_t represented by an r-dimensional random walk so that $F_{it} - F_{i,t-1}$ are i.i.d. random variables with mean zero, unit variance, and finite fourth moment.

Consider matrix

$$\frac{MF\Lambda'\Lambda F'M}{N} = MFF'M.$$

Theorem 1 applied to r-dimensional data represented by F implies that the largest eigenvalue of $MFF'M/r$ is of stochastic order T^2. Hence, $\|\Lambda F'M\|/\sqrt{N} = \|F'M\|$ is of stochastic order $\sqrt{r}T$. This means that when r grows, at whatever slow rate, $\|\Lambda F'M\|/\sqrt{N}$ dominates $\hat{\lambda}_1^{1/2} = O_{\mathrm{P}}(T)$. Therefore, the behavior of the principal eigenvalues and eigenvectors of $MY'YM/N$ is asymptotically the same as that of the principal eigenvalues and eigenvectors of $MFF'M$, governed by Theorem 1. In particular, the principal eigenvector will be asymptotically collinear with

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press

the discretized version $(\varphi_1(1/T),\ldots,\varphi_1(T/T))$ of the principal eigenfunction of the covariance kernel of demeaned Brownian motion instead of any of the rows of the matrix of true factors F.

Intuitively, a few strong factors in the data break the spurious factor phenomenon because they may influence the principal eigenvectors of the sample covariance $\hat{\Sigma}$. However, when there is a *growing* number of such factors, they themselves start to spuriously "correlate" along the eigenfunctions of the covariance kernel of a persistent process. This "correlation" will again overwhelm any genuine common cross-sectional features in the data making PCA relatively useless.

6. CONCLUSION

Extending OW, this paper considers applying PCA to data with heterogeneous local-to-unit roots. A similar spurious factor phenomenon is observed, that is, a few principal components explain most of the data variation, with the estimated spurious factors corresponding to the eigenfunctions of a weighted average of covariance kernels of demeaned OU processes with different decay rates. As in OW, the warning to empirical researchers that a very high explanatory power of a few principal components of persistent data does not necessarily indicate the presence of factors and the suggestion to always first difference such data before conducting factor analysis still apply (for a discussion of the properties of factor estimates obtained from differenced data, see Bai and Ng, 2004).

SUPPLEMENTARY MATERIAL

Onatski, A. and C. Wang (2024): Supplement to "Spurious factors in data with local-to-unit roots", *Econometric Theory Supplementary Material.* To view, please visit https://doi.org/10.1017/S0266466624000094.

REFERENCES

Anselone, P. M. (1967). Collectively compact operator approximations. Technical Report 76, Computer Science Department, Stanford University.

Bai, J., & Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4), 1127–1177.

Bai, J., & Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176, 18–29.

Bai, J., & Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2), 1893–1916.

Boivin, J., Giannoni, M. P., & Mojon, B. (2009). How has the euro changed the monetary transmission mechanism? In D. Acemoglu, K. Rogoff, and M. Woodford (Eds.), *NBER macroeconomics annual 2008*, vol. 23. University of Chicago Press, pp. 77–126.

Brillinger, D. R. (1981). *Time series: Data analysis and theory*. Holden Day.

Bykhovskaya, A., & Gorin, V. (2022a). Asymptotics of cointegration tests for high-dimensional VAR(k). Preprint, arXiv:2202.07150.

Bykhovskaya, A., & Gorin, V. (2022b). Cointegration in large VARs. *Annals of Statistics*, 50(3), 1593–1617.

Crump, R. K., & Gospodinov, N. (2022). On the factor structure of bond returns. *Econometrica*, 90, 295–314.

Elliott, G. (1998). The robustness of cointegration methods when regressors almost have unit roots. *Econometrica*, 66, 149–158.

Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review*, 40, 767–783.

Gohberg, I., & Goldberg, S. (1981). *Basic operator theory*. Birkhäuser.

Gonzalo, J., & Pitarakis, J. (2021). Spurious relationships in high-dimensional systems with strong or mild persistence. *International Journal of Forecasting*, 37, 1480–1497.

Horn, R. A., & Johnson, C. R. (1985). *Matrix analysis*. Cambridge University Press.

Karatzas, I., & Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. (2nd ed.) Springer.

Kato, T. (1980). *Perturbation theory for linear operators*. Springer.

Latała, R. (2004). Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society*, 133, 1273–1282.

Onatski, A. (2015). Asymptotic analysis of the squared estimation error in misspecified factor models. *Journal of Econometrics*, 186, 388–406.

Onatski, A., & Wang, C. (2021). Spurious factor analysis. *Econometrica*, 89(2), 591–614.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, 57, 1361–1401.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica*, 56, 1021–1043.

Phillips, P. C. B. (1998). New tools for understanding spurious regression. *Econometrica*, 66, 1299–1325.

Stock, J. H. (1994). Unit roots, structural breaks and trends. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of econometrics*, vol. 4, ch. 46. Elsevier, pp. 2739–2841.

Stock, J. H., & Watson, M. W. (2016). Factor models and structural vector autoregressions in macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of macroeconomics*, vol. 2A. Elsevier, pp. 415–526.

Uhlig, H. (2009). Comment on "How has the euro changed the monetary transmission mechanism?". In D. Acemoglu, K. Rogoff, and M. Woodford (Eds.), *NBER macroeconomics annual 2008*, vol. 23. National Bureau of Economic Research, Inc., pp. 141–152.

Watson, M. W. (1994). Vector autoregressions and cointegration. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of econometrics*, vol. 4, ch. 47. Elsevier, pp. 2843–2915.

Zhang, B., Gao, J., & Pan, G. (2020). Estimation and testing for high-dimensional near unit root time series. Working paper 12/20, Monash Econometrics and Business Statistics.

Zhang, B., Pan, G., & Gao, J. (2018). CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *Annals of Statistics*, 46, 2186–2215.

https://doi.org/10.1017/S0266466624000094 Published online by Cambridge University Press