Detecting Spurious Factor Models

Yi He[1] and Bo Zhang[2]

[1]Amsterdam School of Economics, University of Amsterdam
[2]School of Management, University of Science and Technology of China

November 1, 2023

Abstract

Spurious factor behaviors arise in large random matrices with high-rank signal components and heavy-tailed spectral distributions. This paper establishes analytical probabilistic limits and distribution theory of these spurious behaviors for high-dimensional non-stationary integrated systems, and stationary systems with near-unit-root spatial processes across cross sections. We transform scree plots into Hill plots to detect spectral patterns in these spurious factor models and develop multiple t-tests to distinguish between spurious and genuine factor models. Numerical analysis indicates that the existing spurious factor models fit some, but not all, economic datasets. In particular, the term structure of interest rates adheres to genuine factor models rather than the local correlation model.

Preliminary Version. Comments Welcome.

1 Introduction

Factor models provide the theoretical foundation for econometric studies of aggregate information from economic variables. Extensive literature in econometrics and statistics has developed for latent factor estimation and inference (Stock and Watson, 2002a,b, Bai, 2003, 2004, Forni et al., 2000, 2004, Hallin and Liška, 2011, Doz, Giannone, and Reichlin, 2011, Fan, Liao and Mincheva, 2013, Bai and Ng, 2023, among others). See also Bai and Wang (2016) for an excellent survey. For simplicity, we focus on the static factor models that decompose a large N, large T panel data matrix, say, $X \in \mathbb{R}^{N\times T}$, into a low rank plus noise form, namely,

$$X = A + e, \quad \text{rank}(A) = r, \tag{1}$$

where A represents a signal matrix of small rank r, and e is a high-rank noise matrix of measurement errors. A number of formal methods have been proposed to select or test the rank of factors r by, for instance, Bai and Ng (2002), Onatski (2009, 2010), and Ahn and Horenstein (2013). When the signal is sufficiently stronger than the noise and r is known, one may consistently recover the signal A by using principal component analysis (PCA) under mild conditions. We rule out the difficult cases in Johnstone and Lu (2009) and Onatski (2012) where the signal is too weak to be distinguished from the noise, but we do allow the factor models to have weaker loadings as in Bai and Ng (2023).

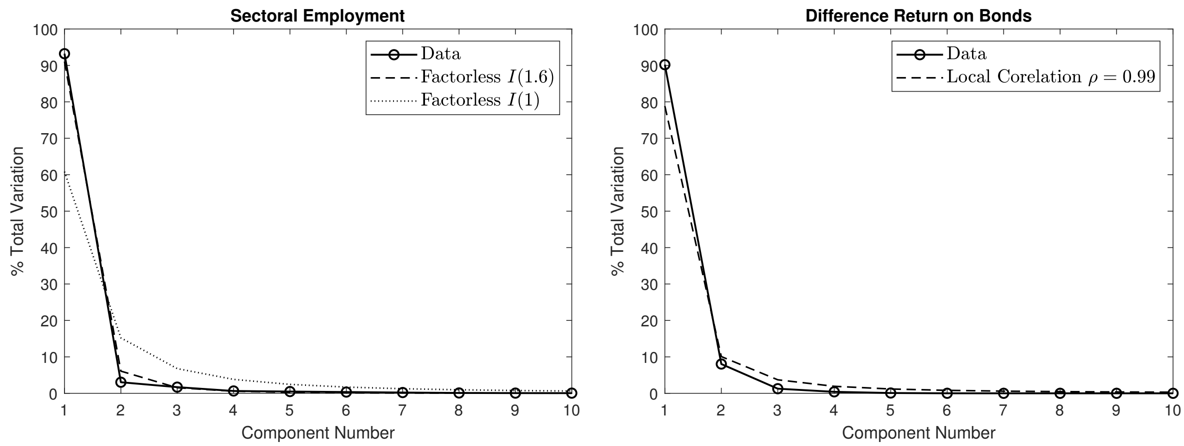


Figure 1: Scree plots for sectoral employment and difference returns on bonds

The scree plots (Cattell, 1966) in Figure 1, showing the large fraction of variations explained by a few principal components, often motivate factor analysis for two running data examples: (1) the US employment data, on log scale, across 58 sectors extracted from the online supplement of Onatski and Wang (2021), hereafter cited as OW; (2) the difference returns across 20 US bonds with maturities up to 20 quarters studied in Crump and Gospodinov (2022), hereafter cited as CG. The sectoral employment data has similar contents of Bai (2004), who suggested choosing $2 \leq r \leq 4$ factors. We downloaded the new vintage of the nominal yield curve data from Fed's website that contains minor modifications to the data in Gürkaynak, Sack and Wright (2007). Tracing back at least to Litterman and Scheinkman (1991), the affine term structure literature often considers $r = 3$ factors for bond returns referred to as level, slope, and curvature; see, e.g., the monograph Piazzesi (2010).

However, recently OW and CG warned the empirical researchers that scree plots could be illusive by theoretically showing that large matrices can generate similar spectral behaviors even in the absence of a low-rank signal plus noise form (1). OW formalized the comments by Uhlig (2009) and derived the analytical limits of "explained" variance ratios for the leading principal components of a factorless large-dimensional unit-root system displayed by the dotted line on the left of Figure 1. In a similar perspective, CG derived the analytical limits for local correlation models with a spatial autoregressive coefficient, say, ρ near unity across the maturity rather than the true time axis. The dashed line on the right of Figure 1 shows these limits for $\rho = 0.99$.

Our first contribution is to generalize OW's findings and derive the analytical limits for the explained variance ratios and spurious factors from a large system of non-stationary integrated $I(d)$ time series of any order $d \geq 1$, where d is not necessarily an integer. This extension is essential because the $I(1)$ model cannot explain all the spurious behaviors: the first PC in $I(1)$ can only explain about 60% of the total variation rather than about 90% in our observations. A better fit is $I(1.6)$ as shown by the dashed line on the left of Figure 1, which can also generate similar principal component estimates as of $I(1)$ that are very close to the observations in our empirical analysis.

Our second contribution is to establish the probabilistic convergence of the sample explained variance ratios and the sample principal component loadings towards their analytical limits according to the local correlation models in CG. We relax their Gaussian assumptions on bond returns and allow for weak temporal dependence. These new limit theorems justify our statistical analysis of the fitness of local correlation models to the actual data by comparing the eigenvalues and eigenvectors.

Like the network aggregating mechanisms in Acemoglu et al. (2012), spurious aggregation variance on principal components emerges through the underlying high-rank dependence structure rather than by low rank random signals. The variance concentration resembles a heavy tail in the spectral distribution such that the largest eigenvalues dominate the aggregate. This novel perspective highlights a close relationship between the spurious factor models and power law models. To visualize the power law exponent of the spectral distribution, we transform the scree plot $(k, \widehat{\lambda}_k)$ into the so-called Hill (1975) plot $(k, \widehat{\gamma}_k)$ via a simple mapping

$$\widehat{\gamma}_k = \frac{1}{k} \sum_{i=1}^{k} \log \widehat{\lambda}_i - \log \widehat{\lambda}_{k+1}, \tag{2}$$

that is, the average exceedance of the eigenvalues $\{\lambda_k\}$ of the sample covariance matrix on the log scale. Note that $\widehat{\gamma}_k$ remain unchanged if we substitute the eigenvalues $\widehat{\lambda}_k$ with their corresponding explained variance ratios.

Figure 2 shows the Hill plots for the same datasets in Figure 1. We can better distinguish their spectral behaviors: the sectoral employment data shows a relatively stable pattern like the integrated systems (again $I(1.6)$ is a better fit than $I(1)$), while the bond data shows an explosive behavior that does not come from the local correlation models.

This remarkable discovery leads to our third and most important contribution: a multiple t-test against these spurious factor models using the Hill plot. Our tests directly apply to the level data that do not require differencing procedures. The asymptotic theory behind our tests builds on a substantial extension of the matrix distribution theory in the earlier work Zhang, Pan and Gao (2018), hereafter abbreviated as ZPG. In particular, we scan over the class of spurious factor models under the null hypothesis so that rejecting their peculiar behaviors may restore our confidence in factor analysis.

What makes the Hill plot unique for our goal is that transformation (2) on the scree plot turns out to be the only solution (up to transformations yielding the same tests) satisfying the following axiomatic properties in our asymptotic theory.

1. Linearity: For each $1 \leq k \leq K$, the statistic $\widehat{\gamma}_k$ is a linear combination of $\{\log \widehat{\lambda}_i\}_{i=1}^{k+1}$.

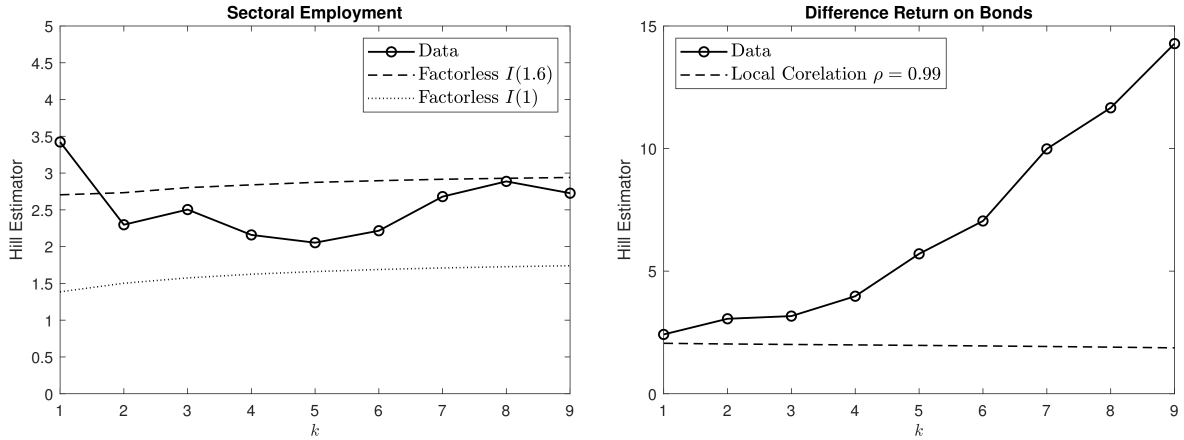


Figure 2: Hill plots for sectoral employment and difference returns on bonds

2. Scale Invariance: $\{\gamma_k\}$ are invariant to the scale of eigenvalues $\{\lambda_k\}$.

3. Independence: $\widehat{\gamma}_k$ is asymptotically independent of $\widehat{\gamma}_{k'}$ for any $k \neq k'$.

The first property allows for testing power by detecting the separation between signal and noise under the genuine factor models (1) over different choices of k. The second property is crucial for removing the nuisance scale parameter for our tests. The last property means that drawing the Hill plots for more eigenvalues brings new information to recognize factor models.

Combining Hill estimators' asymptotically independence and normality, we construct a sequence of t statistics and use their maximum magnitude to detect if the data is generated from a particular class of spurious factor models. To mitigate potential power loss resulting from self-normalization in t statistics, we employ a variance shrinkage approach inspired by Fan, Liao and Yao (2015). This approach ensures that the shrinkage effect remains negligible under the null (spurious factor models), but the test statistic explodes under the alternative (weaker factor models). Our central limit theorems are for eigenvalues rather than eigenvectors. The center of our discussion is the rank of signals defined based on eigenvalues. As pointed out by OW, comparing eigenvectors is a 'simple' but 'inexact' method because the genuine latent factors can still look similar to spurious ones in eigenvector plots, as in the last part of our simulation study. However, comparing eigenvectors may still be helpful if one is interested in distinguishing genuine and spurious factors, but we leave these to future works.

Our random-matrix-theory-based limit theorems for the Hill estimator using a small number of extreme eigenvalues (or their corresponding explained variance ratios) is novel, being very different from that in extreme value statistics requiring an infinite number of observations from the tail (see, e.g., Drees, Resnick, and de Haan, 2000). The only exception is the fixed-k inference by Müller and Wang (2017) among follow-up works, but our extreme eigenvalues do not obey their generalized Pareto models. In contrast, we show that extreme eigenvalues are asymptotically Gaussian under the spurious factor models thanks to our large dimensions across individuals and time.

The rest of the paper is organized as follows. Section 2 develops our asymptotic theory for the spurious factor models through large-dimensional integrated time series. Section 3 develops a parallel approach for the local correlation models with a near-unit-root process across the cross sections. Section 4 presents a simulation study to investigate our tests' size and power performance in small samples. Our empirical analysis in Section 5 does not reject a spurious factor analysis of the sectoral employment data but provides strong evidence against the local correlation models for the bond data. We point out that rejecting the local correlation model can have important implications for mean-variance portfolio optimization because the economic costs of the type-II error, namely, wrongly classifying a genuine factor model as a local correlation model, could be as substantial as the cost of the type-I error. Section 6 offers more remarks and discusses some possible extensions based on more economic examples. All the mathematical proofs are in the appendices.

2 Spurious Factor Models Through Time Structure

In this section, we offer a general mechanism to generate spurious factor models through large-dimensional integrated time series $X_t \in \mathbb{R}^N$, $t = 1, \ldots, T$, with N comparable to T:

Assumption 2.1. $N = N(T) \to \infty$ and $N/T \to c \in (0, \infty)$ as $T \to \infty$.

This framework is often called the random matrices regime, with non-trivial large dimensional effects across individuals and time. We allow for a high dimension $N > T$ yielding singular sample covariance matrices. For presentation convenience, we shall suppress T in the subscript whenever possible: the matrices are all indexed by the sample size T, and the asymptotic results hold as $T \to \infty$. Let $\mathbf{L}$ denote the lag operator and suppose X_t are generated by the model

$$(1 - \mathbf{L})^d (X_t - X_0) = u_t, \tag{3}$$

where the exponent $d \geq 0$ denotes the order of integration, and the innovations $u_t \in \mathbb{R}^N$ are stationary linear time series given by

$$u_t = \Psi(\mathbf{L})\varepsilon_t = \left(\sum_{\ell=0}^{\infty} \Psi_\ell \mathbf{L}^\ell \right) \varepsilon_t = \sum_{\ell=0}^{\infty} \Psi_\ell \varepsilon_{t-\ell} \tag{4}$$

with $u_t = 0$ for $t \leq 0$ for simplicity.

Assumption 2.2. The entries of $\varepsilon_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,N})'$ are independent and identically distributed (i.i.d.) variables from a double array with zero mean $\mathbb{E}\varepsilon_{t,i} = 0$, unit variance $\mathbb{E}\varepsilon_{t,i}^2 = 1$, and bounded kurtosis $\mathbb{E}\varepsilon_{t,i}^4 < \infty$.

Our interest is in the non-stationary system that demonstrate spurious factor model behaviors.

Assumption 2.3. The following conditions hold.

1. The integrated process (3) is non-stationary of some order $d \geq 1$.

2. There exists a summable sequence of constants $\varphi_\ell \geq 0$ such that $\sum_{\ell=0}^{\infty} \varphi_\ell < \infty$ and $\|\Psi_\ell\| \leq \varphi_\ell \|\Psi\|$ for all $\ell \in \mathbb{N}_0$, where $\|\cdot\|$ denotes the Frobenius norm and $\Psi = \Psi(1) = \sum_{\ell=0}^{\infty} \Psi_\ell$.

When $d > 1$ in condition 1, the local Whittle estimator of d, which converges to 1 under mild conditions, is *in*consistent using marginal time series (Phillips and Shimotsu, 2004). Examining many time series simultaneously to identify the order d is crucial in our analysis. We exclude the fractional integration cases $d \in (1/2, 1)$ for technical reasons; otherwise, a workaround is to apply our approach to the cumulative sums $Y_t = \sum_{s=0}^{t}(X_s - X_0)$ with an integrated order of $d + 1 \geq 1$ for any $d \geq 0$. Still, we do not pursue more details as the fractionally integrated models do not generate as obvious concentration behaviors as observed in our economic application on sectoral employment. Condition 2 is similar to that in Liu, Aue and Paul (2015) for large-dimensional linear time series, but we relax the simultaneous diagonalizability of Ψ_ℓ and relax the spectral norms to be Frobenius norms. It ensures the effects of contemporary covariances in (4) vanishes in the long-run covariances through a sequence of local coefficients φ_ℓ invariant with respect to the data scale.

One can invert the autoregressive model (3) into an alternative representation given by

$$X_t = X_0 + (1 - \mathbf{L})^{-d} u_t = X_0 + \sum_{\ell=0}^{t-1} a_\ell u_{t-\ell}, \quad a_\ell = \frac{\Gamma(d+\ell)}{\ell!\,\Gamma(d)}, \tag{5}$$

where Γ denotes the gamma function and $\ell! = s \cdot (s-1) \cdot \ldots \cdot 1$ denotes the factorial of ℓ. Let $X = [X_1, \ldots, X_T] \in \mathbb{R}^{N \times T}$ and $\widetilde{X} = XC \in \mathbb{R}^{N \times T}$ denote the raw and demeaned observations respectively, where C denotes the $T \times T$ centering matrix. Denote by $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_T] \in \mathbb{R}^{N \times T}$ the random matrix of latent noises from (4). We can rewrite X_t further into an Beveridge and Nelson (BN) decomposition form in terms of the random noises ε_t given by

$$\widetilde{X} = \Psi \varepsilon \Phi + \Xi C, \quad \Phi = (I - L_T)^{-d} C \tag{6}$$

where $\Psi = \Psi(1)$ as in Assumption 2.3, I denotes the $T \times T$ identity matrix, and L_T denotes the $T \times T$ upper shift matrix with ones on the superdiagonal and zero elsewhere, and the error matrix $\Xi = [\Xi_1, \ldots, \Xi_T]$ with

$$\Xi_t = \sum_{s=0}^{t-2}(a_s - a_{s+1}) \sum_{j=0}^{s}(u_{t-j} - \Psi \varepsilon_{t-j}) + a_{t-1} \sum_{j=0}^{t-1}(u_{t-j} - \Psi \varepsilon_{t-j}).$$

It is instructive to compare (6) with (1), where the leading term $A = \Psi \varepsilon \Phi$ takes a so-called separable form in random matrix theory (see, e.g., Ding and Yang, 2021), and the remainder $e = \Xi C$ shall be negligible under mild conditions.

What we need for the BN decomposition (6) is the separability on top eigenvalues to identify the spurious factors, where $K \geq 1$ is some user-specified parameter of the maximum number of spurious factors.

Assumption 2.4. Let σ_k^2 denote the k-th largest eigenvalue of the population matrix $\Phi'\Phi$. For all large T, $\sigma_k^2/\sigma_{k+1}^2$ are bounded away from 1 over $1 \leq k \leq K+1$ for some given K.

This condition is comparable to the identification condition for latent factors in, for example, Anderson (1963), Stock and Watson (2002a), Bai (2003), and many follow-up works mentioned in the introduction. If these population eigenvalues are not separable, we cannot identify all individual spurious factors, but we may generalize our theory at the cost of complexity using the notions of subspace convergence in PCA literature; see, for example, Jung and Marron (2009), Shen, Shen and Marron (2016) and supplement of He (2023).

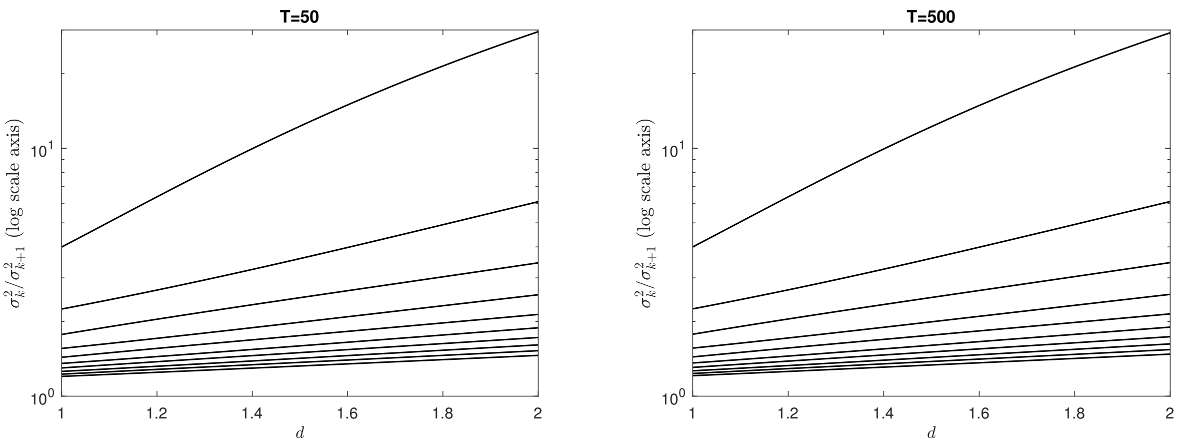


Figure 3: The consecutive eigenvalue ratio $\sigma_k^2/\sigma_{k+1}^2$ as functions of d for $k = 1, 2, \ldots, 10$ (from top to bottom) with sample sizes $T = 50$ (left) and $T = 500$ (right).

Indeed, numerical methods immediately suggest that the gaps between top eigenvalues of $\Phi'\Phi$ are large enough to identify their corresponding principal components. Figure 3 shows the ratio $\sigma_k^2/\sigma_{k+1}^2$ as a function of $d \in [1, 2]$ for $k = 1, \ldots, 10$. The ratios are smaller for larger k, increase with the order of integration d, and are all strictly larger than 1 (the lower limit of the y-axis). A rigorous proof of this separability condition is given in ZPG and OW for the unit root process with $d = 1$. Since the separability even improves as the order of integration d increases according to Figure 3, we consider the assumption trivial and do not observe any violations in numerical analysis. However, a mathematical proof is difficult, and we leave it as an open question.

Following OW, we consider the spurious factor analysis on the level data X_t under the following factorless condition.

Assumption 2.5. The long-run covariance matrix $\Omega = \Psi\Psi'$ is positive definite and factorless in the sense that the population explained variance ratios of the principal components approach zero in such a way that: $\lambda_{\max}(\Omega)/\operatorname{tr}\Omega \to 0$ as $N, T \to \infty$, where $\lambda_{\max}(\Omega)$ denotes the largest eigenvalue of Ω and the sum of eigenvalues of Ω equals to its trace $\operatorname{tr}\Omega$.

We shall show that the spurious phenomenon generalizes towards non-stationary integrated time series beyond unit-root processes. Specifically, let $\widetilde{X} = [X_1 - \bar{X}, \ldots, X_T - \bar{X}] \in \mathbb{R}^{N\times T}$ and $\bar{X} = \frac{1}{T}\sum_{t=1}^T X_t$, then the principal eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_N$ (i.e. "explained variance" of the principal components) of the sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \bar{X}\right)\left(X_t - \bar{X}\right)' = \frac{1}{T}\widetilde{X}\widetilde{X}' \in \mathbb{R}^{N\times N}$$

demonstrates a concentration behavior that is similar to a genuine factor model. The proportion of explained variance

$$\widehat{\xi}_k = \widehat{\lambda}_k / \left(\sum_{i=1}^{N}\widehat{\lambda}_i\right) \tag{7}$$

is large for any $k \geq 1$, via its consistency towards that of the population matrix Φ defined in (6). This limit is not related to the cross-sectional covariance matrix Ψ, and it is spurious in the sense that it determined by the temporal dependence structure Φ rather than common latent variables. Let '$\xrightarrow{\mathbb{P}}$' denote convergence in probability.

Theorem 1. *For all large-dimensional non-stationary integrated time series generated by* (3) *and* (4) *under Assumptions 2.1–2.5, a spurious factor model emerges in such a way that the followings hold for every* $k \geq 1$*:*

(i) $\widehat{\xi}_k - \xi_k \xrightarrow{\mathbb{P}} 0$ *where* $\xi_k := \sigma_k^2/\left(\sum_{i=1}^{T}\sigma_i^2\right)$ *are the explained variance ratios for the eigenvalues* σ_i^2 *of the* $T\times T$ *covariance matrix* $\Phi'\Phi$.

(ii) The population ratio ξ_k *is bounded away from 0.*

(iii) $|\widehat{U}_k' U_k| \xrightarrow{\mathbb{P}} 1$, *where* $\widehat{U}_k$ *is the* k*-th principal eigenvector of* $\widetilde{X}'\widetilde{X} \in \mathbb{R}^{T\times T}$ *and* U_k *denotes the* k*-th principal eigenvector of* $\Phi'\Phi \in \mathbb{R}^{T\times T}$.

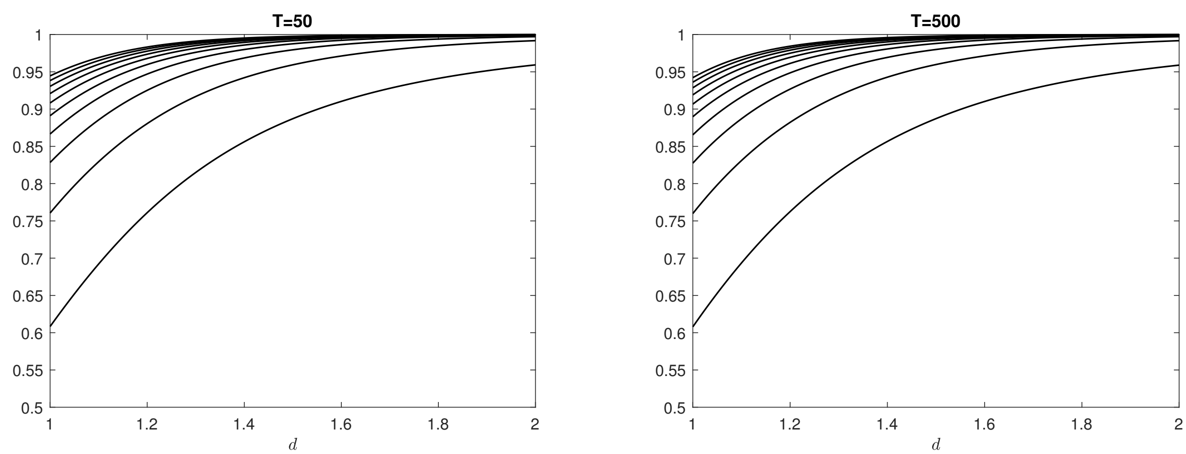


Figure 4: Cumulative explained variance proportions $\sum_{k=1}^{r}\xi_k$ by top r principal components as a function of d for $r = 1, 2, \ldots, 10$ (from bottom to top) with sample sizes $T = 50$ (left) and $T = 500$ (right).

For any given sample size T, the asymptotic explained variance proportions ξ_k of spurious factors in this theorem only depends on d and can be computed numerically. Figure 4 shows the cumulative explained variance proportions $\sum_{k=1}^{r} \xi_k$ for $1 \leq r \leq 10$ and $d \in [1, 2]$ with two sample sizes $T = 50$ and $T = 500$. For every fixed number r of principal components, the cumulative explained variance proportions grows with d. The top 3 components explain about 83% of the total variation for $d = 1$ while 1 component alone suffices to explain 96% of the total variation for $d = 2$. All spurious factors are, however, due to the commonality of time structure.

While the explained variance ratios are sensitive to d, Figure 5 shows that the spurious factors for unit-root $d = 1$ and a higher order $d = 1.6$ look rather similar.

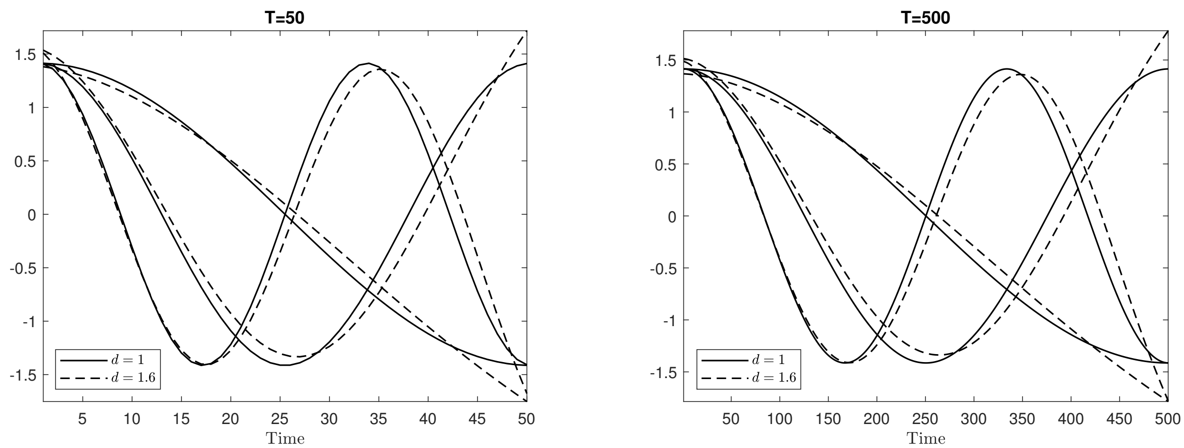


Figure 5: The first three spurious factors $\widehat{F}_k = \sqrt{T}\widehat{U}_k$, $1 \leq k \leq 3$, for $d = 1$ and $d = 1.6$.

Theorem 1 generalizes the first-order asymptotics for spurious factor analysis of an arbitrary order $d \geq 1$. The extreme eigenvalues, properly standardized, can be approximated as quadratic forms in random vectors with a common weight matrix $\underline{\Omega} = \Psi'\Psi$. Since $\underline{\Omega}$ share the same eigenvalues as the long-run covariance matrix Ω in Assumption 2.5, we can immediately impose the spectral condition on Ω for second-order results.

Assumption 2.6. $\lambda_{\max}(\Omega)/\left\|\Omega\right\| \to 0$, where $\left\|\cdot\right\|$ denotes the Frobenius norm.

This condition is stronger than Assumption 2.5 as $\|\Omega\| \leq \operatorname{tr}\Omega$. It is possible to weaken it by controlling the structure of $\underline{\Omega}$ directly rather than that of Ω; see Remark 1. Let '$\Rightarrow$' denote convergence in distribution. The following theorem gives the asymptotic normality of extreme eigenvalues.

Theorem 2. *Under the conditions of Theorem 1 and Assumption 2.6 with any fixed* $K \geq 1$*,*

$$\sqrt{\frac{\nu}{2}}\left(\log\frac{\widehat{\lambda}_1}{\sigma_1^2} - \mu, \ldots, \log\frac{\widehat{\lambda}_{K+1}}{\sigma_{K+1}^2} - \mu\right) \Rightarrow (Z_1, \ldots, Z_{K+1})$$

with independent standard Gaussian variables $Z_1, \ldots, Z_{K+1}$ *provided that* $\sum_{\ell=1}^{\infty} \ell^{1/2}\varphi_\ell < \infty$*, where*

$$\nu = \frac{(\operatorname{tr}\Omega)^2}{\operatorname{tr}\Omega^2}, \quad \mu = \log\left(\frac{\operatorname{tr}\Omega}{T}\right).$$

The theorem remains true by replacing the condition $\sum_{\ell=1}^{\infty} \ell^{1/2}\varphi_\ell < \infty$ *with a weaker version:* $\sum_{\ell=1}^{\infty} \ell^{\iota/2}\varphi_\ell < \infty$ *for some* $\iota > 0$ *such that* $\nu = O(N^\iota)$*. Note that we always have* $\nu \leq N$ *by Cauchy-Schwarz inequality.*

Remark 1. The theorem remains true by replacing Assumption 2.6 with a weaker version:

$$\left\| \widetilde{\underline{\Omega}}' \widetilde{\underline{\Omega}} \right\| = o(\|\underline{\Omega}\|^2),$$

where $\widetilde{\underline{\Omega}}$ denotes the triangular projection of $\underline{\Omega}$ sharing the same lower triangular part but having zeros elsewhere (including the diagonal).

Theorem 2 shows that extreme sample eigenvalues are asymptotically independent due to the orthogonality between eigenvectors. Including more extreme eigenvalues gives more information about the data generating process. There are two nuisance parameters $\mu = \mu(\Omega)$ and $\nu = \nu(\Omega)$ depending on the long-run covariance matrix Ω. Using the equality between trace and sum of eigenvalues, the parameter ν measures the sparseness of the spectrum of Ω via the identity

$$\sqrt{\nu} = \|\lambda(\Omega)\|_1 \,/\, \|\lambda(\Omega)\|_2\,,$$

where $\|\lambda(\Omega)\|_1$ and $\|\lambda(\Omega)\|_2$ denote L_1 and L_2 norms of the vector $\lambda(\Omega)$ collecting all the eigenvalues of Ω respectively.

Whereas the location parameter μ depends only on the long-run marginal variances on the diagonal of Ω, the scale parameter ν depends on the entire cross-sectional structure. When Ω is unknown, these parameters are difficult to estimate. We shall explain how to profile these nuisance parameters, one by one, via self-normalizations using multiple eigenvalues. First, we apply the invariance principal and consider the Hill estimators $\widehat{\gamma}_k$ given by

$$\widehat{\gamma}_k = \frac{1}{k}\sum_{i=1}^{k} \log \widehat{\lambda}_i - \log \widehat{\lambda}_{k+1} = \frac{1}{k}\sum_{i=1}^{k} \log \widehat{\xi}_i - \log \widehat{\xi}_{k+1}. \tag{8}$$

As discussed in the introduction, these Hill estimators are the only invariant statistics with respect to μ that satisfy all the axiomatic properties, up to sign or scale changes. The second representation in terms of $\widehat{\xi}_k$ allows us to compute statistics $\widehat{\gamma}^{(k)}$ directly from the scree plots without knowing the scale of the underlying eigenvalues. Their joint asymptotic distribution follows immediately by the continuous mapping theorem.

Corollary 1. *Under the conditions of Theorem 2 and for any fixed* $K \geq 1$*,*

$$\sqrt{\frac{\nu}{2}}\,(\widehat{\gamma}_1 - \gamma_1, \ldots, \widehat{\gamma}_K - \gamma_K) \Rightarrow (\Gamma_1, \ldots, \Gamma_K)$$

where Γ_k *are independent mean-zero Gaussian variables with* $\mathrm{var}(\Gamma_k) = 1 + k^{-1}$*, and the oracle parameters* γ_k *are given by*

$$\gamma_k = \frac{1}{k}\sum_{i=1}^{k} \log \sigma_i^2 - \log \sigma_{k+1}^2 = \frac{1}{k}\sum_{i=1}^{k} \log \xi_i - \log \xi_{k+1}, \quad k \geq 1, \tag{9}$$

with the explained variance ratios ξ_k *from Theorem 1.*

This theorem characterizes the asymptotic behavior of the Hill plot using K Hill estimators based on $K+1$ leading eigenvalues. As discussed in the introduction, our asymptotic theory of the Hill estimator with a finite k is novel and very different from that in existing extreme value statistics literature.

Next, we profile out the nuisance parameter ν. Suppose d is unspecified, and imagine if we could introduce the *oracle* z-score statistic for each individual Hill estimator in terms of d given by

$$\widehat{z}_k(d) = \sqrt{\frac{\nu}{2}} \cdot \sqrt{\frac{k}{1+k}} \left(\widehat{\gamma}_k - \gamma_k(d)\right),$$

where the benchmark parameter $\gamma_k = \gamma_k(d)$ is defined by (9) via $\sigma_k^2 = \sigma_k^2(d)$, all as functions of d. Under a null hypothesis $d = d_0$ in Theorem 2,

$$(\widehat{z}_1(d_0), \ldots, \widehat{z}_K(d_0)) \Rightarrow (Z_1, \ldots, Z_K), \quad Z_k \overset{iid}{\sim} \mathcal{N}(0,1).$$

Based on these z-scores we can construct t-statistics sequentially given by

$$\widetilde{t}_{k-1}(d) = \frac{\widehat{z}_k(d)}{\sqrt{(\widehat{z}_1^2(d) + \ldots + \widehat{z}_{k-1}^2(d))/(i-1)}} = \sqrt{\frac{k}{k+1}} \frac{\widehat{\gamma}_k - \gamma_k(d)}{\widetilde{s}_{k-1}(d)}, \quad k \geq 2, \tag{10}$$

where the variance statistic

$$\widetilde{s}_j^2(d) = \frac{1}{j} \sum_{i=1}^{j} \frac{i}{i+1} \left(\widehat{\gamma}_i - \gamma_i(d)\right)^2,$$

are computed from the preceding z-scores. Note that ratio statistics $\widetilde{t}_{k-1}(d_0)$ are now all feasible, whose distribution converges to the student t distribution with degrees of freedom $k-1$ under the null $d = d_0$.

However, standard t-tests based on the naive self-normalized statistics (10) may be lack of power. The t-statistics can remain small by canceling out the numerator and denominator even under the alternatives where the oracle z-scores are all very large. To enhance testing power in the spirit of Fan, Liao and Yao (2015), we propose to shrink the scale parameter $\widetilde{s}_{k-1}^2(d)$ by exploiting the fact it must vanish under the null, and replace it with the following shrinkage estimator:

$$\widehat{s}_{k-1}^2(d) = \frac{1}{k-1} \sum_{i=1}^{k-1} \delta_h \left(\frac{i}{i+1} \left(\widehat{\gamma}_i - \gamma_i(d)\right)^2 \right) \tag{11}$$

with some shrinkage δ_h function and bandwidth $h > 0$ in the form of

$$\delta_h(x) = xG\left(\frac{x}{h}\right)/G(0), \quad x > 0 \tag{12}$$

for some continuous positive kernel function $G = G(z)$ such that $\sup_{z \geq 0} zG(z) < \infty$. The bandwidth h controls the magnitude of shrinkage: the smaller h, more sensitive of our test statistic to the bias of the Hill estimators. Finally, define the power-enhanced t-statistics by

$$t_{k-1}(d) = \sqrt{\frac{k}{k+1}} \frac{\widehat{\gamma}_k - \gamma_k(d)}{\widehat{s}_{k-1}(d)}, \tag{13}$$

then we have their joint asymptotic distribution directly via the continuous mapping theorem.

Corollary 2. *Under the conditions of Theorem 1 and the null hypothesis $d = d_0 \geq 1$, for any fixed $K \geq 2$ and bandwidth sequence h such that $h\nu \to \infty$,*

$$(t_1(d_0), \ldots, t_{K-1}(d_0)) \Rightarrow (\mathrm{T}_1, \mathrm{T}_2, \ldots, \mathrm{T}_{K-1}),$$

where T_j are student variables of degrees of freedom j, but they are dependent via a common sequence of independent standard normal variables $Z_1, \ldots, Z_K$ in such a way that

$$\mathrm{T}_j = \frac{Z_{j+1}}{\sqrt{(Z_1^2 + \ldots + Z_j^2)/j}}, \ j = 1, \ldots, K-1.$$

Remark 2. The bandwidth condition $h\nu \to \infty$ is weak and holds for all fixed bandwidth $h \in (0, \infty)$ as $\nu \to \infty$ under Assumption 2.5. In particular, if the eigenvalues of Ω are bounded away from zero and infinity, the nuisance parameter ν diverges at the exact order of N. Then the condition holds for $h = 2\chi_1^2(1-\beta)/N$ with any choice of vanishing $\beta \downarrow 0$ where $\chi_1^2(\cdot)$ denotes the quantile function of the chi-square distribution with unit degree of freedom.

To detect whether the true model is an integrated system with an unspecified d, we propose to scan over a sufficiently large but compact parameter space and use the following minimax statistic

$$H_k = \min_d \max_{1 \leq i \leq k-1} |t_i(d)|.$$

Obtaining the non-degenerate limit distribution of H_k is challenging but a conservative test is available using the fact that $H_k \leq \max_{1 \leq i \leq k-1} |T_i(d_0)|$, where d_0 denotes the true value of d and the upper bound has a simple asymptotic distribution from Corollary 2:

$$\max_{1 \leq i \leq k-1} |t_i(d_0)| \Rightarrow \max_{1 \leq i \leq k-1} |\mathrm{T}_i|. \tag{14}$$

While it is possible to calculate the critical values from the limit in equation (14) using Monte Carlo methods, the Bonferroni method generates almost identical critical values as demonstrated by the left plot in Figure 6 for small size α. Specifically, we control the sum of tail probabilities over all t variables T_i based on Boole's inequality:

$$\mathbb{P}\left(\max_{1 \leq i \leq k-1} |\mathrm{T}_i| > t\right) \leq \sum_{i=1}^{k-1} \mathbb{P}(|\mathrm{T}_i| > t) = 2\sum_{i=1}^{k-1} \mathbb{P}(\mathrm{T}_i > t) = 2 \sum_{1 \leq i \leq k-1} S_i(t),$$

where S_i denotes the student survival function with degrees of freedom i. To summarize, our minimax test rejects the integrated system if

$$H_k > S_{1:k}^{-1}(\alpha/2), \quad S_{1:k} = \sum_{1 \leq i \leq k-1} S_i, \tag{15}$$

where the critical value $S_{1:k}^{-1}(\alpha/2)$ can be easily computed for any $\alpha \in (0, 1)$.

Corollary 3. *Under the conditions of Corollary 2, the minimax test (15) is asymptotically conservative for all $2 \leq k \leq K$, such that for all fixed $\alpha \in (0, 1)$*

$$\limsup_{T \to \infty} \mathbb{P}(H_k > S_{1:k}^{-1}(\alpha/2)) \leq \alpha.$$

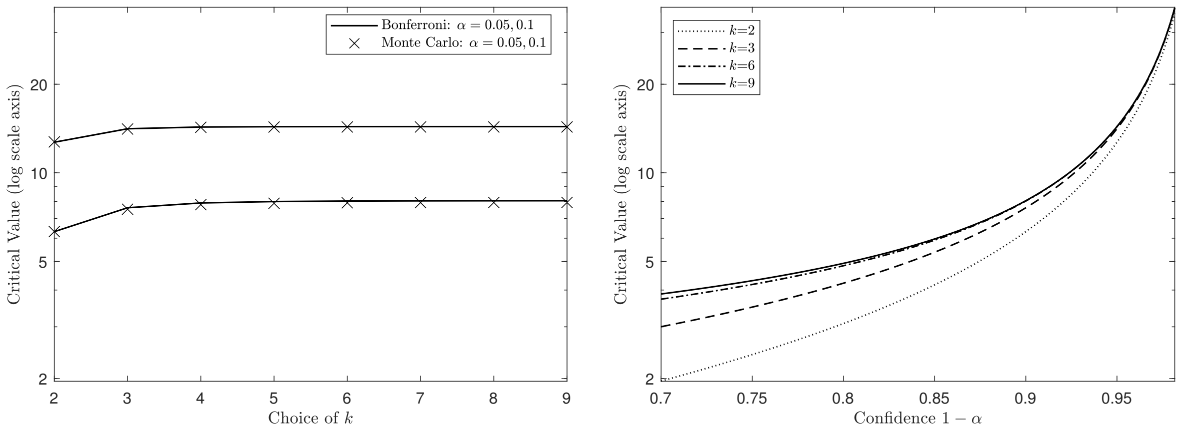


Figure 6: The critical value $S_{1:k}^{-1}(\alpha/2)$ as a function of $1 - \alpha$ for different choices of k.

Furthermore, the type-I error approaches zero with any vanishing level, that is,

$$\mathbb{P}(H_k > S_{1:k}^{-1}(\alpha/2)) \to 0, \quad \text{if } \alpha = \alpha(N, T) \downarrow 0.$$

Remark 3. The corollary remains true for $\alpha \downarrow 0$ if one replaces the critical value $S_{1:k}^{-1}(\alpha/2)$ with $S_1^{-1}(\alpha/2) = \tan[\pi(1/2 - \alpha/2)]$, as the critical values are asymptotically the same at the log-scale as $\alpha \downarrow 0$ for all $k \geq 2$ in Figure 6. However, we use the critical value $S_{1:k}^{-1}(\alpha/2)$ adaptive to k to maintain better finite-sample performance.

The type-II error of our minimax test against genuine factor models tends to zero with an appropriate choice of α. To fix ideas, consider the factor models according to (1) as follows:

$$X_t = \Lambda f_t + e_t, \tag{16}$$

where $f_t \in \mathbb{R}^r$ are latent factors of a finite dimension r, $\Lambda \in \mathbb{R}^{N \times r}$ is a matrix of factor loading coefficients, and $e_t \in \mathbb{R}^N$ are measurement errors. Note that the factors f_t and measurement errors e_t could be either stationary or non-stationary. Recall that the demeaned data matrix $\widetilde{X} = [X_1 - \bar{X}, \ldots, X_T - \bar{X}]$, and define $\widetilde{F} = [f_1 - \bar{f}, \ldots, f_T - \bar{f}] \in \mathbb{R}^{r \times n}$ and $\widetilde{E} = [e_1 - \bar{e}, \ldots, e_T - \bar{e}] \in \mathbb{R}^{N \times T}$. Then we can rewrite (16) into a matrix form given by

$$\widetilde{X} = \Lambda \widetilde{F} + \widetilde{E}. \tag{17}$$

Let $\sigma_i^2(A)$ denote the i-th largest eigenvalue, counting multiplicities, of the matrix $A'A$.

Theorem 3. *Suppose the data is generated from the genuine factor model (17) with $\sigma_1^2(\widetilde{E})/\sigma_r^2(\Lambda\widetilde{F}) = O_{\mathbb{P}}(N^{-\theta})$ for some $\theta > 0$, rather than the integrated system (6). For any sequence of level α with $\alpha^{-1} = o(\log N/h)$, our rejection rule (15) achieves asymptotically full power for any $k \geq r$, that is,*

$$\mathbb{P}(H_k > S_{1:k}^{-1}(\alpha/2)) \to 1.$$

Specifically, the result holds for all fixed $\alpha \in (0, 1)$ with any bounded bandwidth h.

For genuine factor models, our test based on Hill estimators works like a smoothed, cumulative adaptation of the eigenvalue ratio test introduced by Ahn and Horenstein (2013) that exploits the gap between $\widehat{\lambda}_r$ and $\widehat{\lambda}_{r+1}$. The Hill estimator $\widehat{\gamma}_r \xrightarrow{\mathbb{P}} \infty$ diverge at a rate of $\log(N)$, while the benchmark parameters $\gamma_r(d)$ remain uniformly bounded for all $d > 0$ within any compact set. Therefore, any vanishing bandwidth ensures the t-statistics $t_r(d)$ to explode at the rate of at least $\log(N)/h$, uniformly for $d > 0$ within any compact set. However, for any sequence $\alpha \in (0, 1)$, from Remark 3 we know that the critical value $S_{1:k}^{-1}(\alpha/2)$ is bounded by smaller order of $O(S_1^{-1}(\alpha/2)) = O(\alpha^{-1}) = o(\log N/h)$.

3 Spurious Factor Models Through Cross-Sectional Structure

Spurious factor analysis can be directly generated by a spatial near-unity-root process embedded in a cross-sectional structure rather than a time structure. The basic local correlation model from CG is a high-dimensional random Gaussian vector $X_t \sim \mathcal{N}(0, \sigma^2\Sigma)$ with the population covariance matrix $\Sigma = \{\Sigma(i, j) : 1 \leq i, j \leq N\}$ given by

$$\Sigma(i, j) = \frac{\rho^{|i-j|} - \rho^{i+j}}{1 - \rho^2} \tag{18}$$

for some local correlation coefficient $\rho \approx 1$ and unknown scale parameter $\sigma^2 > 0$. As CG points out, the actual dimension of local correlation models is the same as that for covariates, and approximating Σ by a factor-model-based estimator may lead to substantial inefficiency in economic analysis.

We generalize CG's local correlation models to allow for non-Gaussian shocks in the linear time series. Following their setup, we let observations $X_{t,i}$ in $X_t = (X_{t,1}, \ldots, X_{t,N})'$ obey a spatial moving average form, up to a permutation of coordinates, with some correlation coefficient $\rho \in (0, 1)$ given by

$$X_{t,i} = \sum_{s=1}^{i} \rho^{i-s} e_{t,s}, \quad e_t = (e_{t,1}, \ldots, e_{t,N})', \tag{19}$$

and the margins in e_t are N independent copies of linear time series given by

$$e_t - e_0 = \sum_{\ell=0}^{\infty} \psi_\ell \varepsilon_{t-\ell}, \quad \varepsilon_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,N})', \tag{20}$$

where $\{\psi_\ell \in \mathbb{R} : \ell = 0, 1, 2, \ldots\}$ are unknown parameters common for all individuals to maintain the structure of the correlation matrix, and $\{\varepsilon_{t,s}\}$ are i.i.d. standardized random variables satisfying Assumption 2.2. Observe that the cross-sectional covariance matrix still obeys the local correlation model with

$$\mathrm{cov}(X_t) = \sigma^2\Sigma, \quad \sigma^2 = \sum_{\ell=0}^{\infty} \psi_\ell^2,$$

where Σ is the same population covariance matrix defined in (18). Throughout we assume that σ^2 is bounded away from 0.

Let $\Phi \in \mathbb{R}^{N\times N}$ be an upper triangular Toeplitz matrix with first row $(1, \rho, \rho^2, \ldots, \rho^{N-1})$ satisfying the expansion

$$\Phi = \sum_{i=0}^{\infty} \rho^i L_N^i = \sum_{i=0}^{N-1} \rho^i L_N^i, \tag{21}$$

where $L_N \in \mathbb{R}^{N\times N}$ denotes the upper shift matrix like in (6) but of a different dimension. Let $\Psi' = \left(\sum_{\ell=0}^{\infty} \psi_\ell L_T^\ell\right) C \in \mathbb{R}^{T\times T}$ be the product of a upper triangular Toeplitz matrix with first row $(\psi_0, \ldots, \psi_{T-1})$ and the $T\times T$ centering matrix C. Then we can decompose the demeaned data matrix $\widetilde{X} = [X_1 - \bar{X}, \ldots, X_T - \bar{X}] \in \mathbb{R}^{N\times T}$ by

$$\widetilde{X} = \Phi'\varepsilon\Psi' + \Xi C, \quad \text{or} \quad \widetilde{X}' = \Psi\varepsilon'\Phi + C\Xi' \tag{22}$$

where the reminder matrix $\Xi = [\Xi_1, \ldots \Xi_T]$ has columns given by

$$\Xi_t = \sum_{s\leq 0} \psi_{t-s}\Phi'\varepsilon_s.$$

It is instructive to compare the representation of $\widetilde{X}'$ in (22) with the decomposition (6): the leading component $\Psi\varepsilon'\Phi$ is again a separable random matrix, and the reminder shall be negligible under the following assumption.

Assumption 3.1. The following conditions hold.

1. The local correlation $\rho \uparrow 1$ such that $N(1-\rho) \to \omega \in [0, \infty)$.

2. There exists a summable sequence of constants $\varphi_\ell \geq 0$, $\ell \in \mathbb{N}_0$, such that $\sum_{\ell=0}^{\infty} \varphi_\ell < \infty$ and $\psi_\ell^2 \leq \varphi_\ell \sum_{\ell=0}^{T-1} \psi_\ell^2$.

Condition 1 encompasses both scenarios of spatial unit root ($\omega = 0$) and spatial near unit root ($\omega > 0$). Our choice of the convergence rate allows for non-degenerate asymptotic analysis and aligns well with empirical investigations in the near-unit-root literature. When $0 < \rho < 1$ deviates significantly from 1, the spurious phenomenon caused by local correlation diminishes, falling outside the scope of interest for this paper. One may interpret condition 2 as a simple notion of the mixing property, which remains scale-free with respect to $\{\psi_\ell\}$ inflated by an arbitrary scale being divergent or not.

Recall the proportion of explained variance $\widehat{\xi}_k$ defined by (7) via the eigenvalues $\widehat{\lambda}_k$ of the sample covariance matrix. The following theorem gives their spurious probabilistic limits due to local correlations.

Theorem 4. *In large-dimensional locally correlated systems generated by* (19) *and* (20) *under Assumptions 2.1, 2.2, and 3.1, a spurious factor model emerges in such a way that the followings hold for every $k \geq 1$:*

(i) $\widehat{\xi}_k - \xi_k \xrightarrow{\mathbb{P}} 0$ *where* $\xi_k := \sigma_k^2 / \left(\sum_{i=1}^N \sigma_i^2\right)$ *are the explained variance ratios for the eigenvalues* σ_k^2 *of the population covariance matrix* $\Sigma = \Phi'\Phi$ *defined by* (18).

(ii) The population ratio ξ_k is bounded away from 0.

(iii) $|\widehat{V}_k' V_k| \xrightarrow{\mathbb{P}} 1$, where $\widehat{V}_k \in \mathbb{R}^N$ denotes the k-th principal eigenvector of $\widetilde{X}\widetilde{X}' \in \mathbb{R}^{N\times N}$ and V_k denotes the k-th principal eigenvector of $\Sigma = \Phi'\Phi$ defined by (18).

Figure 7 displays the finite-sample oracle cumulative explained variance ratio $\sum_{k=1}^r \xi_k$, $1 \le r \le 10$, for $N = 20$ and $N = 200$ as functions of $\omega = N(1-\rho)$. The variance concentrates heavier on the leading PCs as ω approaches 0. We refer the plots of eigenvectors to Figure 3 of CG.

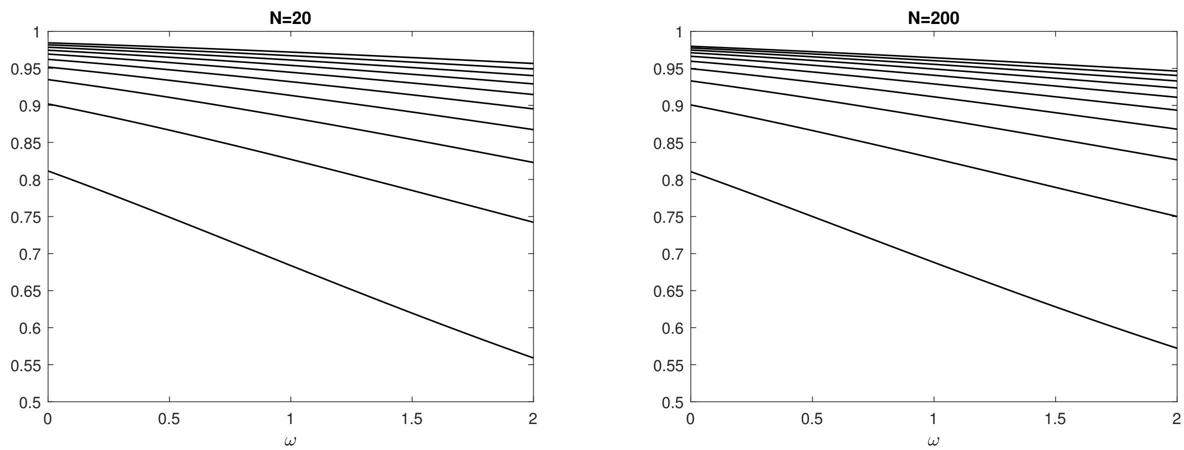


Figure 7: Cumulative oracle explained variance ratio $\sum_{k=1}^r \xi_k$ for $r = 1, \ldots, 10$ as functions of ω with $N = 20$ (left) and $N = 200$ (right).

Recall the eigenvalues $\widehat{\lambda}_1 \ge \ldots \ge \widehat{\lambda}_N$ of the sample covariance matrix $\widehat{\Sigma} = T^{-1}\widetilde{X}'\widetilde{X} \in \mathbb{R}^{N\times N}$. The following is an analogy of Theorem 2 and Corollary 1.

Theorem 5. *Under the conditions of Theorem 4, for any fixed $K \ge 1$,*

$$\sqrt{\frac{\nu}{2}}\left(\log\frac{\widehat{\lambda}_1}{\sigma_1^2} - \mu, \ldots, \log\frac{\widehat{\lambda}_{K+1}}{\sigma_{K+1}^2} - \mu\right) \Rightarrow (Z_1, \ldots, Z_{K+1})$$

with independent standard Gaussian variables $Z_1, \ldots, Z_{K+1}$ provided that $\sum_{\ell=0}^\infty \varphi_\ell^{2/3} < \infty$ and $\sum_{\ell=1}^\infty \ell^{1/2}\varphi_\ell < \infty$, where σ_k^2 denotes the k-th largest eigenvalue of the covariance matrix $\Sigma = \Phi'\Phi$ defined by (18) and

$$\nu = \frac{(\operatorname{tr}\Omega)^2}{\operatorname{tr}\Omega^2}, \quad \mu = \log\left(\frac{\operatorname{tr}\Omega}{T}\right), \quad \Omega = \Psi\Psi'.$$

The theorem remains true by replacing the condition $\sum_{\ell=1}^\infty \ell^{1/2}\varphi_\ell < \infty$ with a weaker version: $\sum_{\ell=1}^\infty \ell^{\iota/2}\varphi_\ell < \infty$ for some $\iota > 0$ such that $\nu = O(T^\iota)$. Note that we always have $\nu \le T$ by Cauchy-Schwarz inequality.

Corollary 1 remains true with the eigenvalues σ_k^2 and ratios ξ_k from Theorem 4.

The parameters $\nu = \nu(\Omega)$ and $\mu = \mu(\Omega)$ are the same functions of Ω as in Theorem 2, except now we substitute a different expression of Ω taking into the demeaning effect. Following the same arguments in the previous section, we can profile out the nuisance parameters μ and ν by considering the feasible t-statistics given by

$$t_{k-1}(\rho) = \sqrt{\frac{k}{k+1}}\frac{\widehat{\gamma}_k - \gamma_k(\rho)}{\widehat{s}_{k-1}(\rho)}, \quad k \geq 2,$$

by substituting the oracle parameters $\gamma_k(\rho)$ via the same formula (9) but with new eigenvalues $\sigma_i^2 = \sigma_i^2(\rho)$ here in Theorem 4 as functions of ρ, and $\widehat{s}_{k-1}(\rho)$ are the corresponding shrunk variance estimators like in (11). Similarly, we then scan over a sufficiently large parameter space to obtain the minimax statistic

$$H_k = \min_{\rho: N(1-\rho)\leq\bar{\omega}} \max_{1\leq i\leq k-1} |t_i(\rho)|, \; \bar{\omega} \in (\omega, \infty). \tag{23}$$

Now we can apply the same minimax test (15) that rejects the local correlation models if $H_k > S_{1:k}^{-1}(\alpha/2)$ for any appropriate choice of level α.

Theorem 6. *Corollary 3 remains true under the conditions of Theorem 5. Theorem 3 for genuine factor models still holds for the minimax statistics (23).*

In what follows, however, we focus on the case $\omega = 0$ that is $\rho = 1 - o(N^{-1})$ in the vicinity of 1 to let the first PC explain more than 80% of the total variation. This setting is most relevant to our bond data; see the calibrations in Figure 1 in the introduction.

Corollary 4. *Under the conditions of Theorem 5 with $N(1-\rho) \to 0$, uniformly for any finite collections of $k \geq 2$*

$$\max_{1\leq i\leq k-1} |t_i(1)| \Rightarrow \max_{1\leq i\leq k-1} |\mathrm{T}_i|,$$

where T_i are student variables of degrees of freedom i as defined in Corollary 2.

Hence, an asymptotically conservative test at any given level $\alpha \in (0,1)$ is to reject the local correlation models when

$$\max_{1\leq i\leq k-1} |t_i(1)| > S_{1:k}^{-1}(\alpha/2). \tag{24}$$

We will explore this test of cross-sectional unit root in simulation studies and empirical applications.

4 Simulation Study

We are considering three sets of Monte Carlo experiments for the integrated time series systems, locally correlated spatial systems, and genuine factor models, respectively. For simplicity, we choose a common level of $\alpha = 5\%$. We use the standard normal density function as the kernel G and set the bandwidth to be $h = 2\chi_1^2(1 - (\log N \cdot \log T)^{-1})/\max\{N, T\}$ where χ_1^2 denotes quantile function of chi-squared distribution with unit degrees of freedom. We report the empirical rejection rates over 50000 replications.

4.1 Factorless Integrated System

We generate the data from the integrated system (3) for a range of order $d \in \{1, 1.1, \ldots, 1.9\}$. We calibrate the dimensions $(N, T) = (58, 53)$ and an autoregressive model for the errors

$$u_t = 0.304u_{t-1} + 0.057u_{t-2} + 0.086u_{t-3} + 0.126u_{t-4} + v_t, \quad v_t \overset{iid}{\sim} \mathcal{N}(0, \Omega)$$

from the sectoral employment data in the introduction. The autoregressive coefficients are the pooled least-squares estimates for the log growth in sectoral employment. The dense covariance matrix Ω of the Gaussian vectors v_t is fitted to the idiosyncratic errors without the first three components in the level data of sectoral employment, using the nonlinear shrinkage estimator of Ledoit and Wolf (2012) in the QuEST package from the University of Zurich faculty website of Michael Wolf.

Table 1: Size for integrated systems at the level $\alpha = 5\%$ with $(N, T) = (58, 53)$

	$d = 1$	$d = 1.1$	$d = 1.2$	$d = 1.3$	$d = 1.4$	$d = 1.5$	$d = 1.6$	$d = 1.7$	$d = 1.8$	$d = 1.9$
$k = 2$	0.011	0.007	0.004	0.003	0.001	0.000	0.000	0.000	0.000	0.000
$k = 3$	0.013	0.009	0.005	0.003	0.002	0.001	0.001	0.001	0.001	0.002
$k = 4$	0.014	0.010	0.007	0.005	0.004	0.003	0.004	0.005	0.007	0.008
$k = 5$	0.017	0.013	0.010	0.009	0.009	0.011	0.013	0.016	0.019	0.022
$k = 6$	0.020	0.019	0.018	0.020	0.023	0.025	0.029	0.032	0.037	0.041
$k = 7$	0.025	0.026	0.028	0.033	0.037	0.041	0.045	0.049	0.053	0.058
$k = 8$	0.032	0.036	0.041	0.048	0.053	0.057	0.062	0.065	0.069	0.074
$k = 9$	0.040	0.047	0.055	0.061	0.066	0.072	0.077	0.080	0.084	0.088

Table 1 shows the empirical size of our minimax test (15) in nominal values, which tends to increase with our choice of k and the population order of integration d. Our tests are conservative in general, according to our asymptotic theory. The exceptions are the most challenging cases for large k and d, where we observe a minor oversized issue.

4.2 Locally Correlated System

We calibrate the dimensions $(N, T) = (19, 112)$ and generate the data from the local correlation model in Section 3 for a correlation $\rho \in \{0.98, 0.985, 0.99, 0.995, 0.999, 1\}$ close to 1. We generate the shocks from an autoregressive model

$$e_t = -0.017e_{t-1} - 0.079e_{t-2} + 0.100e_{t-3} + 0.041e_{t-4} + v_t, \quad v_t \overset{iid}{\sim} \mathcal{N}(0, I_N)$$

fitted to the standardized difference bond returns using the pooled least-squares estimator across all maturities.

Table 2 shows the empirical size of our cross-sectional unit root test (24) at the level $\alpha = 5\%$. The sizes are almost correct, especially for ρ closer to 1, even though the test is only conservative. For each

Table 2: Size for locally correlated systems at the level $\alpha = 5\%$ with $(N, T) = (19, 112)$

	$\rho = 0.98$	$\rho = 0.985$	$\rho = 0.99$	$\rho = 0.995$	$\rho = 0.999$	$\rho = 1$
$k = 2$	0.047	0.043	0.047	0.050	0.050	0.049
$k = 3$	0.045	0.042	0.046	0.050	0.049	0.048
$k = 4$	0.045	0.042	0.046	0.050	0.049	0.048
$k = 5$	0.045	0.042	0.046	0.050	0.049	0.048
$k = 6$	0.045	0.042	0.046	0.050	0.049	0.048
$k = 7$	0.045	0.042	0.046	0.050	0.049	0.048
$k = 8$	0.045	0.042	0.046	0.050	0.049	0.048
$k = 9$	0.045	0.042	0.046	0.050	0.049	0.048

given ρ, the rejection rates are similar over large k because the maximum of t statistics is most likely to be obtained at smaller k due to the ascending degrees of freedom of the limiting t distributions.

4.3 Genuine Factor Models

We consider the following genuine factor models that resemble the eigenvector plots of the spurious limits in Onatski and Wang (2021) and Crump and Gospodinov (2022).

1. We calibrate the dimensions $(N, T) = (58, 53)$ from the employment data, and let the $r = 3$ genuine factors $F = \sqrt{T}[U_1, U_2, U_3]' \in \mathbb{R}^{3\times 53}$ equal to the spurious factors, with the principal eigenvectors U_k of the matrix $\Phi'\Phi$ in Theorem 1 for $d = 1.6$. We generate independent entries in the loading matrix $\Lambda \in \mathbb{R}^{58\times 3}$ from the standard normal distribution and then permute the columns of Λ to ensure a descending order of total factor loadings.

2. We calibrate the dimensions $(N, T) = (19, 112)$ and the factor strengths $\{\sigma_k\}$ from the bond data. We set the genuine factor loadings $\Lambda = [\sigma_1 V_1, \sigma_2 V_2, \sigma_3 V_3] \in \mathbb{R}^{19\times 3}$ on $r = 3$ latent factors equal to the spurious limits, with the principal eigenvectors V_k of Σ in Theorem 4 for $\rho = 0.99$. We generate independent entries in the factor matrix $F = [f_1, \dots, f_T] \in \mathbb{R}^{3\times 112}$ from the standard normal distribution.

In both cases, we generate the measurement errors $e_t \in \mathbb{R}^N$ in (16) from an ARMA system:

$$e_t = 0.5e_{t-1} + v_t + 0.5v_{t-1}$$

with the entries of v_t as independent mean-zero Gaussian variables with a proper scale parameter yielding a (weaker) separation between factor and noise parts such that $\sigma_r^2(\Lambda\widetilde{F}) = N^\theta \sigma_1^2(\widetilde{E})$ in (17) with $\theta \in \{0.5, 0.6, \dots, 1\}$.

Tables 3 and 4 show that the empirical power of our tests increases with the separation rate θ between the signal and noise components, with a lift up for $k \geq 3$ according to our asymptotic theory.

Table 3: Power against integrated systems at the level $\alpha = 5\%$ with $(N, T, r) = (58, 53, 3)$

	$\theta = 0.5$	$\theta = 0.6$	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 1$
$k = 2$	0.637	0.635	0.634	0.634	0.633	0.633
$k = 3$	0.772	0.834	0.887	0.929	0.965	0.993
$k = 4$	0.768	0.831	0.883	0.926	0.962	0.991
$k = 5$	0.768	0.831	0.883	0.925	0.961	0.991
$k = 6$	0.768	0.830	0.883	0.925	0.961	0.991
$k = 7$	0.768	0.830	0.883	0.925	0.961	0.991
$k = 8$	0.768	0.830	0.883	0.925	0.961	0.991
$k = 9$	0.768	0.830	0.883	0.925	0.961	0.991

Table 4: Power against locally correlated systems at the level $\alpha = 5\%$ with $(N, T, r) = (19, 112, 3)$

	$\theta = 0.5$	$\theta = 0.6$	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 1$
$k = 2$	0.208	0.211	0.212	0.214	0.216	0.217
$k = 3$	0.564	0.720	0.936	0.998	1.000	1.000
$k = 4$	0.554	0.703	0.918	0.998	1.000	1.000
$k = 5$	0.552	0.700	0.914	0.998	0.999	1.000
$k = 6$	0.552	0.699	0.914	0.998	0.999	1.000
$k = 7$	0.552	0.699	0.913	0.998	0.999	1.000
$k = 8$	0.552	0.699	0.913	0.998	0.999	1.000
$k = 9$	0.552	0.699	0.913	0.998	0.999	1.000

The powers of our tests are almost full for the standard linear separate rate with θ close to unity while remaining arguably satisfactory even for the weaker factor models with smaller θ.

5 Empirical Analysis

Table 5 compares the inner products, in absolute value, of the sample eigenvectors and their spurious limits for the sectoral employment and bond datasets, respectively. Subject to sign change, the products on the left equal the unadjusted sample correlations between the sample and spurious factors (plotted in Figure 5) from integrated systems of order $d = 1.6$. Similar to OW, we observe correlations close to unity, in absolute value, for all three leading components, suggesting a resemblance of spurious factor models. The inner products on the right are for the sample and spurious principal component loadings for the locally correlated systems of $\rho = 0.99$. The first sample component (often called the level) is similar to the spurious limit, but the second and third sample components (often called the slope and curvature) are different. The angle between subspaces spanned by $\{\widehat{V}_1, \widehat{V}_2, \widehat{V}_3\}$ and

$\{V_1, V_2, V_3\}$ is more than 60 degrees, suggesting the sample space of these leading principal component loadings are different from that generated by the local correlation model.

Table 5: The absolute value of inner products between sample estimates and spurious limits

	Sectoral Employment				Difference Return on Bonds		
	U_1	U_2	U_3		V_1	V_2	V_3
$\widehat{U}_1$	0.986	0.080	0.021	$\widehat{V}_1$	0.937	0.326	0.099
$\widehat{U}_2$	0.083	0.939	0.096	$\widehat{V}_2$	0.332	0.726	0.404
$\widehat{U}_3$	0.023	0.073	0.941	$\widehat{V}_3$	0.107	0.551	0.289

Table 6 shows the test statistics for different choices of k based on the Hill plots in Figure 2. For the sectoral employment data, the null models are the collection of integrated systems with $d \geq 1$ in Corollary 2, and we report the minimax statistics H_k from (15). For the bond data, the null models are local correlation models with ρ very close to 1 in Corollary 4, and we report the cross-sectional unit root test statistics $\max_{1\leq i\leq k-1} |t_i(1)|$ from (24). The last row shows the conservative critical values at the level $\alpha = 5\%$ that are the same for both tests. Like the Hill plot, the test statistics stabilize for large k the sectoral employment data, and we cannot reject the integrated system. It is essential to regard integrated systems, nevertheless, solely as tools for providing the foundation for the 'problem of distribution', rather than as the actual data-generating processes. Further discussions will be deferred to the next section.

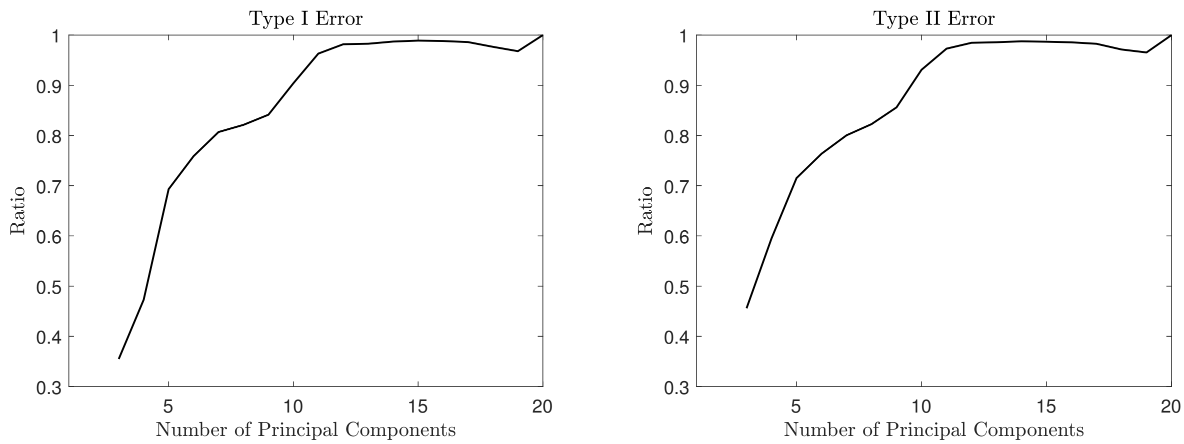
Table 6: Test statistics and critical values for spurious factor models

	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$
Employment	1.891	1.891	2.673	3.762	3.762	3.762	3.762	3.762
Bond	5.102	8.883	19.835	44.523	69.128	121.352	160.160	219.150
Critical Value	12.706	14.117	14.311	14.346	14.353	14.355	14.356	14.356

On the contrary, the difference return on bonds rejects the local correlation models for all $k \geq 4$, whereas the statistics show a similar explosive behavior as the Hill plot. To illustrate the economic implications, we revisit the mean-variance optimization example in CG. On the left of Figure 8 is a replicate of that in Figure 5 of CG showing the Sharpe ratios of our portfolio relative to the optimal value when suffering the type I errors in our test procedures, that is, using the incorrectly specified variance matrix based on a factor model rather than correctly specified variance matrix from the local correlation models.

We also examine the relative Sharpe ratios in cases of type II errors, wherein managers use incorrectly specified local correlation models instead of the correct variance matrix based on the factor model. We find that both types of errors incur substantial and comparable reduction of economic efficiency. This pattern holds across the range of values for ρ we considered in our simulation studies

Figure 8: Sharpe ratios from the optimal portfolio allocation using the factor model based variance matrix (alternative hypothesis) versus that based on the local correlation models (null hypothesis) with $\rho = 0.99$.



(i.e., $\rho \in [0.98, 1)$), with nearly identical results to what we are showing for $\rho = 0.99$. These findings highlight the importance for portfolio managers to rigorously test the hypothesis of local correlations to mitigate efficiency losses. Alternatively, at a possible cost of statistical efficiency, portfolio managers may consider employing a model-free procedure robust to an unknown factor structure, as suggested by CG.

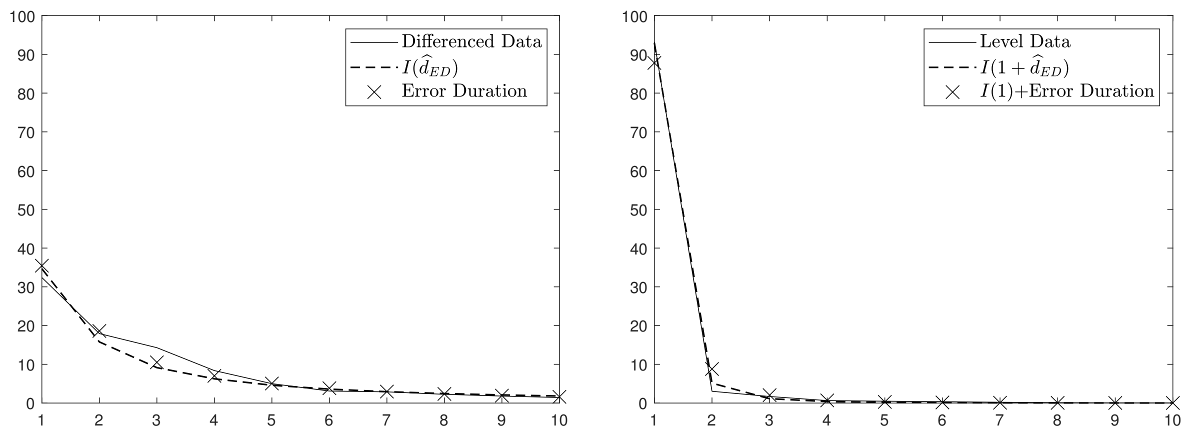
6 Discussions and Extensions

6.1 Why You Should Avoid Accepting Spurious Factor Models

Empirical researchers ought to be concerned when they are unable to reject spurious factor models. However, this does not necessarily mean one should accept these models as accurate representations of the data-generating processes. Our analysis of eigenvectors and tests on eigenvalues for the sectoral employment data solely indicate that the spectral characteristics of the data matrix cannot be distinguished with statistical significance from those generated by a high-dimensional integrated system of order around $d \approx 1.6$.

On the left of Figure 9 compares the scree plots for the differenced data simulated from the error duration model of Parke (1999) and a fitted integrated model, averaged over 500 sample paths, with that from observations. We calibrate the survival probabilities from total private establishments opened in 1994 provided by the Bureau of Labor Statistics (`https://www.bls.gov/bdm/bdmage.htm`), assuming the durations are the same over sectors, and fit an integrated order as in Parke (1999) given

Figure 9: Scree Plots for Differenced and Level Data of Sectoral Employment



by

$$\widehat{d}_{\mathrm{ED}} = \frac{1}{2} + \frac{1}{2}\frac{\log(p_5/p_{10})}{\log(5/10)} \approx 0.7,$$

where p_k denotes the probability of the event that the shocks ε_s survives until period $s+k$. We obtain the corresponding scree plots on the right when we aggregate the growth data into level data. These scree plots incorporate a fitted integrated system of order $1+\widehat{d}_{\mathrm{ED}} \approx 1.7$, which our test does not reject. The resemblance observed between the models with and without factors, encompassing the error duration model and the integrated system, serves as a cautionary signal to economists for potential spurious factor analysis, even after differencing.

6.2 Controlling for Time Trends

Let $\{Z_t\}$ be a high dimensional latent time series from one of our spurious factor models, but our observations $\{X_t\}$ contain time trends such as

$$X_t = X_0 + \beta^{(1)}t + \beta^{(2)}t^2 + Z_t, \quad \beta^{(j)} = (\beta_1^{(j)}, \ldots, \beta_N^{(j)})', \tag{25}$$

or, more generally, with a fixed r number of non-stochastic observable trends

$$X_t = \sum_{j=0}^{r} \beta^{(j)} W_{j,t} + Z_t, \quad W_t = (1, W_{1,t}, \ldots, W_{r,t})' \in \mathbb{R}^{1+r}, \quad \beta^{(0)} = X_0.$$

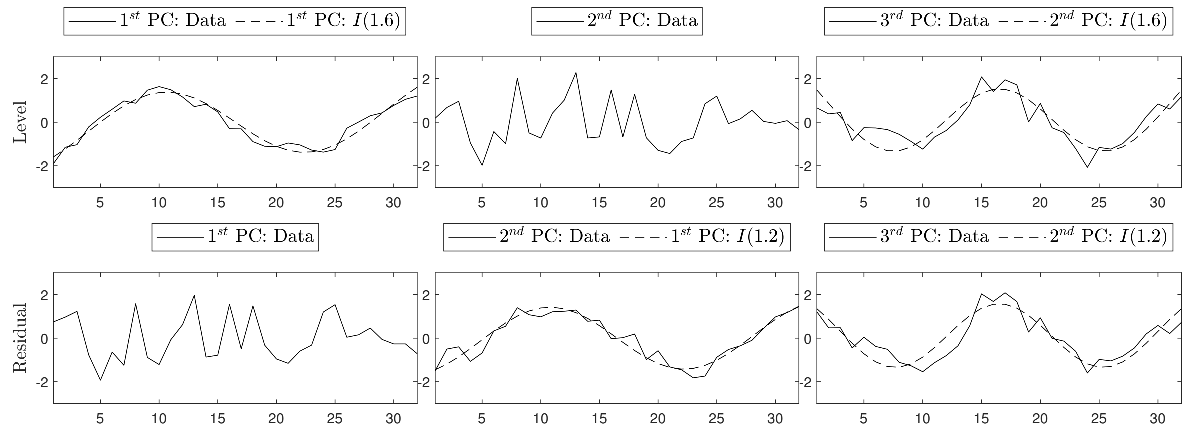
Note that we consider $r=0$ in previous sections, and (25) is a special case with $W_t = (1, t, t^2)'$. All our results remain true in Section 2 by substituting the demeaned matrix $\widetilde{X}$ everywhere with the demeaned and detrended matrix

$$\widetilde{X} = X(I - P_W) = \Psi\varepsilon\Phi + \Xi(I - P_W), \quad \Phi = (I - L_T)^{-d}(I - P_W) \tag{26}$$

where the projection matrix P_W on the trends is given by

$$P_W = W'(WW')^{-1}W, \quad W = [W_1, \ldots, W_T].$$

Figure 10: Estimated and Spurious Factors Controlling for Time Trends



Note that (26) generalizes the special form (6) where $r = 0$ and $I - P_W = C$, the centering matrix.

In Figure 10, we compare the top three estimated factors for U.S. divorce rates across $N = 48$ states and $T = 32$ years, controlling for trends and fixed effects in (25), as studied by Wolfers (2006), Kim and Oka (2014), and Moon and Weidner (2015), with spurious factors generated by an integrated system of the best fitted order according to our eigenvalue test. The residuals are from the linear regression on lagged observations and biannual dummy matrices in the aforementioned papers. For both level and residual data, two estimated factors correlate highly with the leading spurious factors, leaving only one that seems genuine and almost identical (the second estimated factor for level and the first for residual).

Our eigenvalue test aligns with the challenge Moon and Weidner (2015) highlighted in distinguishing between factor and nonfactor eigenvalues. While both the ED (edge distribution) criterion by Onatski (2010) and the ER (eigenvalue ratio) criterion by Ahn and Horenstein (2013) select one factor for the residual data, as reported by Moon and Weidner (2015), this genuine factor still exhibits a mild correlation of 46% in absolute value with the third spurious factor (not shown in the figure). This correlation calls for further investigation in future studies that should consider both genuine and spurious factors for the inference of linear regression coefficients.

6.3 When to Employ the Local Correlation Model

While the local correlation model may not be a perfect fit for the bond data discussed in Section 5, it can still serve a valuable purpose in explaining spurious factor analysis. For instance, consider the inflation forecast errors by the Survey of Professional Forecasters in the United States, currently maintained by the Federal Reserve Bank of Philadelphia. We directly obtained the forecast error data spanning $N = 8$ horizons (0.5,1.5,. . .,7.5 quarters) and $T = 40$ years (1982–2021) from Fabian Krüger's GitHub page (`https://github.com/FK83/gdp_intervals`). Table 7 shows that the top three sample principal component loadings correlate highly with the spurious limits derived from the local correlation model with $\rho = 1$. The right panel of the table indicates that the first principal component alone explains more than 90% of the variance. While these forecast errors display significant persistence across horizons (the pooled OLS estimate of the spatial AR1 coefficient is close to 1), we argue that there are no underlying factors. Otherwise, forecasters would have already extracted such factors.

Table 7: Inner Products Between Sample Loadings and Spurious Limits In Absolute-Value (Left) and Explained Variance Ratios (Right) for Professional Forecast Errors of Inflation.

	V_1	V_2	V_3
$\widehat{V}_1$	0.987	0.139	0.078
$\widehat{V}_2$	0.090	0.884	0.398
$\widehat{V}_3$	0.130	0.420	0.879

PC	Explained Variance (%)
1	90.909
2	4.305
3	2.357

7 Conclusion

Misidentifying factor and principal component analyses can lead to incorrect economic interpretations and substantial consequences. As concerns grow regarding the impact of non-stationarity and local correlations in generating spurious limits within high-dimensional data, our contribution is establishing a distribution theory that enables the formal testing of such spurious factor models. We introduce a self-normalized multiple t-test, using a finite sequence of Hill estimators to assess the tail heaviness of the spectral distribution. We show that the existing spurious factor models fit some but not all economic datasets. Our highly adaptive test procedure requires only sequences of sample statistics and benchmarks derived from the spurious limits. This adaptability suggests the potential for extending our methodology to broader categories of spurious factor models in future research, provided that corresponding spurious limits can be determined for data matrices that approximate a so-called separable form in random matrix theory.

A Proofs from Section 2

Let us first present some necessary lemmas but postpone their proofs to the supplement. The first lemma provides the order of the population eigenvalues σ_k^2 in Theorem 1 that generate the spurious concentrations ξ_k.

Lemma 1. *Denote the singular values of the $T \times T$ matrix $A = (I - L_T)^{-1}$ in descending order by $s_1 \geq \ldots \geq s_T \geq 0$ given by*

$$s_k^2 = \frac{1}{2(1 + \cos\theta_k)}, \quad \theta_k = \frac{2(T + 1 - k)}{2T + 1}\pi, \quad k = 1, \ldots, T.$$

There exists a constant $M = M(d) > 0$, depending only on true value of $d \geq 1$, satisfying the followings hold for all $k \geq 1$:

(i) $M^{-1} s^{2d}_{k+1+(\lceil d\rceil-1)\lceil d\rceil} \left(s_{k+1+(\lceil d\rceil-1)\lceil d\rceil}/T\right)^{\lceil 2d\rceil-2d} \leq \sigma_k^2 \leq MT^{2d}$.

(ii) $\sigma_k^2 \leq M s_k^2 T^{2d-2}$.

where $\lceil\cdot\rceil$ denotes the ceiling function. Furthermore, part (i) extends to all $d > 0$.

Part i shows that σ_k^2 diverge at same order of T^{2d} for each fixed k, whereas part ii shows the decay rate of the σ_k^2 over k is at least as fast as s_k^2.

The second lemma controls the perturbation from the reminder matrix Ξ in the BN decomposition (6). It is a high-dimensional analogy of Lemma A.5 in Phillips and Shimotsu (2004) for univariate non-stationary time series.

Lemma 2. *There exists a constant $M = M(d) > 0$ such that for all t*

$$\mathbb{E}\left\|\Xi_t\right\|^2 \leq M\left\|\Psi\right\|^2 \left\{t^{2d-3}\sum_{s=0}^{t-2} s\bar{\Lambda}_s + t^{2d-1}\bar{\Lambda}_t\right\}$$

where

$$\bar{\Lambda}_t = \sum_{\ell\geq t+1}\varphi_\ell + \frac{1}{t+1}\sum_{\ell=0}^{t}\ell\varphi_\ell,$$

and $\|\cdot\|$ denotes the Frobenius norm.

Proof of Theorem 1. Our proof is similar to that of Theorem 1 in Onatski and Wang (2021), abbreviated as OW, and thus we only sketch the differences. Let $d \geq 1$ be fixed. Observe that the eigenvalues s_k^2 defined in Lemma 1 can be bounded by a power law in the sense that $s_k^2 \leq Mk^{-2}T^2$ for all k and some constant $M > 0$. Then Lemma 1 implies the existence of a common constant M and a particular $M_k > 0$ for every k such that

$$M_k T^{2d} \leq \sigma_k^2 \leq \sum_{i=1}^{T}\sigma_i^2 \leq M\zeta(2)T^{2d}, \tag{27}$$

where $\zeta(\cdot)$ is the Riemann zeta function given by $\zeta(s) = \sum_{i=1}^{\infty} i^{-s}$ with $\zeta(2) = \pi^2/6 < \infty$. This immediately gives part (ii) of the theorem because σ_k^2 and $\sum_{i=1}^{T}\sigma_i^2$ are diverging at the same order of T^{2d} for every k. On the other hand, the small eigenvalues are negligible by part ii of Lemma 1 in the sense that: for any $\delta > 0$, one can find a $K(\delta)$ such that for all large N, T

$$\sum_{i\geq K(\delta)+1}\sigma_i^2 \leq \delta T^{2d}. \tag{28}$$

Next, compare the $T \times T$ sample covariance matrix of the leading term in (6) denoted by $\widetilde{\Sigma} := \Phi'\varepsilon'\Psi'\Psi\varepsilon\Phi/T = \Phi'\varepsilon'\underline{\Omega}\varepsilon\Phi/T$ here with the matrix $\widetilde{\Sigma}$ in equation (12) of OW, where we replace their matrices Ψ and $U = (I - L_T)^{-1}$ with a more general definitions of Ψ and $(I - L_T)^{-d}$ here. We also renamed their matrix W with $\underline{\Omega}$ here and replaced their N with T without influencing the results. Let $\widetilde{\sigma}_k^2$ and $\widetilde{U}_k$ denote the k-th largest eigenvalue and the associated eigenvector of $\widetilde{\Sigma}$, respectively. The consistency of $\widetilde{\xi}_k^2 := \widetilde{\sigma}_k^2 / \sum_{i=1}^n \widetilde{\sigma}_i^2$ and $\widetilde{U}_k$ then follows by the same arguments in OW using (27), (28) and the separability Assumption 2.4. In particular, for all $1 \leq k \leq K+2$

$$\frac{\widetilde{\sigma}_k^2}{\sigma_k^2 \cdot \operatorname{tr}\Omega / T} - 1 \xrightarrow{\mathbb{P}} 0, \tag{29}$$

and

(i) $\widetilde{\xi}_k = \xi_k(1 + o_{\mathbb{P}}(1))$.

(ii) $\left|\widetilde{U}_k' U_k\right| \xrightarrow{\mathbb{P}} 1$.

Let $\|\cdot\|_{\mathrm{sp}}$ denote the spectral norm and $\|\cdot\|$ denote the Frobenius norm. Moreover, as explained in OW, extending these results to $\widehat{\xi}_k$ and $\widehat{U}_k$ only requires verifying that the perturbation term in (6) is negligible in the sense that

$$\|\Xi C\|_{\mathrm{sp}}^2 = o_{\mathbb{P}}(T^{2d} \|\Psi\|^2), \quad \text{or} \quad \|\Xi\|_{\mathrm{sp}}^2 = o_{\mathbb{P}}(T^{2d} \|\Psi\|^2),$$

where the order T^{2d} comes from (27) for leading eigenvalues σ_k^2 and $\|\Psi\|^2 = \operatorname{tr}(\Psi'\Psi) = \operatorname{tr}\Omega$. It suffices to show that

$$\mathbb{E}\|\Xi\|^2 = \sum_{t=1}^T \mathbb{E}\|\Xi_t\|^2 = o(T^{2d} \|\Psi\|^2).$$

The last equality follows by summing up the upper bound on $\mathbb{E}\|\Xi_t\|^2$ from Lemma 2, using the summability $\sum_{\ell=0}^\infty \varphi_\ell < \infty$ and the Kronecker's lemma. □

To establish the distribution theory in Theorem 2, we first verify a Lyapunov condition on the eigenvectors of the matrix $\Phi\Phi' = (I - L_T)^{-d} C (I - L_T')^{-d} \in \mathbb{R}^{T\times T}$. From now on we assume all the conditions of Theorem 2.

Lemma 3. *Let $w_k = (w_{k,1}, \ldots, w_{k,T})'$ denote the k-th principal eigenvector of $\Phi\Phi'$. For every $k \geq 1$, $\sum_{j=1}^T w_{k,j}^4 / (\sum_{j=1}^T w_{k,j}^2)^2 = \sum_{j=1}^T w_{k,j}^4 = O(T^{-1})$.*

With this Lyapunov condition we then establish the asymptotic approximation of sample eigenvalues, similar to that in Proposition 5 of Zhang, Pan and Gao (2018). Our proof is even more involved and available in the supplement. In particular, the negligibility of the reminder term ΞC in (6) in the next lemma relies on the extra condition $\sum_{\ell=1}^\infty \ell^{\iota/2}\varphi_\ell < \infty$ that allows for a polynomial decay of $\bar{\Lambda}_t$ as $t \to \infty$:

$$\bar{\Lambda}_t \leq \frac{1}{(t+1)^{\iota/2}} \sum_{\ell \geq t+1} \ell^{\iota/2}\varphi_\ell + \frac{1}{(t+1)^{\iota/2}} \cdot \frac{1}{(t+1)^{1-\iota/2}} \sum_{\ell=0}^t \ell^{1-\iota/2} \cdot \ell^{\iota/2}\varphi_\ell = o(t^{-\iota/2}), \tag{30}$$

by the summability of $\ell^{1/2}\varphi_\ell$ and Kronecker's lemma. This decay rate is fast enough, relative to $\nu = O(N^\iota)$, to make the reminder term ΞC negligible in probability via the Markov inequality and Lemma 2.

Lemma 4. *Recall that* $\underline{\Omega} = \Psi'\Psi$ *and* $\nu = \frac{(\operatorname{tr}\Omega)^2}{\operatorname{tr}\Omega^2}$. *For all* $1 \le k \le K+1$,

$$\sqrt{\frac{\nu}{2}}\left(\frac{T}{\operatorname{tr}\Omega}\frac{\widehat{\lambda}_k}{\sigma_k^2} - 1\right) = \widehat{Z}_k + o_{\mathbb{P}}(1), \quad \widehat{Z}_k = \sqrt{\frac{\nu}{2}}\left(\frac{1}{\operatorname{tr}\Omega} w_k' \varepsilon' \underline{\Omega} \varepsilon w_k - 1\right),$$

where

$$\mathbb{E}\widehat{Z}_k = 0, \quad \operatorname{var}\left(\widehat{Z}_k\right) = 1 + o(1). \tag{31}$$

Note that $\widehat{Z}_k$ are quadratic forms, subject to normalization, of $\bar{\varepsilon}_k := \varepsilon w_k = (\bar{\varepsilon}_{k,1}, \ldots, \bar{\varepsilon}_{k,N})'$. For every k, $\bar{\varepsilon}_k$ has independent entries with zero mean and unit variance. The following corollary of Lemma 3, particularly part i, shows that our standardization of the quadratic forms are asymptotically correct in view of (31). The other parts of this corollary are useful for establishing the asymptotic independence between $\widehat{Z}_k$.

Corollary 5. *The following holds uniformly for* $1 \le i \le N$ *and all* $k \ge 1$.

(i) *The excess kurtosis of* $\bar{\varepsilon}_{k,i}$ *vanish in such a way that* $\mathbb{E}\bar{\varepsilon}_{k,i}^4 - 3 = O(T^{-1})$.

(ii) *For all* $k' \neq k$, $\operatorname{cov}(\bar{\varepsilon}_{k,i}^2, \bar{\varepsilon}_{k',i}^2) = O(T^{-1})$.

(iii) *For all* k', $\operatorname{cov}(\bar{\varepsilon}_{k,i}^2, \bar{\varepsilon}_{k',i}) = O(T^{-1/2})$.

For Theorem 2 it remains to show that $(\widehat{Z}_1, \ldots, \widehat{Z}_{K+1}) \Rightarrow (Z_1, \ldots, Z_{K+1})$ with independent standard normal variables Z_k. Represent the standardized quadratic forms $\widehat{Z}_k = \sum_{i=1}^N D_{k,i}$ from Lemma 4 in a martingale form with

$$D_{k,i} = \frac{1}{\sqrt{2\operatorname{tr}\Omega^2}}\left\{\underline{\Omega}_{i,i}(\bar{\varepsilon}_{k,i}^2 - 1) + 2\bar{\varepsilon}_{k,i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s}\right\}. \tag{32}$$

Let $\mathcal{F}_{p,i}$ denote the sigma-algebra generated by $\{\bar{\varepsilon}_{k,s} : 1 \le k \le K+1, 1 \le s \le i\}$. For every $1 \le k \le K+1$, $\{D_{k,i}, \mathcal{F}_{p,i} : i = 1, \ldots, N\}$ is an adapted array and $D_{k,i}$ is a martingale difference array. Note that the 'time' axis of these martingales is the individual index $1 \le i \le N$ rather than the real time axis for our data collection. We shall establish the joint convergence of $\widehat{Z}_k$ by martingale central limit theorems and Cramér-Wold device. The following lemma will allow us to verify the conditions in Corollary 3.1 in Hall and Heyde (1980) for linear combinations of $Z_1, \ldots, Z_{K+1}$.

Lemma 5. *The following hold for the martingale difference arrays* $\{D_{k,i} : 1 \le i \le N\}$ *over* $1 \le k \le K+1$.

(i) $\sum_{i=1}^N \mathbb{E}\left[D_{k,i}D_{k',i} \mid \mathcal{F}_{p,i-1}\right] \xrightarrow{\mathbb{P}} 0$ *for all* $k \neq k'$.

(ii) $\sum_{i=1}^N \mathbb{E}\left[D_{k,i}^2 \mid \mathcal{F}_{p,i-1}\right] \xrightarrow{\mathbb{P}} 1$.

(iii) $\max_{1 \le i \le N} \mathbb{E}[D_{k,i}^2 \mid \mathcal{F}_{p,i-1}] \xrightarrow{\mathbb{P}} 0$.

Now we are ready to prove Theorem 2.

Proof of Theorem 2. We shall prove the theorem using Cramér-Wold device. Let $c = (c_1, \dots, c_{K+1})' \in \mathbb{R}^{K+1}$ be any fixed projection vector with unit length $\|c\| = 1$, without loss of generality. We use Corollary 3.1 in Hall and Heyde (1980) to establish that

$$\sum_{k=1}^{K+1} c_k \widehat{Z}_k = \sum_{i=1}^{N} c' D_i \Rightarrow \mathcal{N}(0,1),$$

where $D_i = (D_{1,i}, \dots, D_{K+1,i})'$ and its entries are defined by (32), and $\mathcal{N}(0,1)$ denotes a standard normal variable. In particular, we verify the following two conditions:

(a) $\sum_{i=1}^{N} \mathbb{E}[c' D_i D_i' c \mid \mathcal{F}_{p,i-1}] \xrightarrow{\mathbb{P}} 1.$

(b) For all $\delta > 0$, $\sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2 \mathbb{1}[|c'D_i| > \delta] \mid \mathcal{F}_{p,i-1}\right] \xrightarrow{\mathbb{P}} 0,$

with the sigma-algebras $\mathcal{F}_{p,i-1}$ defined in Lemma 5. Combining parts i and ii of Lemma 5 gives that

$$\sum_{i=1}^{N} \mathbb{E}\left[D_i D_i' \mid \mathcal{F}_{p,i-1}\right] \xrightarrow{\mathbb{P}} I_{K+1}$$

where the limit is an identity matrix. Then condition a follows from the Slutsky theorem.

Now take any $\delta > 0$. By the law of iterated expectations, almost surely

$$\begin{aligned}
&\sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2 \mathbb{1}[|c'D_i| > \delta] \mid \mathcal{F}_{p,i-1}\right] \\
\le &\sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2 \cdot 1 \mid \mathcal{F}_{p,i-1}, |c'D_i| > \delta\right] \mathbb{P}\left(|c'D_i| > \delta \mid \mathcal{F}_{p,i-1}\right) \\
\le &\sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2 \cdot 1 \mid \mathcal{F}_{p,i-1}, |c'D_i| > \delta\right] \cdot \max_{1 \le i \le N} \mathbb{P}\left(|c'D_i| > \delta \mid \mathcal{F}_{p,i-1}\right).
\end{aligned}$$

The first term

$$\sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2 \cdot 1 \mid \mathcal{F}_{p,i-1}, |c'D_i| > \delta\right] = O_{\mathbb{P}}(1)$$

by Markov inequality and the law of iterated expectations

$$\mathbb{E}\left[\sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2 \cdot 1 \mid \mathcal{F}_{p,i-1}, |c'D_i| > \delta\right]\right] = \sum_{i=1}^{N} \mathbb{E}\left[(c'D_i)^2\right] = O(1).$$

To bound the probability term $\max_{1 \le i \le N} \mathbb{P}(|c'D_i| > \delta \mid \mathcal{F}_{p,i-1})$, use the Cauchy–Schwarz inequality and Bonferroni method to get that

$$\mathbb{P}\left(|c'D_i| > \delta \mid \mathcal{F}_{p,i-1}\right) \le \mathbb{P}\left(\|D_i\| > \delta \mid \mathcal{F}_{p,i-1}\right) \le \sum_{k=1}^{K+1} \mathbb{P}\left(|D_{k,i}| > \delta/\sqrt{K+1} \mid \mathcal{F}_{p,i-1}\right).$$

Exchanging the maximum and summation operations and applying the Markov inequlity,

$$\begin{aligned}
\max_{1 \le i \le N} \mathbb{P}\left(|c'D_i| > \delta \mid \mathcal{F}_{p,i-1}\right) &\le \sum_{k=1}^{K+1} \max_{1 \le i \le N} \mathbb{P}\left(|D_{k,i}| > \delta/\sqrt{K+1} \mid \mathcal{F}_{p,i-1}\right) \\
&\le \sum_{k=1}^{K+1} \frac{K+1}{\delta^2} \max_{1 \le i \le N} \mathbb{E}\left[D_{k,i}^2 \mid \mathcal{F}_{p,i-1}\right] \xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

using part iii of Lemma 5 in the end. This completes the proof of condition b. □

B Proofs from Section 3

The proofs are completely analogous to that in the last subsection, but we need to establish the corresponding conditions. We begin with the following lemma from the unpublished manuscript by Zhang, Gao and Pan (2020) for the population eigenvalues of the covariance matrix Σ from (18).

Lemma 6. *The eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_N^2 \geq 0$ of Σ are given by*

$$\sigma_k^2 = \frac{1}{1+\rho^2 - 2\rho\cos(\theta_k)}$$

with $\theta_1 < \theta_2 < \ldots < \theta_N$ being the solutions of the equation

$$\rho \sin(N(\pi-\theta)) + \sin((N+1)(\pi-\theta)) = 0, \quad \theta \in (0,\pi),$$

such that $\theta_k \in \left[\frac{2k-1}{2N+1}\pi, \frac{k}{N+1}\pi\right]$.

Combining this lemma with part 1 of Assumption 3.1, we find the bounds for all extreme population eigenvalues by expanding the cosine function at the origin. We can also obtain the decay rate of σ_k^2 over k for any given N.

Corollary 6. *For every $k \geq 1$ and any $\delta > 0$, for all large N, T,*

$$(1-\delta)N^2\left(\omega^2 + (k-1/2)^2\pi^2/2\right)^{-1} \leq \sigma_k^2 \leq (1+\delta)N^2\left(\omega^2 + k^2\pi^2/2\right)^{-1}.$$

Moreover, there exists a constant $M > 0$ such that uniformly for all $1 \leq k \leq N$

$$\sigma_k^2 \leq MN^2k^{-2}.$$

The next lemma establishes the necessary bounds of the trace, spectral norm and Frobenius norm of $\underline{\Omega} = \Psi'\Psi$ for our limit theorems, subject to normalization, via the local coefficients $\{\varphi_\ell\}$ from Assumption 3.1.

Lemma 7. *Let $\underline{\Omega}_{(T)} = \underline{\Omega}/\sum_{\ell=0}^{T-1}\psi_\ell^2$ be an adaptive normalization of $\underline{\Omega}$. The followings hold for all N, T.*

(i) $\frac{1}{4}T \leq \operatorname{tr}\underline{\Omega}_{(T)} \leq T$

(ii) $\lambda_{\max}\left(\underline{\Omega}_{(T)}\right) \leq \left(\sum_{\ell=0}^{T-1}\sqrt{\varphi_\ell}\right)^2$

(iii) $\left\|\underline{\Omega}_{(T)}\right\| \geq \frac{1}{4}T^{1/2}$ *where $\|\cdot\|$ denotes the Frobenius norm.*

Combining this lemma with the next one yields the desired orders of the spectral and Frobenius norms of $\underline{\Omega}$.

Lemma 8. $\sum_{\ell=0}^{T-1}\sqrt{\varphi_\ell} = o\left(T^{\frac{2\iota-1}{2\iota}}\right)$ *if $\sum_{\ell=0}^{\infty}\varphi_\ell^\iota < \infty$ for any $\iota \in (1/2, 1]$*

In particular, for $\iota = 1$, Assumption 2.5 holds with $\underline{\Omega}$ substituting for Ω:

$$\frac{\lambda_{\max}(\underline{\Omega})}{\operatorname{tr}\underline{\Omega}} = \frac{\lambda_{\max}(\underline{\Omega}_{(T)})}{\operatorname{tr}\underline{\Omega}_{(T)}} = o\left(\frac{T}{T}\right) = o(1) \to 0. \tag{33}$$

For $\iota = 2/3$, Assumption 2.6 holds with $\underline{\Omega}$ substituting for Ω:

$$\frac{\lambda_{\max}(\underline{\Omega})}{\|\underline{\Omega}\|} = \frac{\lambda_{\max}(\underline{\Omega}_{(T)})}{\left\|\underline{\Omega}_{(T)}\right\|} = o\left(\frac{T^{1/2}}{T^{1/2}}\right) = o(1) \to 0. \tag{34}$$

Proof of Theorem 4. The proof is completely analogous to that of Theorem 1, even easier, and we only sketch the differences. Corollary 6 re-establishes the inequalities (27) and (28) by substituting $d = 1$ therein and the separability condition, Assumption 2.4, for all fixed K. Together with the factorless condition (33), the stated results hold for the explained ratios $\widetilde{\xi}_k$ and eigenvectors $\widetilde{V}_k$ of the Gram matrix $\Phi'\varepsilon'\underline{\Omega}\varepsilon\Phi$ of the leading term in (22). Extending the results to $\widehat{\xi}_k$ and $\widehat{V}_k$ requires the reminder in (22) to be negligible. Following the arguments in the end of the proof of Theorem 1, a sufficient condition is that $\mathbb{E}\|\Xi\|^2 = o(N^2 \operatorname{tr}\underline{\Omega})$, where N^2 comes from the order of extreme eigenvalues σ_k^2 for all fixed k given in Corollary 6. A direct calculation yields that

$$\mathbb{E}\|\Xi\|^2 = \sum_{t=1}^{T} \mathbb{E}\|\Xi_t\|^2 = \|\Phi\|^2 \cdot \sum_{t=1}^{T}\sum_{\ell \geq t}^{\infty} \psi_\ell^2 = \|\Phi\|^2 \cdot \sum_{\ell=1}^{\infty} \min\{\ell, T\}\psi_\ell^2.$$

The second part of Corollary 6 gives that $\|\Phi\|^2 = \sum_{k=1}^{N} \sigma_k^2 = O(N^2 \sum_{k=1}^{N} k^{-2}) = O(N^2)$. Moreover, recall the definition of $\bar{\Lambda}_{T-1}$ from Lemma 2 given by

$$\bar{\Lambda}_{T-1} = \sum_{\ell=T}^{\infty} \varphi_\ell + \frac{1}{T}\sum_{\ell=1}^{T-1} \ell\varphi_\ell,$$

such that $\bar{\Lambda}_{T-1} = o(1)$ by the summability of $\{\varphi_\ell\}$ and the Kronecker's lemma. Hence, again using part 2 of Assumption 3.1,

$$\sum_{\ell=1}^{\infty} \min\{\ell, T\}\psi_\ell^2 \leq \bar{\Lambda}_{T-1} T \left(\sum_{\ell=1}^{T-1} \psi_\ell^2\right) = o\left(T \sum_{\ell=0}^{T-1} \psi_\ell^2\right) = o(\operatorname{tr}\underline{\Omega})$$

where the last step is due to part i of Lemma 7. □

From now on we assume all the conditions of Theorem 5. The following lemma is an analogy of Lemma 3 for the local correlation models.

Lemma 9. *Let $w_k = (w_{k,1}, \ldots, w_{k,N})'$ denote the k-th principal eigenvector of $\Phi\Phi'$. For every $k \geq 1$, $\sum_{i=1}^{N} w_{k,i}^4 / (\sum_{i=1}^{N} w_{k,i}^2)^2 = \sum_{i=1}^{N} w_{k,i}^4 = O(N^{-1})$.*

Proof of Theorem 5. Since the Gram matrices $\widehat{\Sigma} = T^{-1}\widetilde{X}'\widetilde{X}$ and $\underline{\widehat{\Sigma}} = T^{-1}\widetilde{X}\widetilde{X}'$ share the same eigenvalues, we only need to work on the transpose of $\widetilde{X}$, denoted by $\widetilde{X}' \in \mathbb{R}^{T\times N}$, obeying BN decomposition (22) given by

$$\widetilde{X}' = \Psi\varepsilon'\Phi + C\Xi'$$

with $\underline{\Omega} = \Psi'\Psi$. The rest is completely analogous to that of Theorem 2, swapping individual index $1 \le i \le N$ with the time index $1 \le t \le T$, with Assumption 2.6 satisfied via (34). Specifically, apply Lemma 3 instead of Lemma 9 for the Lyapunov condition to re-establish Lemma 4 (swapping the dimensions T and N). To make the reminder term $C\Xi'$ (or, equivalently, its transpose ΞC) negligible it suffices to bound its spectral norm via its Frobenius norm as follows:

$$\|C\Xi'\|_{\text{sp}}^2 \le \|\Xi'\|_{\text{sp}}^2 \le \|\Xi'\|^2 = \|\Xi\|^2 = N^2 \operatorname{tr}\Omega \cdot o_{\mathbb{P}}(\nu^{-1/2}),$$

where the last step follows from the Markov inequality, because from the proof of Theorem 5 we know that

$$\mathbb{E}\,\|\Xi\|^2 = N^2 \operatorname{tr}\Omega \cdot O(\bar{\Lambda}_{T-1}) = N^2 \operatorname{tr}\Omega \cdot o(\nu^{-1/2})$$

according to (30) with $\nu = O(T^{\iota})$. □

C Proofs of Lemmas 1– 5 and Corollary 5

C.1 Proof of Lemma 1

We denote by $s_k(A)$ the k-th largest singular values (counting multiplicities) of matrix A and set $s_k(A) = 0$ if $k > \operatorname{rank}(A)$, where $\operatorname{rank}(A)$ denotes the rank of matrix A. Note that oracle parameters $\sigma_k = s_k(\Phi)$ in Theorem 1. The following lemma is due to Fan (1951).

Lemma S1. *Let A_1 and A_2 be two real matrices. Then for any positive integers $i, j \ge 1$,*

(i) $s_{i+j-1}(A_1A_2) \le s_i(A_1)s_j(A_2)$

(ii) $s_{i+j-1}(A_2) \le s_i(A_1) + s_j(A_2 - A_1)$

In particular, for $j > \operatorname{rank}(A_2 - A_1)$, $s_{i+j-1}(A_2) \le s_i(A_1)$.

Lemma S2. *Let Γ denote the gamma function. For any $\alpha \in \mathbb{R}$, $\Gamma(x+\alpha) \sim \Gamma(x)x^{\alpha}$ as $x \to \infty$. It follows that, for any $d > 0$, there exists a constant $M = M(d) \ge 1$, such that $a_\ell \le M \max\{1, \ell\}^{d-1}$ for all $\ell \in \mathbb{N}_0$, where a_ℓ are defined in (5).*

Proof. The first part is a well-known corollary of Watson's lemma; see, e.g., Chapter 4 of Bleistein and Handelsman (1975). The second part follows as for all $\ell \ge 1$

$$a_\ell = \frac{\Gamma(d+\ell)}{\ell!\Gamma(d)} = \frac{\Gamma(d+\ell)}{\Gamma(\ell)\ell^d}\frac{\ell^d}{\ell\Gamma(d)} \le M\ell^{d-1}.$$

The case for $k = 0$ trivial with $a_0 = 1 \le M$. □

Lemma S3. *Let $A = (I - L_T)^{-1}$ denote the $T \times T$ upper triangular matrix with ones on the diagonal and upper part. Define the matrix $H = 2I - L_T - L_T'$ as the $T \times T$ symmetric Toeplitz tridiagonal matrix with 2 on the diagonal, -1 on the sub and super diagonals, and 0 elsewhere. Then*

$$A'A = H^{-1} + \frac{1}{1+T}A'A\mathbf{e}_T\mathbf{e}_T'A'A$$

and, similarly,

$$AA' = H^{-1} + \frac{1}{1+T} AA'\mathbf{e}_1\mathbf{e}_1' AA'$$

where $\mathbf{e}_t = (0, \ldots, 0, 1, 0, \ldots, 0)' \in \mathbb{R}^T$ *denote the basic vector with* 1 *on the* t*-th entry but zeros elsewhere. It follows that* $\operatorname{rank}(A'A - AA') \leq 2$.

Proof. Write L in short of L_T. Observe that

$$A'A = (I - L')^{-1} (I - L)^{-1} = (I - L - L' + I - \mathbf{e}_T\mathbf{e}_T')^{-1}$$

and

$$AA' = (I - L)^{-1} (I - L')^{-1} = (I - L - L' + L'L)^{-1} = (I - L - L' + I - \mathbf{e}_1\mathbf{e}_1')^{-1}.$$

The rest follows from the Sherman-Morrison formula and the identities $\mathbf{e}_T'A'A\mathbf{e}_T = \mathbf{e}_1'AA'\mathbf{e}_1 = T$. □

Lemma S4. *For all integers* $T, d \geq 1$,

$$\operatorname{rank}\left((A')^d A^d - (A'A)^d\right) \leq 2\sum_{i=1}^{d-1} i = (d-1)d.$$

Proof. We prove by induction over d. The statement is trivial for $d = 1$. The statement holds for $d = 2$ by Lemma S3. Suppose the statement holds for $d = k \geq 2$, and consider the case for $d = k + 1$. We denote by a symmetric matrix $M = O_{\text{rank}}(r)$ if $\operatorname{rank}(M) \leq r$. By the induction hypothesis,

$$\begin{aligned}
(A')^{k+1}A^{k+1} =& A'\left[(A')^k (A)^k\right] A \\
=& A'(A'A)^k A + O_{\text{rank}}((k-1)k) \\
=& A'\left[\prod_{i=1}^{k} (AA' + O_{\text{rank}}(2))\right] A + O_{\text{rank}}((k-1)k) \\
=& (A'A)^{(k+1)} + O_{\text{rank}}(2k) + O_{\text{rank}}((k-1)k) \\
=& (A'A)^{(k+1)} + O_{\text{rank}}(k(k+1)).
\end{aligned}$$

This is the statement for $d = k + 1$. □

Lemma S5. *For all positive integers* $d, k \geq 1$,

$$s_1^{2(d-1)} s_k^2 \geq \sigma_k^2 \geq s_{k+1}^2\left(A^d\right) \geq s_{k+1+(d-1)d}^{2d}.$$

Proof. Applying part i of Lemma S1 yields the upper bound, namely

$$\sigma_k = s_k(A^d C) \leq s_k(A^d)s_1(C) \leq s_k(A)[s_1(A)]^{d-1}s_1(C) = s_k s_1^{d-1},$$

where we used the fact that $s_1(C) = 1$ in the last step. Next, note that $I - C = T^{-1}\mathbf{1}_T\mathbf{1}_T$ has only rank 1 where $\mathbf{1}_T$ denotes the T-dimensional all-ones vector. Applying part ii of Lemma S1 yields one lower bound:

$$\sigma_k = s_k(A^d - A^d(I - C)) \geq s_{k+1}\left(A^d\right).$$

Let $\lambda_k(W)$ denote the k-th largest eigenvalue of any symmetric matrix W, and observe that $\lambda_k(W) = s_k(W)$ if W is positive semi-definite. Combining Lemma S4 with part ii of Lemma S1 then gives the second lower bound:

$$\begin{aligned} s_{k+1}^2\left(A^d\right) = s_{k+1}\left((A^d)'A^d\right) \geq & s_{k+1+(d-1)d}\left((A'A)^d\right) \\ & = \left(s_{k+1+(d-1)d}\left(A'A\right)\right)^d = s_{k+1+(d-1)d}^{2d}. \end{aligned}$$

□

Proof of Lemma 1, Part i. The bounds follows from Lemma S5 for all integers d. Consider any non-integer d and thus $\lceil d \rceil - d > 0$. For the upper bound we only need to show $s_1^2(A^d) \leq MT^{2d}$ since $\sigma_k^2 \leq \lambda_{\max}(C)\lambda_{\max}((A^d)'A^d) \leq \lambda_{\max}((A^d)'A^d) = s_1^2(A^d)$. Observe that A^d is a triangular Toeplitz matrix with first row equals to $a = [a_0, a_1, \ldots, a_{T-1}]'$, where the coefficients a_ℓ are from (5). Using part ii of Lemma S1,

$$s_1(A^d) = s_1\left(\sum_{\ell=0}^{T-1} a_\ell L_T^\ell\right) \leq \sum_{\ell=0}^{T-1} a_\ell s_1(L_T^\ell) \leq \sum_{\ell=0}^{T-1} a_\ell.$$

But by Lemma S2,

$$\begin{aligned} \sum_{\ell=0}^{T-1} a_\ell = & (\Gamma(d))^{-1} \sum_{\ell=0}^{T-1} (\ell!)^{-1}\Gamma(d+\ell) \\ \leq & \sqrt{M}(\Gamma(d))^{-1}\left(1 + \sum_{\ell=1}^{T-1}\left(\frac{(\ell-1)!}{\ell!}\right)\ell^d\right) \\ = & \sqrt{M}(\Gamma(d))^{-2}\left(1 + \sum_{\ell=1}^{T-1}\ell^{d-1}\right) \leq MT^d. \end{aligned}$$

Next, we show the lower bound in part i of the lemma via

$$\sigma_k = s_k(A^dC) \geq s_{k+1}(A^d) \geq \frac{s_{k+1}(A^{\lceil d \rceil})}{s_1\left(A^{\lceil d \rceil - d}\right)}.$$

Using the upper bound above (replacing d with $\lceil d \rceil - d$), $s_1\left(A^{\lceil d \rceil - d}\right) \leq MT^{\lceil d \rceil - d}$. Hence, applying Lemma S5 to $A^{\lceil d \rceil}$ gives that

$$\sigma_k \geq \frac{s_{k+1+(\lceil d \rceil - 1)\lceil d \rceil}^{\lceil d \rceil}}{MT^{\lceil d \rceil - d}} = \frac{1}{M} s_{k+1+(\lceil d \rceil - 1)\lceil d \rceil}^d \left(\frac{s_{k+1+(\lceil d \rceil - 1)\lceil d \rceil}}{T}\right)^{\lceil d \rceil - d}.$$

This completes the proof for part i of the lemma.

□

Proof of Lemma 1, Part ii. Use part i of Lemma S1 for $d \geq 1$ to get that

$$\begin{aligned} \sigma_k^2 = & [s_k((I-L)^{-d+1}(I-L)^{-1})]^2 \\ \leq & [s_1((I-L)^{-d+1}) \cdot s_k((I-L)^{-1})]^2 \leq M^{1/2}T^{2(d-1)}s_k^2. \end{aligned}$$

But

$$s_k^2 = \left(2\sin\left(\frac{2k-1}{2(2T+1)\pi}\right)\right)^{-2} \leq M^{1/2}\left(\frac{k}{T}\right)^{-2},$$

and thus the stated result follows. □

C.2 Proof of Lemma 2

By the triangle inequality,

$$\|\Xi_t\| \leq \sum_{s=0}^{t-2} |a_s - a_{s+1}| \left\| \sum_{j=0}^{s} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\| + a_{t-1} \left\| \sum_{j=0}^{t-1} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|,$$

and then by the Cauchy-Schwarz inequality,

$$\|\Xi_t\|^2 \leq 2\sum_{s=0}^{t-2}(a_s - a_{s+1})^2 \sum_{s=0}^{t-2} \left\| \sum_{j=0}^{s} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2 + 2a_{t-1}^2 \left\| \sum_{j=0}^{t-1} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2.$$

Taking expectations on both sides,

$$\begin{aligned}
&\mathbb{E}\|\Xi_t\|^2 \\
\leq& 2\sum_{s=0}^{t-2}(a_s - a_{s+1})^2 \sum_{s=0}^{t-2} \mathbb{E}\left\| \sum_{j=0}^{s} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2 + 2a_{t-1}^2 \mathbb{E}\left\| \sum_{j=0}^{t-1} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2 \\
\leq& C\sum_{s=0}^{t-2}|s|_+^{2d-4} \sum_{s=0}^{t-2} \mathbb{E}\left\| \sum_{j=0}^{s} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2 + C|t-1|_+^{2d-2} \mathbb{E}\left\| \sum_{j=0}^{t-1} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2, \qquad (35)
\end{aligned}$$

where $|s|_+ = \max\{s, 1\}$.

For all $0 \leq s \leq t-1$, decompose that

$$\begin{aligned}
&\mathbb{E}\left\| \sum_{j=0}^{s} (u_{t-j} - \Psi\varepsilon_{t-j}) \right\|^2 \\
=& \mathbb{E}\left\| \sum_{j=0}^{s} u_{t-j} \right\|^2 + \mathbb{E}\left\| \Psi \sum_{j=0}^{s} \varepsilon_{t-j} \right\|^2 - 2\mathbb{E}\left[\left(\sum_{j=0}^{s} u_{t-j} \right)' \Psi \sum_{j=0}^{s} \varepsilon_{t-j} \right] \\
=:& T_{t,s,1} + T_{t,s,2} - 2T_{t,s,3}.
\end{aligned}$$

Define $\bar{\Psi}_\ell = \mathbb{E}[u_t' u_{t-\ell}] = \sum_{s=0}^{\infty} \operatorname{tr}(\Psi_s' \Psi_{s+\ell})$, and decompose further that

$$\begin{aligned}
T_{t,s,1} =& \sum_{\ell=-s}^{s} (s+1-|\ell|)\bar{\Psi}_\ell = (s+1)\sum_{\ell=-s}^{s} \bar{\Psi}_\ell - \sum_{\ell=-s}^{s} |\ell|\bar{\Psi}_\ell \\
T_{t,s,2} =& (s+1)\operatorname{tr}\Omega, \quad \operatorname{tr}\Omega = \sum_{j=-\infty}^{\infty} \bar{\Psi}_j = \sum_{s=0}^{\infty} \operatorname{tr}(\Psi_s' \Psi) \\
T_{t,s,3} =& \sum_{\ell=0}^{s} (s+1-\ell)\operatorname{tr}(\Psi_\ell' \Psi) = (s+1)\sum_{\ell=0}^{s} \operatorname{tr}(\Psi_\ell' \Psi) - \sum_{\ell=0}^{s} \ell \operatorname{tr}(\Psi_\ell' \Psi).
\end{aligned}$$

It follows that

$$
\begin{aligned}
&\mathbb{E}\left\|\sum_{j=0}^{s}(u_{t-j}-\Psi\varepsilon_{t-j})\right\|^2 \\
=&-(s+1)\sum_{|\ell|\geq s+1}\bar{\Psi}_\ell-2\sum_{\ell=1}^{s}\ell\bar{\Psi}_\ell+2(s+1)\sum_{\ell\geq s+1}\mathrm{tr}(\Psi_\ell'\Psi)+2\sum_{\ell=0}^{s}\ell\,\mathrm{tr}(\Psi_\ell'\Psi) \\
\leq&2(s+1)\sum_{\ell\geq s+1}\left[|\bar{\Psi}_\ell|+\|\Psi_\ell\|\,\|\Psi\|\right]+2\sum_{\ell=0}^{s}\ell\left[|\bar{\Psi}_\ell|+\|\Psi_\ell\|\,\|\Psi\|\right].
\end{aligned}
$$

But by the triangle inequality and the Cauchy-Schwarz inequality,

$$
\begin{aligned}
\sum_{\ell\geq s+1}|\bar{\Psi}_\ell|\leq\sum_{\ell\geq s+1}\sum_{s'=0}^{\infty}|\mathrm{tr}(\Psi_{s'}'\Psi_{s'+\ell})|\leq&\sum_{\ell\geq s+1}\sum_{s'=0}^{\infty}\|\Psi_{s'}\|\,\|\Psi_{s'+\ell}\| \\
\leq&\sum_{\ell=0}^{\infty}\|\Psi_\ell\|\sum_{\ell\geq s+1}\|\Psi_\ell\|\leq\|\Psi\|^2\sum_{\ell=0}^{\infty}\varphi_\ell\cdot\sum_{\ell\geq s+1}\varphi_\ell.
\end{aligned}
$$

In particular,

$$
\sum_{\ell\geq 0}|\bar{\Psi}_\ell|\leq\|\Psi\|^2\left(\sum_{\ell=0}^{\infty}\varphi_\ell\right)^2.
$$

Similarly,

$$
\begin{aligned}
\sum_{\ell=0}^{s}\ell|\bar{\Psi}_\ell|\leq&\sum_{\ell=0}^{s}\sum_{s'=0}^{\infty}\ell\,\|\Psi_{s'}\|\,\|\Psi_{s'+\ell}\| \\
\leq&\sum_{\ell=0}^{\infty}\|\Psi_\ell\|\sum_{\ell=0}^{s}\ell\,\|\Psi_\ell\|\leq\|\Psi\|^2\sum_{\ell=0}^{\infty}\varphi_\ell\cdot\sum_{\ell=0}^{s}\ell\varphi_\ell
\end{aligned}
$$

To summarize, for all $1\leq s\leq t-1$

$$
\mathbb{E}\left\|\sum_{j=0}^{s}(u_{t-j}-\Psi\varepsilon_{t-j})\right\|^2\leq 2\|\Psi\|^2\left(\sum_{\ell=0}^{\infty}\varphi_\ell+1\right)(s+1)\left\{\sum_{\ell\geq s+1}\varphi_\ell+\frac{1}{s+1}\sum_{\ell=0}^{s}\ell\varphi_\ell\right\}
$$

Plugging these bounds back to the upper bound (35) completes the proof.

C.3 Proof of Lemma 3

Let $\{\mathbf{e}_j\in\mathbb{R}^T:1\leq j\leq T\}$ denote the standard basis of $\mathbb{R}^T$ such that $\mathbf{e}_j$ has one in the j-th entry and zeros elsewhere. By the definition of eigenvector,

$$
\Phi\Phi' w_k=\sigma_k^2 w_k,\quad\text{or}\quad\mathbf{e}_j'\Phi\Phi' w_k=\sigma_k^2 w_{k,j}\ \forall 1\leq j\leq T.
$$

Hence,

$$
\sum_{j=1}^{T}w_{k,j}^4=\sigma_k^{-8}\sum_{j=1}^{T}(\mathbf{e}_j'\Phi\Phi' w_k)^4\leq\sigma_k^{-8}\sum_{j=1}^{T}(\mathbf{e}_j'\Phi\Phi'\Phi\Phi'\mathbf{e}_j)^2,
$$

where the last inequality is due to the Cauchy-Schwarz inequality and the construction that $w_k' w_k = 1$. Using the spectral inequality and the maximum inequality,

$$\sum_{j=1}^{T} w_{k,j}^4 \leq \sigma_k^{-8} \lambda_{\max}^2(\Phi'\Phi) \sum_{j=1}^{T} (\mathbf{e}_j' \Phi\Phi' \mathbf{e}_j) \cdot \max_{1\leq j\leq T} \mathbf{e}_j' \Phi\Phi' \mathbf{e}_j$$
$$= \sigma_k^{-8} \cdot \sigma_1^4 \cdot \mathrm{tr}(\Phi\Phi') \cdot \max_{1\leq j\leq T} \mathbf{e}_j' \Phi\Phi' \mathbf{e}_j = O(T^{-2d}) \cdot \max_{1\leq j\leq T} \mathbf{e}_j' \Phi\Phi' \mathbf{e}_j,$$

where the last step follows from part i of Lemma 1.

Finally, using Lemma S2, uniformly for $1 \leq j \leq T$

$$\mathbf{e}_j' \Phi\Phi' \mathbf{e}_j = \mathbf{e}_j' (I - L)^{-d} C (I - L')^{-d} \mathbf{e}_j$$
$$\leq \mathbf{e}_j' (I - L)^{-d} (I - L')^{-d} \mathbf{e}_j \leq \sum_{s=0}^{T-1} a_s^2 \leq M \sum_{s=1}^{T} s^{2d-2} = M T^{2d-1}.$$

The lemma follows by combining all bounds.

C.4 Proof of Corollary 5

For all k, k', $\{(\bar{\varepsilon}_{k,i}, \bar{\varepsilon}_{k',i}) : 1 \leq i \leq N\}$ are i.i.d. and the entries have zero mean and unit variance. Therefore, it suffices to check the results for each individual i. Part i follows from

$$\mathbb{E}\bar{\varepsilon}_{k,i}^4 - 3 = O\left(\sum_{j=1}^{T} w_{k,j}^4\right) = O(T^{-1}).$$

Part ii follows from

$$\mathbb{E}[\bar{\varepsilon}_{k,i}^2 \bar{\varepsilon}_{k',i}^2] = 1 + O\left(\sum_{j=1}^{T} w_{k,j}^2 w_{k',j}^2\right) = 1 + O(T^{-1}),$$

where the last step is due to the Cauchy-Schwarz inequality

$$\sum_{j=1}^{T} w_{k,j}^2 w_{k',j}^2 \leq \sqrt{\sum_{j=1}^{T} w_{k,j}^4} \cdot \sqrt{\sum_{j=1}^{T} w_{k',j}^4} = O(T^{-1}).$$

Part iii follows from

$$\mathbb{E}\left[\bar{\varepsilon}_{k,i}^2 \bar{\varepsilon}_{k',i}\right] = O\left(\sum_{j=1}^{T} w_{k,j}^2 w_{k',j}\right) = O\left(\sqrt{\sum_{j=1}^{T} w_{k,j}^4} \cdot \sqrt{\sum_{j=1}^{T} w_{k',j}^2}\right) = O(T^{-1/2}),$$

where the penultimate step is due to the Cauchy-Schwarz inequality.

C.5 Proof of Lemma 4

To establish (31), observe $\sqrt{\frac{2}{\nu}}\widehat{Z}_k + 1$ is a quadratic form given by

$$\sqrt{\frac{2}{\nu}}\widehat{Z}_k + 1 = \frac{1}{\mathrm{tr}\,\Omega} \bar{\varepsilon}_k' \underline{\Omega} \bar{\varepsilon}_k, \quad \bar{\varepsilon}_k = \varepsilon w_k$$

where $\bar{\varepsilon}_k$ has i.i.d. entries with zero mean, unit variance and excess kurtosis of $O(T^{-1})$ according to Corollary 5. A direct calculation yields that

$$\sqrt{\frac{2}{\nu}}\mathbb{E}\widehat{Z}_k + 1 = \frac{\operatorname{tr}\underline{\Omega}}{\operatorname{tr}\Omega} = w_i' w_i = 1, \text{ or } \mathbb{E}\widehat{Z}_k = 0,$$

and

$$\begin{aligned}\operatorname{var}(\widehat{Z}_k) =& \frac{\nu}{2}\operatorname{var}\left(\sqrt{\frac{2}{\nu}}\widehat{Z}_k + 1\right)\\ =& \frac{(\operatorname{tr}\Omega)^2}{2\operatorname{tr}\Omega^2}\frac{1}{(\operatorname{tr}\Omega)^2}\left\{O(T^{-1})\cdot\operatorname{tr}(\operatorname{diag}(\underline{\Omega})^2) + 2\operatorname{tr}\underline{\Omega}^2\right\} = O(T^{-1}) + 1.\end{aligned}$$

Recall the eigenvalues $\widetilde{\sigma}_k^2$ of the approximate covariance matrix $\widetilde{\Sigma} = \Phi'\varepsilon'\underline{\Omega}\varepsilon\Phi/T$ of the leading term in (6) that are used in the proof of Theorem 1. We first show that the stated results holds for $\widetilde{\sigma}_k^2$ before extending them to the perturbed statistics $\widehat{\lambda}_k$. Consider the singular value decomposition of Φ given by

$$\Phi = WDV' = W_1 D_1 V_1' + W_2 D_2 V_2',$$

where $D = \operatorname{diag}(\sigma_1, \ldots, \sigma_T)$, $D_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_{K+1})$, $D_2 = \operatorname{diag}(\sigma_{K+2}, \ldots, \sigma_T)$, $W = [W_1, W_2]$, and $V = [V_1, V_2]$. It suffices to consider the eigenvalues of

$$\begin{aligned}V'\widetilde{\Sigma}V =& \frac{1}{T}\begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}\begin{bmatrix} W_1' \\ W_2' \end{bmatrix}\varepsilon'\underline{\Omega}\varepsilon[W_1, W_2]\begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}\\ =& \frac{1}{T}\begin{bmatrix} D_1 W_1'\varepsilon'\underline{\Omega}\varepsilon W_1 D_1 & D_1 W_1'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2 \\ D_2 W_2'\varepsilon'\underline{\Omega}\varepsilon W_1 D_1 & D_2 W_2'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2 \end{bmatrix} =: \begin{bmatrix} Q_{1,1} & Q_{1,2} \\ Q_{2,1} & Q_{2,2} \end{bmatrix}.\end{aligned}$$

Apply the consistency result (29) of the extreme eigenvalues from the proof of Theorem 1 to the submatrix $Q_{2,2}$ yields that

$$\lambda_{\max}(Q_{2,2}) = \frac{\operatorname{tr}\Omega}{T}\sigma_{K+2}^2(1 + o_{\mathbb{P}}(1)) = \exp(\mu)\sigma_{K+2}^2(1 + o_{\mathbb{P}}(1)), \quad \mu = \log\left(\frac{\operatorname{tr}\Omega}{T}\right).$$

Let $\|\cdot\|_{\mathrm{sp}}$ denote the spectral norm. Combining with the separability Assumption 2.4, for all $\lambda \geq \exp(\mu)\sigma_{K+1}\sigma_{K+2}$

$$\begin{aligned}\left\|(\lambda I_{T-K-1} - Q_{2,2})^{-1}\right\|_{\mathrm{sp}} =& O_{\mathbb{P}}((\exp(\mu)\sigma_{K+1}\sigma_{K+2} - \exp(\mu)\sigma_{K+2}^2)^{-1})\\ =& O_{\mathbb{P}}\left(\sigma_1^{-2}\exp(-\mu)\right),\end{aligned} \tag{36}$$

where the last equality is due to (27). With the consistency result (29) in the proof of Theorem 1, the $K+1$ largest eigenvalues $\widetilde{\sigma}_1^2 \geq \ldots \geq \widetilde{\sigma}_{k+1}^2$ are therefore the solutions of the following characteristic equation with probability approaching 1:

$$\det\left(\lambda I_{K+1} - Q_{1,1} - Q_{1,2}(\lambda I_{T-K-1} - Q_{2,2})^{-1}Q_{2,1}\right) = 0, \quad \frac{\lambda}{\exp(\mu)} \in [\sigma_{K+1}\sigma_{K+2}, 2\sigma_1^2]. \tag{37}$$

We shall show that the reminder $Q_{1,2}(\lambda I_{T-K-1} - Q_{2,2})^{-1}Q_{2,1}$ is asymptotic negligible. From (36) we already know that uniformly for $\lambda \geq \exp(\mu)\sigma_{K+1}\sigma_{K+2}$

$$
\begin{aligned}
&\left\|Q_{1,2}(\lambda I_{T-K-1} - Q_{2,2})^{-1}Q_{2,1}\right\|_{\mathrm{sp}} \\
\leq &\left\|(\lambda I_{T-K-1} - Q_{2,2})^{-1}\right\|_{\mathrm{sp}} \cdot \left\|Q_{1,2}Q_{2,1}\right\|_{\mathrm{sp}} \\
= &O_{\mathbb{P}}\left(\sigma_1^{-2}\exp(-\mu)\right) \cdot T^{-2}\left\|D_1 W_1'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2^2 W_2'\varepsilon'\underline{\Omega}\varepsilon W_1 D_1\right\|_{\mathrm{sp}}.
\end{aligned}
$$

Since D_1 is a diagonal matrix with positive diagonal entries bounded by σ_1,

$$
\left\|Q_{1,2}(\lambda I_{T-K-1} - Q_{2,2})^{-1}Q_{2,1}\right\|_{\mathrm{sp}} = O_{\mathbb{P}}\left(\exp(-\mu)T^{-2}\right) \cdot \left\|W_1'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2^2 W_2'\varepsilon'\underline{\Omega}\varepsilon W_1\right\|_{\mathrm{sp}}. \tag{38}
$$

Note that the matrix $W_1'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2^2 W_2'\varepsilon'\underline{\Omega}\varepsilon W_1$ is a $(K+1)\times(K+1)$ positive semidefinite matrix of finite dimension. We only need to control the stochastic order of all the diagonals (hence the trace and spectral norm), namely, $w_k'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2^2 W_2'\varepsilon'\underline{\Omega}\varepsilon w_k$ with the eigenvectors w_k of $\Phi\Phi'$ defined in Lemma 3. Since every diagonal element is non-negative, it suffices to control its expected value via

$$
\mathbb{E}\left[w_k'\varepsilon'\underline{\Omega}\varepsilon W_2 D_2^2 W_2'\varepsilon'\underline{\Omega}\varepsilon w_k\right] = \sum_{j=K+2}^{T} \sigma_j^2 \mathbb{E}\left(w_j'\varepsilon'\underline{\Omega}\varepsilon w_k\right)^2 = \sum_{j=K+2}^{T} \sigma_j^2 \operatorname{var}\left(w_j'\varepsilon'\underline{\Omega}\varepsilon w_k\right), \tag{39}
$$

where the last step is due to the orthogonality between w_j and w_k, $j \neq k$, such that

$$
\mathbb{E}\left[w_j\varepsilon'\underline{\Omega}\varepsilon w_k\right] = \operatorname{tr}\underline{\Omega} \cdot w_j' w_k = 0.
$$

Let $\{\mathbf{e}_j \in \mathbb{R}^T : 1 \leq j \leq T\}$ denote the standard basis of $\mathbb{R}^T$ such that $\mathbf{e}_j$ has one in the j-th entry and zero elsewhere. Denote by $\underline{\Omega}_{s,r}$ the (s,r)-entry of $\underline{\Omega}$. Decompose that

$$
\begin{aligned}
\operatorname{var}\left(w_j'\varepsilon'\underline{\Omega}\varepsilon w_k\right) &= \operatorname{var}\left(\sum_{s,r=1}^{N} \underline{\Omega}_{s,r} w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r' w_i\right) \\
&= \operatorname{var}\left(\sum_{s=1}^{N} \underline{\Omega}_{s,s} w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_s' w_k + \sum_{s\neq r} \underline{\Omega}_{s,r} w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r' w_k\right) \\
&= \sum_{s=1}^{N} \underline{\Omega}_{s,s}^2 \operatorname{var}(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_s'\varepsilon w_k) \\
&\quad + \sum_{s\neq r} \underline{\Omega}_{s,r}^2 \left\{\operatorname{var}(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r'\varepsilon w_k) + \operatorname{cov}\left(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r'\varepsilon w_i, w_j'\varepsilon'\mathbf{e}_r\mathbf{e}_s'\varepsilon w_k\right)\right\}.
\end{aligned}
$$

We shall show that all the (co)variance terms are bounded. First,

$$
\begin{aligned}
&\operatorname{var}(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_s'\varepsilon w_k)\\
=&\mathbb{E}\left[(w_j'\varepsilon'\mathbf{e}_s)^2(\mathbf{e}_s'\varepsilon w_k)^2\right]\\
=&\mathbb{E}\varepsilon_{s,l}^4\cdot\sum_{l=1}^N w_{j,l}^2w_{k,l}^2+\sum_{l_1\neq l_2}(w_{j,l_1}^2w_{k,l_2}^2+w_{j,l_1}w_{j,l_2}w_{k,l_1}w_{k,l_2}+w_{j,l_1}w_{j,l_2}w_{k,l_2}w_{k,l_1})\\
=&O(1)\cdot\left\{\left(\sum_{l=1}^N w_{j,l}^2\right)\left(\sum_{l=1}^N w_{k,l}^2\right)+\left(\sum_{l=1}^N w_{j,l}w_{k,l}\right)^2\right\}\\
=&O\left(\left(\sum_{l=1}^N w_{j,l}^2\right)\left(\sum_{l=1}^N w_{k,l}^2\right)\right)=O(1),
\end{aligned}
$$

where the last line is due to the Cauchy–Schwarz inequality. Second, for $s\neq r$

$$
\operatorname{var}(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r'\varepsilon w_k)=\mathbb{E}(w_j'\varepsilon'\mathbf{e}_s)^2\mathbb{E}(\mathbf{e}_r'\varepsilon w_k)^2=w_j'w_j\cdot w_k'w_k=1,
$$

and thus, by Cauchy–Schwarz inequality,

$$
\left|\operatorname{cov}\left(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r'\varepsilon w_k, w_j'\varepsilon'\mathbf{e}_r\mathbf{e}_s'\varepsilon w_k\right)\right|\leq\sqrt{\operatorname{var}\left(w_j'\varepsilon'\mathbf{e}_s\mathbf{e}_r'\varepsilon w_k\right)\cdot\operatorname{var}\left(w_j'\varepsilon'\mathbf{e}_r\mathbf{e}_s'\varepsilon w_k\right)}=1.
$$

Then we can deduce that

$$
\operatorname{var}\left(w_j\varepsilon'\underline{\Omega}\varepsilon w_k\right)=O(1)\cdot\sum_{s,r=1}^N\underline{\Omega}_{s,r}^2=O(1)\cdot\operatorname{tr}\underline{\Omega}^2=O(1)\cdot\operatorname{tr}\Omega^2. \tag{40}
$$

Substituting into (39) and noting that the $O(1)$ term is uniform,

$$
\mathbb{E}\left[w_k'\varepsilon'\underline{\Omega}\varepsilon W_2D_2^2W_2'\varepsilon'\underline{\Omega}\varepsilon w_k\right]=O(\operatorname{tr}\Omega^2)\cdot\sum_{j=K+2}^T\sigma_j^2=O\left(\operatorname{tr}\Omega^2\cdot\operatorname{tr}D_2^2\right).
$$

Hence, using the Markov inequality, for all fixed K

$$
\begin{aligned}
\left\|W_1'\varepsilon'\underline{\Omega}\varepsilon W_2D_2^2W_2'\varepsilon'\underline{\Omega}\varepsilon W_1\right\|_{\mathrm{sp}}&\leq\operatorname{tr}\left(W_1'\varepsilon'\underline{\Omega}\varepsilon W_2D_2^2W_2'\varepsilon'\underline{\Omega}\varepsilon W_1\right)\\
&=\sum_{k=1}^{K+1}w_k'\varepsilon'\underline{\Omega}\varepsilon W_2D_2^2W_2'\varepsilon'\underline{\Omega}\varepsilon w_k=O_{\mathbb{P}}\left(\operatorname{tr}\Omega^2\cdot\operatorname{tr}D_2^2\right).
\end{aligned}
$$

Combining with (38), uniformly for $\lambda\geq\exp(\mu)\sigma_{K+1}\sigma_{K+2}$

$$
\left\|Q_{1,2}(\lambda I_{T-K-1}-Q_{2,2})^{-1}Q_{2,1}\right\|_{\mathrm{sp}}=O_{\mathbb{P}}\left(\exp(-\mu)T^{-2}\right)\cdot O(\operatorname{tr}\Omega^2\cdot\operatorname{tr}D_2^2)
$$

Observe that $\exp(-\mu)T^{-2}\operatorname{tr}\Omega^2=\nu^{-1}\exp(\mu)$ and recall from (27) that $\operatorname{tr}D_2^2=O(\sigma_1^2)$. Then,

$$
\left\|Q_{1,2}(\lambda I_{T-K-1}-Q_{2,2})^{-1}Q_{2,1}\right\|_{\mathrm{sp}}=O_{\mathbb{P}}\left(\nu^{-1}\exp(\mu)\cdot\sigma_1^2\right). \tag{41}
$$

Denote the matrix function of λ in the characteristic equation (37) by $G(\lambda)=\lambda I_{K+1}-Q_{1,1}-Q_{1,2}(\lambda I_{T-K-1}-Q_{2,2})^{-1}Q_{2,1}$ with the (i,j)th element $G_{i,j}(\lambda)$. Let $Q_{1,1}(i,j)=\sigma_i\sigma_jw_i'\varepsilon'\underline{\Omega}\varepsilon w_j$ denotes the (i,j)th element of $Q_{1,1}$. We can approximate the off-diagonal elements in $G(\lambda)$ via (41) as follows: uniformly for $\lambda \geq \exp(\mu)\sigma_{K+1}\sigma_{K+2}$ and $i \neq j$

$$\begin{aligned} G_{i,j}(\lambda) =& 0 - Q_{1,1}(i,j) + O_{\mathbb{P}}\left(\nu^{-1}\exp(\mu)\sigma_1^2\right) \\ =& 0 - O_{\mathbb{P}}\left(T^{-1}\sqrt{\operatorname{tr}\Omega^2}\sigma_i\sigma_j\right) + O_{\mathbb{P}}\left(\nu^{-1}\exp(\mu)\sigma_1^2\right) = O_{\mathbb{P}}\left(\nu^{-1/2}\exp(\mu)\sigma_1^2\right) \end{aligned}$$

where the second step is due to (40) and the construction that

$$\mathbb{E}Q_{1,1}(i,j) = \frac{1}{T}\mathbb{E}\left[\sigma_i\sigma_j w_j'\varepsilon'\underline{\Omega}\varepsilon w_i\right] = \sigma_i\sigma_j w_i' w_j \operatorname{tr}\underline{\Omega} = 0, \quad i \neq j.$$

Similarly, for all i, $\mathbb{E}[Q_{1,1}(i,i)] = \exp(\mu)\sigma_i^2$ and the diagonal elements

$$G_{i,i}(\lambda) = \lambda - Q_{1,1}(i,i) + O_{\mathbb{P}}\left(\nu^{-1}\exp(\mu)\sigma_1^2\right) = \lambda - \exp(\mu)\sigma_i^2 + O_{\mathbb{P}}\left(\nu^{-1/2}\exp(\mu)\sigma_1^2\right). \tag{42}$$

Denote by $\mathcal{P}_{K+1}$ the collection of all the permutation functions τ on the set $\{1, 2, \cdots, K+1\}$, and let $\tau_0 \in \Theta_{K+1}$ denotes the identity function $\tau_0(i) = i$. By (37) and the Laplace formula for a determinant,

$$\begin{aligned} 0 = \det(G(\lambda)) =& \sum_{\tau \in \mathcal{P}_{K+1}} \operatorname{sgn}(\tau) \prod_{i=1}^{K+1} G_{i,\tau(i)}(\lambda) \\ =& \prod_{i=1}^{K+1} G_{i,i}(\lambda) + \sum_{\tau \in \mathcal{P}_{K+1}, \tau \neq \tau_0} \operatorname{sgn}(\tau) \prod_{i=1}^{K+1} G_{i,\tau(i)}(\lambda), \end{aligned}$$

where $\operatorname{sgn}(\tau) \in \{-1, +1\}$ denotes the sign of permutation τ.

We first consider the reminder term. Uniformly for all $\lambda \leq 2\exp(\mu)\sigma_1^2$, all diagonal elements of $G(\lambda)$ is $O_{\mathbb{P}}(\exp(\mu)\sigma_1^2)$ but all off-diagonal elements of $G(\lambda)$ are $O_{\mathbb{P}}\left(\nu^{-1/2}\exp(\mu)\sigma_1^2\right)$ as known above. For all $\tau \neq \tau_0$, there are at least two different values of i such that $\tau(i) \neq i$, and therefore

$$\exp(-(K+1)\mu)\sigma_1^{-2(K+1)} \prod_{i=1}^{K+1} G_{i,\tau(i)}(\lambda) = O_{\mathbb{P}}\left(\nu^{-\frac{1}{2}\cdot 2}\right).$$

Since K is bounded, summing up the stochastic bounds does not change the order:

$$\exp(-(K+1)\mu)\sigma_1^{-2(K+1)} \sum_{\tau \in \mathcal{P}_{K+1}, \tau \neq \tau_0} \operatorname{sgn}(\tau) \prod_{i=1}^{K+1} G_{i,\tau(i)}(\lambda) = O_{\mathbb{P}}(\nu^{-1}). \tag{43}$$

Now consider the leading term $\prod_{i=1}^{K+1} G_{i,i}(\lambda)$. For any given λ and $i \neq j$, the approximation (42) gives that

$$G_{j,j}(\lambda) - G_{i,i}(\lambda) = \exp(\mu)(\sigma_j^2 - \sigma_i^2) + O_{\mathbb{P}}(\nu^{-1/2}\exp(\mu)\sigma_1^2),$$

where the first term is of order $\exp(\mu)\sigma_1^2$ and second term is negligible as, under Assumption 2.5

$$\nu^{-1} = \frac{\operatorname{tr}\Omega^2}{(\operatorname{tr}\Omega)^2} \leq \frac{\lambda_{\max}(\Omega)}{\operatorname{tr}\Omega} \to 0.$$

On the other hand, using the separation Assumption 2.4 and the fact σ_j^2 and σ_i^2 all diverge at the same order of σ_1^2 from (27), there exists some $M > 0$ such that uniformly for λ

$$\sum_{1\le i\neq j\le K+1} \mathbb{P}\left(|G_{i,i}(\lambda) - G_{j,j}(\lambda)| < M^{-1}\exp(\mu)\sigma_1^2\right) \to 0.$$

Hence, with probability tending to 1, $G_{i,i}(\lambda)$ and $G_{j,j}(\lambda)$ maintain a distance of the exact order $\exp(\mu)\sigma_1^2$ for all $i \neq j$. Thus, (43) implies that when $\det(G(\lambda)) = 0$ there exists $1 \le k \le K+1$ such that

$$\sigma_1^{-2}\exp(-\mu)G_{k,k}(\lambda) = O_{\mathbb{P}}\Big(\nu^{-1}\Big) \quad \text{or} \quad \sigma_k^{-2}\exp(-\mu)G_{k,k}(\lambda) = O_{\mathbb{P}}\Big(\nu^{-1}\Big).$$

Then invoking (42) gives that

$$\sigma_k^{-2}\exp(-\mu)\lambda = \frac{1}{\operatorname{tr}\Omega} w_k'\varepsilon'\underline{\Omega}\varepsilon w_k + O_{\mathbb{P}}\left(\nu^{-1}\right).$$

Since there are $K+1$ different solutions for $\det(G(\lambda)) = 0$ satisfying (29),

$$\sigma_k^{-2}\exp(-\mu)\widetilde{\sigma}_k^2 = \frac{1}{\operatorname{tr}\Omega} w_k'\varepsilon'\underline{\Omega}\varepsilon w_k + O_{\mathbb{P}}\left(\nu^{-1}\right), \quad k = 1,\ldots,K+1.$$

To substitute $\widetilde{\sigma}_k^2$ with $\widehat{\lambda}_k$, we need the reminder in (6) to be negligible. Recall that $\|\cdot\|_{\text{sp}}$ and $\|\cdot\|$ denote the spectral and Frobenius norms respectively. It suffices to show that

$$\|\Xi C\|_{\text{sp}}^2 \le \|\Xi\|_{\text{sp}}^2 \le \|\Xi\|^2 = o_{\mathbb{P}}\left(\operatorname{tr}\Omega\sigma_{K+1}^2\nu^{-1/2}\right), \tag{44}$$

The last step follows from the Markov inequality, because combining Lemma 2 with the decay rate of $\bar{\Lambda}_t$ in (30) yields that

$$\mathbb{E}\|\Xi\|^2 = \sum_{t=1}^{T}\mathbb{E}\|\Xi_t\|^2 = o(\operatorname{tr}\Omega\cdot T^{2d-\iota/2}) = o(\operatorname{tr}\Omega\sigma_{K+1}^2\nu^{-1/2}).$$

with $\nu = O(N^\iota) = O(T^\iota)$ according to Assumption 2.1.

C.6 Proof of Lemma 5

Recall the matrix $\underline{\Omega} = \Psi'\Psi$ and denote its (i,j)th entry by $\underline{\Omega}_{i,j}$. Our proofs rely on the following decomposition.

$$\begin{aligned}
&D_{k,i}D_{k',i}\\
=&\frac{1}{2\operatorname{tr}\Omega^2}\left\{\underline{\Omega}_{i,i}(\bar{\varepsilon}_{k,i}^2 - 1) + 2\bar{\varepsilon}_{k,i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s}\right\}\left\{\underline{\Omega}_{i,i}(\bar{\varepsilon}_{k',i}^2 - 1) + 2\bar{\varepsilon}_{k',i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k',s}\right\}\\
=&\frac{1}{2\operatorname{tr}\Omega^2}\underline{\Omega}_{i,i}^2(\bar{\varepsilon}_{k,i}^2 - 1)(\bar{\varepsilon}_{k',i}^2 - 1) + \frac{1}{\operatorname{tr}\Omega^2}\underline{\Omega}_{i,i}(\bar{\varepsilon}_{k,i}^2 - 1)\bar{\varepsilon}_{k',i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k',s}\\
&+\frac{1}{\operatorname{tr}\Omega^2}\underline{\Omega}_{i,i}(\bar{\varepsilon}_{k',i}^2 - 1)\bar{\varepsilon}_{k,i}\cdot\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s} + \frac{2}{\operatorname{tr}\Omega^2}\bar{\varepsilon}_{k,i}\bar{\varepsilon}_{k',i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k',s}
\end{aligned}$$

Proof of Part i. Since $\bar{\varepsilon}_{k,i}$ and $\bar{\varepsilon}_{k',i}$ are independent of $\mathcal{F}_{p,i-1}$ and using the orthogonality $\mathbb{E}[\bar{\varepsilon}_{k,i}\bar{\varepsilon}_{k',i}] = w_k' w_{k'} = 0$ for $k \neq k'$,

$$\begin{aligned}&\mathbb{E}\left[D_{k,i}D_{k',i} \mid \mathcal{F}_{p,i-1}\right]\\=&\frac{\underline{\Omega}_{i,i}^2}{2\operatorname{tr}\Omega^2}\mathbb{E}\left[\bar{\varepsilon}_{k,i}^2\bar{\varepsilon}_{k',i}^2-1\right]+\frac{\underline{\Omega}_{i,i}}{\operatorname{tr}\Omega^2}\mathbb{E}\left[\bar{\varepsilon}_{k,i}^2\bar{\varepsilon}_{k',i}\right]\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k',s}+\frac{\underline{\Omega}_{i,i}}{\operatorname{tr}\Omega^2}\mathbb{E}\left[\bar{\varepsilon}_{k',i}^2\bar{\varepsilon}_{k,i}\right]\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s}\end{aligned}$$

Using parts ii and iii of Corollary 5,

$$\begin{aligned}\sum_{i=1}^N\mathbb{E}\left[D_{k,i}D_{k',i}' \mid \mathcal{F}_{p,i-1}\right]=&\frac{\sum_{i=1}^N\underline{\Omega}_{i,i}^2}{\operatorname{tr}\Omega^2}\cdot O(T^{-1})+O(T^{-1/2})\cdot\frac{1}{\operatorname{tr}\Omega^2}\sum_{i=1}^N\underline{\Omega}_{i,i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k',s}\\&+O(T^{-1/2})\cdot\frac{1}{\operatorname{tr}\Omega^2}\sum_{i=1}^N\underline{\Omega}_{i,i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s}.\end{aligned}$$

Exchanging the order of summations,

$$\mathbb{E}\left(\sum_{i=1}^N\underline{\Omega}_{i,i}\sum_{s=1}^{i-1}\Omega_{k,s}\bar{\varepsilon}_{k',s}\right)^2=\mathbb{E}\left(\sum_{s=1}^{N-1}\bar{\varepsilon}_{k',s}\sum_{i=s+1}^N\underline{\Omega}_{i,i}\underline{\Omega}_{i,s}\right)^2=\sum_{s=1}^{N-1}\left(\sum_{i=s+1}^N\underline{\Omega}_{i,i}\underline{\Omega}_{i,s}\right)^2$$

By the Cauchy-Schwarz inequality, the last term is bounded by

$$\sum_{s=1}^{N-1}\sum_{i=s+1}^N\underline{\Omega}_{i,i}^2\sum_{i=s+1}^N\underline{\Omega}_{i,s}^2\le\sum_{i=1}^N\underline{\Omega}_{i,i}^2\cdot\operatorname{tr}\Omega^2.$$

Therefore,

$$\sum_{i=1}^N\underline{\Omega}_{i,i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k',s}=O_{\mathbb{P}}\left(\sqrt{\sum_{i=1}^N\underline{\Omega}_{i,i}^2}\sqrt{\operatorname{tr}\Omega^2}\right). \tag{45}$$

Similarly, replacing the index k' with k,

$$\sum_{i=1}^N\underline{\Omega}_{i,i}\sum_{s=1}^{i-1}\underline{\Omega}_{i,s}\bar{\varepsilon}_{k,s}=O_{\mathbb{P}}\left(\sqrt{\sum_{i=1}^N\underline{\Omega}_{i,i}^2}\sqrt{\operatorname{tr}\Omega^2}\right).$$

Combining all bounds yields that

$$\begin{aligned}\sum_{i=1}^N\mathbb{E}\left[D_{k,i}D_{k',i} \mid \mathcal{F}_{p,i-1}\right]=&O(T^{-1})+O_{\mathbb{P}}\left(T^{-1/2}\sqrt{\sum_{i=1}^N\underline{\Omega}_{i,i}^2/\operatorname{tr}\Omega^2}\right)\\=&O(T^{-1})+O_{\mathbb{P}}(T^{-1/2}).\end{aligned}$$

□

Proof of Part ii. See the proof of Theorem 2.1, equation (3.4), in Bhansali, Giraitis, and Kokoszka (2007). We do not repeat the details. Note that the limiting variance 1 coincides with the limiting variance of $\widehat{Z}_k$ in (31). □

Proof of Part iii. Invoking the expansion at the beginning with $k' = k$,

$$D_{k,i}^2 = \frac{1}{2 \operatorname{tr} \Omega^2} \underline{\Omega}_{i,i}^2 (\bar{\varepsilon}_{k,i}^2 - 1)^2 + \frac{2}{\operatorname{tr} \Omega^2} \underline{\Omega}_{i,i} (\bar{\varepsilon}_{k,i}^2 - 1) \bar{\varepsilon}_{k,i} \cdot \sum_{s=1}^{i-1} \underline{\Omega}_{i,s} \bar{\varepsilon}_{k,s}$$
$$+ \frac{2}{\operatorname{tr} \Omega^2} \bar{\varepsilon}_{k,i}^2 \cdot \left(\sum_{s=1}^{i-1} \underline{\Omega}_{i,s} \bar{\varepsilon}_{k,s} \right)^2 .$$

Using parts i and iii of Corollary 5,

$$\mathbb{E}\left[D_{k,i}^2 \mid \mathcal{F}_{p,i-1} \right] = \frac{\underline{\Omega}_{i,i}^2}{\operatorname{tr} \Omega^2} \cdot (1 + T^{-1} \cdot O(1)) + T^{-1/2} \cdot O(1) \cdot \frac{\underline{\Omega}_{i,i}}{\operatorname{tr} \Omega^2} \sum_{s=1}^{i-1} \underline{\Omega}_{i,s} \bar{\varepsilon}_{k,s}$$
$$+ \frac{2}{\operatorname{tr} \Omega^2} \cdot \left(\sum_{s=1}^{i-1} \underline{\Omega}_{i,s} \bar{\varepsilon}_{k,s} \right)^2$$

where the $O(1)$ terms are uniform. Now use Assumption 2.6 to get that

$$\max_{1 \le i \le N} \frac{\underline{\Omega}_{i,i}^2}{\operatorname{tr} \Omega^2} \le \frac{\lambda_{\max}^2(\underline{\Omega})}{\operatorname{tr} \Omega^2} \to 0.$$

Moreover, following the proof of Lemma 4 in He et al. (2023),

$$\max_{1 \le i \le N} \frac{2}{\operatorname{tr} \Omega^2} \cdot \left(\sum_{s=1}^{i-1} \underline{\Omega}_{i,s} \bar{\varepsilon}_{k,s} \right)^2 \xrightarrow{\mathbb{P}} 0.$$

Then the stated result follows. □

D Proofs of Lemmas 7 and 8

D.1 Proof of Lemma 7

Let $\Gamma' = \sum_{\ell=0}^{\infty} \psi_\ell L_T^\ell$ such that $\Psi' = \Gamma' C$ and $\underline{\Omega} = \Psi' \Psi = \Gamma' C \Gamma$.

Proof of Part i. The upper bound of $\operatorname{tr} \underline{\Omega}$ is straightforward through

$$\operatorname{tr} \underline{\Omega} \le \|\Gamma\|^2 \le \sum_{\ell=0}^{T-1} (T - \ell) \psi_\ell^2 \le T \sum_{\ell=0}^{T-1} \psi_\ell^2$$

To find its lower bound, first we introduce the the T dimensional all-ones vector $\mathbf{1}_T$. By the Cauchy-Schwarz inequality

$$\|\Gamma' \mathbf{1}_T\|^2 = \sum_{t=1}^{T} \left(\sum_{\ell=0}^{T-t} \psi_\ell \right)^2 \le \sum_{t=1}^{T} \sum_{\ell=0}^{T-t} (T - t + 1) \psi_\ell^2$$
$$= \sum_{\ell=0}^{T-1} \sum_{t=1}^{T-\ell} (T - t + 1) \psi_\ell^2 = \sum_{\ell=0}^{T-1} \frac{(T - \ell)(T + \ell + 1)}{2} \psi_\ell^2.$$

It follows that

$$\begin{aligned}
\operatorname{tr} \underline{\Omega} &= \|\Gamma\|^2 - \frac{1}{T}\|\Gamma \mathbf{1}_T\|^2 \\
&\geq \sum_{\ell=0}^{T-1}(T-\ell)\psi_\ell^2 - \sum_{\ell=0}^{T-1}\frac{(T-\ell)(T+\ell+1)}{2T}\psi_\ell^2 \\
&= \sum_{\ell=0}^{T-1}(T-\ell)\psi_\ell^2 - \sum_{\ell=0}^{T-1}\frac{(T-\ell)(T+\ell+1)}{2T}\psi_\ell^2 = \frac{1}{2}\sum_{\ell=0}^{T-1}\frac{(T-\ell)(T-\ell-1)}{T}\psi_\ell^2.
\end{aligned}$$

We shall show that the lower bound is close to $\frac{1}{2}T\sum_{\ell=0}^{T-1}\psi_\ell^2$ in large samples, the hence strictly bounded away from $\frac{1}{4}T\sum_{\ell=0}^{T-1}\psi_\ell^2$. Let $\delta > 0$ be arbitrary. There exists a large $K = K(\delta)$ such that for all large T

$$\sum_{\ell=0}^{K}\frac{(T-\ell)(T-\ell-1)}{T}\psi_\ell^2 \geq (1-\delta)T\sum_{\ell=0}^{K}\psi_\ell^2 \geq (1-\delta)^2 T\sum_{\ell=0}^{T-1}\psi_\ell^2.$$

The last step is due to the summability of φ_ℓ, which implies that for all large K

$$\frac{\sum_{\ell=0}^{K}\psi_\ell^2}{\sum_{\ell=0}^{T-1}\psi_\ell^2} = 1 - \frac{\sum_{\ell=K+1}^{T-1}\psi_\ell^2}{\sum_{\ell=0}^{T-1}\psi_\ell^2} \geq 1 - \sum_{\ell=K+1}^{T-1}\varphi_\ell \geq 1-\delta.$$

Combining all these bounds yields that

$$\operatorname{tr} \underline{\Omega} \geq \frac{1}{2}\sum_{\ell=0}^{K}\frac{(T-\ell)(T-\ell-1)}{T}\psi_\ell^2 \geq \frac{1}{2}(1-\delta)^2 T\sum_{\ell=0}^{T-1}\psi_\ell^2.$$

Taking δ small enough gives the desired lower bound. □

Proof of Part ii. Recall the upper shift matrix L_T and $\underline{\Omega} = \Psi'\Psi = \Gamma' C \Gamma$. Let $\|\cdot\|_{\text{sp}}$ denotes the spectral norm. By the triangle inequality,

$$\lambda_{\max}(\underline{\Omega}) \leq \|\Psi\|_{\text{sp}}^2 \leq \|\Gamma\|_{\text{sp}}^2 \leq \left(\sum_{\ell=0}^{T-1}|\psi_\ell|\left\|L_T^\ell\right\|_{\text{sp}}\right)^2 \leq \left(\sum_{\ell=0}^{T-1}|\psi_\ell|\right)^2.$$

Using part 2 of Assumption 3.1,

$$\lambda_{\max}(\underline{\Omega}) \leq \left(\sum_{\ell=0}^{T-1}\sqrt{\varphi_\ell}\right)^2\left(\sum_{\ell=0}^{T-1}\psi_\ell^2\right)$$

where the last step used the lower bound of $\operatorname{tr}\underline{\Omega}$ from above. □

Proof of Part iii. By the Cauchy-Schwarz inequality,

$$T\left\|\underline{\Omega}_{(T)}\right\|^2 \geq \left(\operatorname{tr}\underline{\Omega}_{(T)}\right)^2 \geq \frac{1}{16}T^2,$$

where the last lower bound comes from part i. □

D.2 Proof of Lemma 8

Denote the floor function by $\lfloor\cdot\rfloor$. By the Hölder inequality, for all small $\delta > 0$

$$\sum_{\ell=0}^{\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor -1} \sqrt{\varphi_\ell} \le \delta T^{\frac{2\iota-1}{2\iota}} \left(\sum_{\ell=0}^{\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor -1} \varphi_\ell^\iota \right)^{1/(2\iota)} \le \delta T^{\frac{2\iota-1}{2\iota}} \left(\sum_{\ell=0}^{\infty} \varphi_\ell^\iota \right)^{1/\iota}$$

where the last bound can be an arbitrarily small proportion of $T^{\frac{2\iota-1}{2\iota}}$. Furthermore, by the Hölder inequality again, for any fixed $\delta \in (0,1)$

$$\begin{aligned} T^{-1} \sum_{\ell=\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor}^{T-1} \sqrt{\varphi_\ell} &\le \sum_{\ell=\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor}^{T-1} \ell^{-1} \sqrt{\varphi_\ell} \\ &\le \sqrt{\sum_{\ell=\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor}^{T-1} \ell^{-\frac{2\iota}{2\iota-1}}} \cdot \left(\sum_{\ell=\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor}^{T-1} \varphi_\ell^\iota \right)^{\frac{1}{2\iota}} = O\left(T^{-\frac{1}{2\iota}}\right) \cdot o(1), \end{aligned}$$

meaning that $\sum_{\ell=\lfloor \delta^{\frac{2\iota}{2\iota-1}} T\rfloor}^{T-1} \sqrt{\varphi_\ell} = o\left(T^{\frac{2\iota-1}{2\iota}}\right)$.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80 (5), 1977–2016.

Ahn, Seung C., and Alex R. Horenstein (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81 (3), 1203–1227.

Anderson, Theodore W. (1963): "Asymptotic Theory for Principal Component Analysis." *The Annals of Mathematical Statistics*, 34 (1), 122-148.

Bai, Jushan (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71 (1), 135–171.

Bai, Jushan (2004): "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics*, 122(1), 137–183.

Bai, Jushan, and Serena Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70 (1), 191–221.

Bai, Jushan, and Serena Ng (2023): "Approximate Factor Models With Weaker Loadings," *Journal of Econometrics*, 235 (2), 1893–1916.

Bai, Jushan, and Peng Wang (2016): "Econometric Analysis of Large Factor Models," *Annual Review of Economics*, 8 (1), 53–80.

Bhansali, Raj J., Liudas Giraitis, and Piotr S. Kokoszka (2007): "Convergence of Quadratic Forms With Nonvanishing Diagonal," *Statistics & Probability Letters*, 77 (7), 726–734.

Bleistein, Norman, and Richard A. Handelsman (1975): *Asymptotic Expansions of Integrals.* New York: Holt, Rinehart, Winston, 1975.

Cattell, Raymond B. (1966): "The Scree Test For The Number Of Factors," *Multivariate Behavioral Research*, 1 (2), 245–276.

Crump, Richard K., and Nikolay Gospodinov (2022): "On the Factor Structure of Bond Returns," *Econometrica*, 90 (1), 295–314.

Ding, Xiucai and Fan Yang (2021), "Spiked Separable Covariance Matrices and Principal Components," *The Annals of Statistics*, 49 (2), 1113–1138.

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin (2011): "A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering," *Journal of Econometrics*, 164 (1), 188–205.

Drees, Holger, Sidney Resnick, and Laurens de Haan (2000): "How to make a Hill plot," *The Annals of Statistics*, 28 (1), 254–274.

Fan, Ky (1951): "Maximum Properties and Inequalities for the Eigenvalues of Completely Continuous Operators," *Proceedings of the National Academy of Sciences*, 37 (11), 760–766.

Fan, Jianqing, Yuan Liao, and Martina Mincheva (2013): "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75 (4), 603–680.

Fan, Jianqing, Yuan Liao and Jiawei Yao (2015): "Power Enhancement in High-Dimensional Cross-Sectional Tests," *Econometrica*, 83 (4), 1497–1541.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin (2000): "The Generalized Dynamic-Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82 (4), 540–554.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin (2004): "The Generalized Dynamic Factor Model Consistency and Rates," *Journal of Econometrics*, 119 (2), 231–255.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007): "The US Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, 54 (8), 2291-2304.

Hall, Peter, and Christopher Charles Heyde (1980): *Martingale Limit Theory and its Application.* New York: Academic Press.

Hallin, Marc, and Roman Liška (2011): “Dynamic Factors in the Presence of Blocks,” *Journal of Econometrics*, 163 (1), 29–41.

He, Yi, Sombut Jaidee, and Jiti Gao (2023): “Most Powerful Test Against a Sequence of High Dimensional Local Alternatives,” *Journal of Econometrics*, 234 (1), 151–177.

He, Yi (2023): “Ridge Regression Under Dense Factor Augmented Models,” Forthcoming at *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2023.2206082.

Hill, Bruce M. (1975): “A Simple General Approach to Inference About the Tail of a Distribution,” *The Annals of Statistics*, 3 (5), 1163–1174.

Johnstone, Iain M., and Arthur Yu Lu (2009): “On Consistency and Sparsity for Principal Components Analysis in High Dimensions,” *Journal of the American Statistical Association*, 104 (486), 682–693.

Jung, Sungkyu, and James S. Marron (2009): “PCA Consistency in High Dimension, Low Sample Size Context,” *The Annals of Statistics*, 37 (6B), 4104–4130.

Kim, Dukpa, and Tatsushi Oka (2014): “Divorce Law Reforms and Divorce Rates in the USA: An Interactive Fixed-Effects Approach,” *Journal of Applied Econometrics*, 29, 231–245.

Ledoit, Olivier and Michael Wolf (2012): “Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices,” *The Annals of Statistics*, 40 (2), 1024–1060.

Litterman, Robert, and José Scheinkman (1991): “Common factors affecting bond returns,” *Journal of Fixed Income*, 1 (1), 54–61.

Liu, Haoyang, Alexander Aue, and Debashis Paul (2015): “On the Marčenko–Pastur Law for Linear Time Series,” *The Annals of Statistics*, 43 (2), 675–712.

Moon, Hyungsik Roger, and Martin Weidner (2015): “Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects,” *Econometrica*, 83 (4), 1543–1579.

Müller, Ulrich K., and Yulong Wang (2017): “Fixed-k Asymptotic Inference About Tail Properties,” *Journal of the American Statistical Association*, 112 (519), 1334–1343.

Onatski, Alexei (2009): “Testing Hypotheses About the Number of Factors in Large Factor Models,” *Econometrica*, 77 (5), 1447–1479.

Onatski, Alexei (2010): “Determining the Number of Factors from Empirical Distribution of Eigenvalues,” *The Review of Economics and Statistics*, 92 (4), 1004–1016.

Onatski, Alexei (2012): “Asymptotics of the Principal Components Estimator of Large Factor Models With Weakly Influential Factors,” *Journal of Econometrics*, 168 (2), 244–258.

Onatski, Alexei, and Chen Wang (2021): “Spurious Factor Analysis,” *Econometrica*, 89 (2), 591–614.

Parke, William R. (1999): "What is Fractional Integration?," *The Review of Economics and Statistics*, 81 (4), 632–638.

Phillips, Peter C. B., and Katsumi Shimotsu (2004): "Local Whittle Estimation in Nonstationary and Unit Root Cases," *The Annals of Statistics*, 32 (2), 656–692.

Piazzesi, Monika (2010): "Affine Term Structure Models," In *Handbook of Financial Econometrics*, Vol.1, Elsevier, Chapter 12, 691–758.

Shen, Dan, Haipeng Shen, and James S. Marron (2016): "A General Framework for Consistency of Principal Component Analysis," *Journal of Machine Learning Research*, 17 (150), 1–34.

Stock, James H., and Mark W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97 (460), 1167–1179.

Stock, James H., and Mark W. Watson (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20 (2), 147–162.

Uhlig, Harald (2009): "Comment on 'How Has the Euro Changed the Monetary Transmission Mechanism?' " in *NBER Macroeconomic Annual 2008*, Vol. 23, ed. by D. Acemoglu, K. Rogoff, and M. Woodford.

Wolfers, Justin (2006): "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results," *American Economic Review*, 96 (5), 1802–1820.

Zhang, Bo, Guangming Pan, and Jiti Gao (2018): "CLT for Largest Eigenvalues and Unit Root Testing for High-Dimensional Nonstationary Time Series," *The Annals of Statistics*, 46 (5), 2186–2215.

Zhang, Bo, Jiti Gao, and Guangming Pan (2020): "Estimation and Testing for High–Dimensional Near Unit Root Time Series," Department of Econometrics and Business Statistics Working Paper Series 12/20, Monash University.