UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge Working Papers in Economics

Cambridge-INET Working Paper Series No: 2020/01

Cambridge Working Papers in Economics: 2003

SPURIOUS FACTOR ANALYSIS

Alexei Onatski
(University of Cambridge)

Chen Wang
(University of Hong Kong)

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

Spurious Factor Analysis

Alexei Onatski* and Chen Wang†

*Faculty of Economics, University of Cambridge.

†Department of Statistics and Actuarial Science,

University of Hong Kong.

January 13, 2020

Abstract

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

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

1 Introduction

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

Both strategies are used in practice. For example, Moon and Perron (2007), Eickmeier (2009), Wang and Wu (2015), von Borstel et al. (2016), and Barigozzi et al. (2018) fit factor models to non-stationary data after first-differencing them. Stock and Watson (2016) not only first-difference most of the series entering their dynamic factor model of the US economy, but also locally demean the variables to minimize problems associated with low-frequency variability. On the other hand, Bai (2004), Corielli and Marcellino (2006), Ghate and Wright (2012), West and Wong (2014), and Engel et al. (2015) estimate factor models on non-stationary data in levels.

Factor estimation in levels relies on the assumption of stationary errors. Banerjee et al. (2017, section 4.1) give "several reasons for making the hypothesis of I(0) idiosyncratic errors" in macroeconomic applications. One of their reasons is a very high rejection rate of the hypothesis of a unit root in the estimated idiosyncratic components of the 114 nonstationary monthly US macroeconomic series for the 1959-2014 period (see their Footnote 5).

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

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

Uhlig (2009) attributes the high explanatory power of the spurious factors to the fact that the simulated data are considerably autocorrelated. Many of the simulated

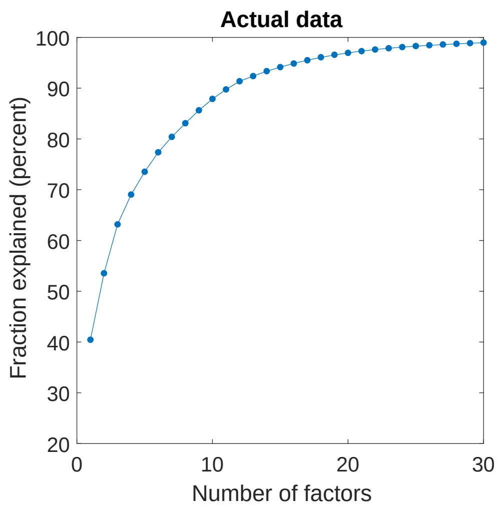

Figure 1: Factor contribution to the overall variance. Left panel: actual Boivin et al.'s (2009) data. Right panel: factorless simulated data with similar autocorrelation properties.

series' first-order autocorrelation coefficients are close to unity. In a finite sample (in his setting, 83 observations), the series may appear to be correlated, which will be picked up by the principal components. Although this explanation is intuitive, Uhlig admits that it is "perhaps tricky to formalize".

In this paper, we do such a formalization at different levels of generality. In our basic setting, the data are generated by a high-dimensional integrated system with an increasing number of common stochastic trends, none of which is dominating the rest asymptotically. An extreme example would be a panel of cross-sectionally *independent* difference-stationary processes. The setting also covers more empirically relevant situations with any types of cross-sectional dependence except those caused by the presence of a fixed number of genuine strong nonstationary factors in the data.

We prove that in such a setting the fraction of the data variation explained by the first principal component converges in probability to $6/\pi^2 \approx 0.61$ even when the data do not contain any common factors. The first three principal components together asymptotically explain $100\% \sum_{j=1}^{3} 6/(j\pi)^2 \approx 83\%$ of the variation in the *factorless* nonstationary data. The corresponding "factor estimates" converge to deterministic cosine waves that resemble linear, quadratic, and cubic time trends.

The flavour of these results is preserved in a more general setting of a local level model, where the data are represented by a weighted sum of I(1) and I(0) processes with the weights on the former possibly decaying to zero as the sample size increases. Furthermore, our conclusions do not change qualitatively when data contain local-to-unit roots, and when data are not only demeaned but also standardized before the principal components analysis (PCA) is applied.

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

The peculiar results of the PCA of factorless nonstationary data are relatively easy to explain in the extreme case where the data are given by cross-sectionally i.i.d. random walks. In such a case, the sample covariance matrix used by the PCA to extract "factors" can be interpreted as a discrete time approximation for the covariance operator of a demeaned Wiener process. As the data dimensions grow, the PCA estimates of the "factors" converge to the eigenfunctions of the covariance operator, which happen to be the cosine waves. The explanatory power of the estimated "j-th factor" converges to the j-th largest eigenvalue of the covariance operator, which equals $6/(j\pi)^2$.

A somewhat different explanation relates to the Karhunen-Loève expansion of the demeaned Wiener process (e.g. Shorack and Wellner, 1986). The expansion represents the process in the form of an infinite sum of trigonometric functions with uncorrelated random coefficients whose variances are quickly decaying. Since difference-stationary series can be approximated by Wiener processes, much of the variation in a nonstationary panel can be captured by a few of the trigonometric functions corresponding to the first terms in the Karhunen-Loève expansion.

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

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

These functions are similar to the deterministic polynomial trends. Hence, the intercept-only Dickey-Fuller statistic computed on the basis of estimated idiosyncratic terms asymptotically behaves similarly to the intercept-only Dickey-Fuller statistic for the regression that includes several deterministic polynomial time trends. This leads to a substantial size distortion[2] and a potentially confused conclusion that the factors soak up all or most of the nonstationarity in the data.

All in all, the results of the principal components analysis of the levels of non-stationary data may be very misleading. We recommend to always compare the first differences of factors estimated from the levels with factors estimated from the first-differenced data. A mismatch indicates a spurious factor analysis in levels. In Section 5, we derive a theory-based threshold for the amount of the mismatch which must raise the alarm.

The remainder of the paper is structured as follows. In Section 2, we formally introduce our setting and present our main results. Section 3 discusses various extensions to the basic setting. Section 4 studies the workings of the information criteria for the determination of the number of factors in the context of spurious factor analysis. Section 5 discusses ways to detect spurious results. Section 6 concludes. Monte Carlo results are reported in the Appendix while all proofs are given in the Supplementary Material (SM).

2 Basic setup and main results

Consider an N-dimensional integrated system

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

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

[2] We confirm this claim using Monte Carlo analysis reported in the Appendix.

where ε_t is N_ε-dimensional, and matrix $\Psi(1)$ may be of deficient rank so that cointegration is allowed. Suppose that data are summarized by the $N \times T$ matrix $X = [X_1, ..., X_T]$. Our goal is to study the workings of the PCA of these data as both N and T go to infinity, without any constraints on the relative speed of growth.

In contemporary economic applications, the PCA is often used to estimate factors F and loadings Λ in the factor model for the temporarily demeaned data[3]

$$X - \bar{X} = \Lambda F' + e. \tag{2}$$

The common factors are often interpreted as a few important latent variables affecting a vast number of economic indicators (rows of X). See Stock and Watson (2016) for a review of the related literature. Of course, in general, data generated from (1) do not have a factor structure. For example, if $\Psi(L)$ is diagonal, then the data are cross-sectionally independent and there are clearly no common factors.

Suppose that a researcher, nevertheless, models the data by (2). The PCA estimates of the first r factors are then defined as the r principal eigenvectors, $\hat{F}_1, ..., \hat{F}_r$, of

$$\hat{\Sigma} = \left(X - \bar{X}\right)' \left(X - \bar{X}\right)/N. \tag{3}$$

The corresponding principal eigenvalues $\hat{\lambda}_1 \geq ... \geq \hat{\lambda}_r$ estimate the explanatory power of the factors. Precisely, $\hat{\lambda}_j / \operatorname{tr} \hat{\Sigma}$ is interpreted as the fraction of the data variation explained by the j-th factor.

Below, we show that such a principal components analysis may be spurious in the sense that $\hat{\Sigma}$ has a few eigenvalues that dominate the rest, but the corresponding eigenvectors do not represent any latent economic factors driving the dynamics of the data. Instead, they capture deterministic trends that explain a large share of variation in any time series that are integrated of order one.

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

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

Note that ε_{it} may have different distributions, although they have to be independent. Further, the normalization $\mathbb{E}\varepsilon_{it}^2 = 1$ is not restrictive as it may be accommo-

[3] We consider the case of demeaned and standardized data in the next section.

dated by the lag polynomial $\Psi(L)$.

Assumption A2. *As* $N \to \infty$, $\sum_{k=0}^{\infty} (1+k) \|\Psi_k\| = O(N^{\alpha})$ *for some* $\alpha \geq 0$, *where* $\|\cdot\|$ *denotes the spectral norm of a matrix.*

This assumption mildly restricts the form of temporal and cross-sectional dependence in the data. Although our setting *does not imply* the existence of common factors in the data, it does allow for them when $\alpha > 0$. For a simple example, consider a basic factor model

$$X_t = \Lambda F_t + e_t, \tag{4}$$

where the N_F factors follow independent random walks, the loadings are normalized so that $\Lambda'\Lambda/N = I_{N_F}$, and the idiosyncratic component is white noise. Such X_t satisfies (1) with $\varepsilon_t' = \left(F_t' - F_{t-1}', e_t'\right)$ and $\Psi(L) = [\Lambda, I_N] - [0, I_N] L$. We have $\sum_{k=0}^{\infty} (1+k) \|\Psi_k\| = \sqrt{N+1} + 2$, and therefore A2 is satisfied with $\alpha = 1/2$.

Assumption A3. *The so-called effective rank*[4] *of the long-run covariance matrix* $\Omega = \Psi(1) \Psi(1)'$, *defined as* $\operatorname{tr} \Omega / \|\Omega\|$, *diverges to infinity as* $N \to \infty$.

If the effective rank diverges, the rank, defined in the standard way, must diverge too. Hence, an immediate consequence of A3 is that the rank of $\Psi(1)$ diverges to infinity as $N \to \infty$. In other words, the number of stochastic trends in the data is increasing with the dimensionality.

The assumption does not allow a finite number of such trends to dominate the rest, so that $\|\Omega\|$ is not allowed to dominate $\operatorname{tr} \Omega$ asymptotically. In particular, A3 precludes the existence of a fixed number of strong nonstationary factors in the data. However, the existence of a growing number of such factors as well as the total absence of any factors is allowed.

To illustrate this, consider example (4) again. There we have $\Psi(1) = [\Lambda, 0]$, $\operatorname{tr} \Omega = N N_F$ and $\|\Omega\| = N$. Hence, the effective rank of Ω equals N_F, and A3 is satisfied if the number of strong factors grows with N. The rate of the growth may be arbitrarily slow. For another example, let the data consist of N independent pure random walks, so there are no common factors whatsoever. Then $\Psi(1) = I_N$, the effective rank of Ω equals N, and A3 is satisfied again.

We would like to stress that A3 fails in situations where data contain a fixed number of strong factors. Moreover, this failure does not depend on whether the idio-

[4] The concept of the effective rank, or effective dimension, has been used in several recent studies of high-dimensional problems (e.g. Vershynin (2012), Koltchinskii and Lounici (2016)).

syncratic terms are stationary or not. For example, if e_t in (4) consists of independent random walks instead of white noises, X_t satisfies (1) with $\varepsilon_t' = \left(F_t' - F_{t-1}', e_t' - e_{t-1}'\right)$ and $\Psi(L) = [\Lambda, I_N]$. Hence, $\Omega = \Lambda\Lambda' + I_N$ and the effective rank of Ω equals $N(N_F + 1)/(N+1)$, which remains bounded with fixed N_F, so that A3 is violated.

On the other hand, A3 still holds when the data contain a fixed number of weaker factors (such that $\|\Lambda\Lambda'\| = o(N)$). Then, the effective rank of $\Omega = \Lambda\Lambda' + I_N$ is no smaller than $N/(1 + o(N))$, which obviously diverges as required by A3.

Theorem 1 *Let "$\xrightarrow{\mathrm{P}}$" denote convergence in probability. Suppose A1-A3 hold. If*

$$N^{2\alpha}(T + N_\varepsilon) / \left(T^2 \operatorname{tr} \Omega\right) \to 0 \tag{5}$$

as $N, T \to \infty$, then for any fixed positive integer k,
(i) $\left|\hat{F}_k' d_k\right| \xrightarrow{\mathrm{P}} 1$, where $d_k = (d_{k1}, ..., d_{kT})'$ with $d_{kt} = \sqrt{2/T}\cos(\pi k t/T)$.
(ii) $\hat{\lambda}_k / (\gamma_N T^2) \xrightarrow{\mathrm{P}} (k\pi)^{-2}$, where $\gamma_N = \operatorname{tr}\Omega/N$.

$$\textit{If } \min\{N, T\} N^{2\alpha}(T + N_\varepsilon) / \left(T^2 \operatorname{tr}\Omega\right) \to 0, \textit{ then} \tag{6}$$

(iii) $\hat{\lambda}_k / \operatorname{tr}\hat{\Sigma} \xrightarrow{\mathrm{P}} 6/(k\pi)^2$.

Let us first interpret the theorem's results, and then discuss conditions (5) and (6) that link N, N_ε, T, and $\operatorname{tr}\Omega$.

Part (i) of the theorem reveals that the "factor" estimates converge in probability to deterministic cosine functions in the sense that the angle between the vector of estimates and the vector of uniform grid values of the corresponding cosine function converges in probability to zero. Figure 2 plots the cosine functions corresponding to the first three "factors". They may be interpreted as the trigonometric versions of the linear, quadratic, and cubic trends.

As mentioned in the Introduction, the functions can be linked to the Karhunen-Loève expansion of the demeaned Wiener process $\tilde{W}(x) = W(x) - \int_0^1 W(x)\,\mathrm{d}x$. Its covariance kernel has eigenfunctions $\sqrt{2}\cos(\pi k x)$, $k = 1, 2, ...$, corresponding to eigenvalues $(\pi k)^{-2}$ (e.g. Müller and Watson (2008, Thm. 1)). Therefore, the Karhunen-Loève expansion of $\tilde{W}(x)$ has the following form

$$\tilde{W}(x) = \sqrt{2}\sum_{k=1}^{\infty}(\pi k)^{-1}\cos(\pi k x) z_k, \tag{7}$$

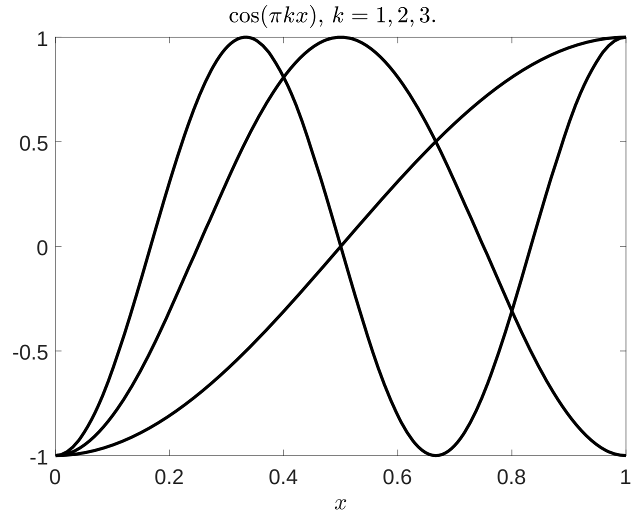

Figure 2: The probability "limits" of the first three spurious factor estimates.

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

For each of the data series X_{jt} that are difference-stationary, define $Y_{jT}(x) = (f_j(0)T)^{-1/2} X_{j[xT]}$, where $f_j(0)$ is the spectral density of $X_{jt} - X_{j,t-1}$ at frequency zero. As is well-known (e.g. Phillips, 1986), functions $Y_{jT}(x)$ weakly converge to $W(x)$ and thus,

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

weakly converge to $\tilde{W}(x)$. Therefore, the demeaned series X_{jt}, divided by $(f_j(0)T)^{1/2}$, can asymptotically be represented by the Karhunen-Loève expansion of $\tilde{W}(x)$. In particular, functions $\cos(\pi kt/T)$ with $k = 1, 2, ...$ capture much of the variation in each of $X_{jt} - \bar{X}_j$, which agrees with the theorem's first result intuitively.

The above arguments suggest that we should expect a flavour of spurious factor analysis to be present even in the PCA of the nonstationary data of fixed dimension N. The cosines would still be capturing much of the common variation in the data (regressing the data on them would produce high R^2), although the PCA estimators of the "factors" would no longer converge to these cosines.

Figure 3 illustrates statement (ii) of the theorem by showing the asymptotic scree plot for data satisfying the theorem's assumptions. The height of the plot is scaled so that the largest eigenvalue equals one. A typical interpretation of such a plot would be that the data "obviously" contain at least one strong factor, but perhaps two, or

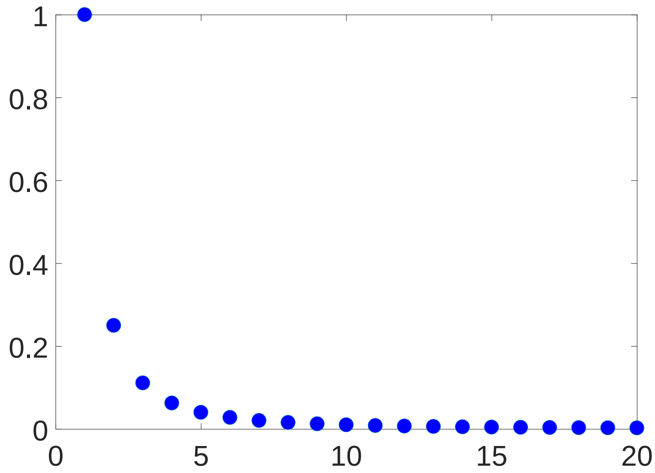

Figure 3: The asymptotic scree plot for possibly factorless persistent data (the first 20 normalized eigenvalues only). The horizontal axis shows the order k of the eigenvalue λ_k. The vertical axis shows the probability limit of λ_k/λ_1.

even three of them. Theorem 1 (ii) shows that such an interpretation may potentially be very misleading as the data may be totally factorless.

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

Let us now discuss conditions of the theorem that link N, N_ε, T, and $\operatorname{tr}\Omega$. For an extreme example where $\Psi(L) = I_N$, so the data consist of independent pure random walks, we have $\alpha = 0$, $N_\varepsilon = N$, and $\operatorname{tr}\Omega = N$. Hence, conditions (5) and (6) are trivially satisfied. It is easy to see that the conditions continue to hold for non-diagonal $\Psi(L)$ (so the data consist of cross-sectionally dependent I(1) processes) as long as $\Psi(L)$ satisfies A2 with $\alpha = 0$, $\|\Psi(1)\|$ and $\|\Psi(1)^{-1}\|$ remain bounded, and $N_\varepsilon = N$.

For example, let the first differenced data follow an autoregression $\Delta X_t = \rho \Delta X_{t-1} + e_t$ with $|\rho| < 1$, where $e_t = (e_{1t}, ..., e_{Nt})'$ are generated by "cross-sectional autoregressions" $e_{it} = \gamma e_{i-1,t} + \varepsilon_{it}$ with $|\gamma| < 1$ and $e_{0t} = 0$. Then $\Psi(L) = (1 - \rho L)^{-1} \Gamma_N$, where Γ_N is an N-dimensional lower triangular Toeplitz matrix with ones on the main diagonal and γ^j on the j-th sub-diagonal. As is well known (e.g. Böttcher and Silbermann (1999, Corollary 4.19)),

$$\lim_{N\to\infty} \|\Gamma_N\| = (1-\gamma)^{-1} \text{ and } \lim_{N\to\infty} \left\|\Gamma_N^{-1}\right\| = (1+\gamma)^{-1}.$$

Therefore, $\Psi(L)$ satisfies A2 with $\alpha = 0$, whereas $\|\Psi(1)\|$ and $\left\|\Psi(1)^{-1}\right\|$ converge to finite positive numbers.

Note that the conditions of the theorem do not require all data to be integrated. Suppose, for example, that $\Psi(L)$ is diagonal with first n diagonal elements equal one, and the rest equal $1-L$. Then, the first n data series are random walks whereas the last $N-n$ series are white noise. Obviously, $N_\varepsilon = N$, A2 holds with $\alpha = 0$, and A3 holds when $\operatorname{tr}\Omega = n \to \infty$. Condition (5) becomes equivalent to $(T+N)/(T^2 n) \to 0$. Hence, *for any* $n\to\infty$, it holds with $N = O(T^2)$, whereas (6) holds with $N = O(T)$.

The fact that a relatively small number n of I(1) series so strongly influence the PCA results can be partially blamed on the different scale of I(1) and I(0) series. The effect of such a scale difference would be eliminated by standardizing the data. We study consequences of the standardization in the next section.

Here, we point out that the effect of the scale difference can also be eliminated by dividing I(1) series by $\sqrt{T}$. Such an adjustment transforms (5) to $(T+N)/(Tn) \to 0$. For $N = O(T)$, this constraint is not binding under the maintained assumption that $n\to\infty$. In particular, if the data contain any increasing number of I(1) series, the PCA estimate of the first "factor" would converge to a deterministic cosine wave, *even after dividing* the I(1) series by $\sqrt{T}$.

Finally, consider the basic factor model example (4) with the number of factors $N_F \to \infty$. In that example, sufficient condition (5) for statements (i-ii) of the theorem to hold becomes $(T+N+N_F)/(T^2 N_F) \to 0$. In particular, if $N = O(T^2)$, the PCA estimates of a few of the strongest factors converge to deterministic cosine waves even though the data do contain an increasing number of genuine strong factors,[5] which may be different from the cosine waves.

[5] As discussed above, A3 is violated and our theorem does not hold when the number of factors N_F is fixed.

3 Extensions

In this section, we consider three extensions to our basic setting.

3.1 Local level model

Suppose that data $Y_t, t = 1, ..., T$, are weighted sum of $I(1)$ and $I(0)$ components

$$Y_t = \omega_T X_t + Z_t, \tag{8}$$

where $\omega_T \neq 0$ is possibly decreasing with the sample size T, X_t is generated by the integrated system (1) as in the previous section, and Z_t is an N-dimensional linear stationary process. Specifically, $Z_t = \Pi(L)\eta_t$, where $\Pi(L) = \sum_{k=0}^{\infty} \Pi_k L^k$ and η_t is an N_η-dimensional random vector with components η_{it}. We make the following assumption.

Assumption A4. *Random variables η_{it} with $i \in \mathbb{N}$ and $t \in \mathbb{Z}$ are independent and such that $\mathbb{E}\eta_{it} = 0$, $\mathbb{E}\eta_{it}^2 = 1$, and $\tau_4 = \sup_{i\in\mathbb{N},t\in\mathbb{Z}} \mathbb{E}\eta_{it}^4 < \infty$. Further, $\sum_{k=0}^{\infty} (1 + k) \|\Pi_k\| = O\left(N^\beta\right)$ for some $\beta \geq 0$ as $N \to \infty$.*

The part of the assumption describing properties of η_{it} parallels assumption A1 for ε_{it}. We do not assume that η_{it} and ε_{it} are mutually independent so Z_t and X_t may depend on each other. The second part of A4 parallels A2. The constant β is introduced to allow component Z_t of the data to contain some genuine common factors.

Let $Y = [Y_1, ..., Y_T]$ be the $N \times T$ data matrix. Let $\check{\lambda}_1 \geq ... \geq \check{\lambda}_T$ and $\check{F}_1, ..., \check{F}_T$ be the eigenvalues and corresponding eigenvectors of $\check{\Sigma} = \left(Y - \bar{Y}\right)' \left(Y - \bar{Y}\right)/N$.

Theorem 2 *Under A1-A4, if (5) holds and*

$$N^{2\beta} (T + N_\eta) / \left(\omega_T^2 T^2 \operatorname{tr}\Omega\right) \to 0 \tag{9}$$

as $N, T \to \infty$, then for any fixed positive integer k,
(i) $\left|\check{F}_k' d_k\right| \xrightarrow{\text{P}} 1$, where $d_k = (d_{k1}, ..., d_{kT})'$ with $d_{kt} = \sqrt{2/T} \cos(\pi k t/T)$.
(ii) $\check{\lambda}_k / (\omega_T^2 \gamma_N T^2) \xrightarrow{\text{P}} (k\pi)^{-2}$, where $\gamma_N = \operatorname{tr}\Omega/N$.
(iii) If (6) holds and

$$\min\{N, T\} N^{2\beta} (T + N_\eta) / \left(\omega_T^2 T^2 \operatorname{tr}\Omega\right) \to 0, \tag{10}$$

then $\check{\lambda}_k / \operatorname{tr} \check{\Sigma} \xrightarrow{\text{P}} 6/(k\pi)^2$.

As an illustration, consider a simple situation where the components of X_t and Z_t are independent random walks and white noises, respectively. Then A2 holds with $\alpha = 0$ and A4 holds with $\beta = 0$. Furthermore, $\operatorname{tr} \Omega = N_\varepsilon = N_\eta = N$. Therefore, (5) trivially holds, whereas (9) holds if $(T + N) / (\omega_T^2 T^2 N) \to 0$. If T and N diverge to infinity proportionally, the latter convergence holds as long as ω_T goes to zero slower[6] than $1/T$.

For another example, suppose that the components of X_t are independent random walks, as in the previous example. However, Z_t now contains a strong stationary factor so that $Z_t = \Gamma f_t + e_t$, where $\Gamma'\Gamma = N$ while the factor f_t and the components of e_t are independent white noises. Then $\beta = 1/2$ and $N_\eta = N + 1$. Hence, for (9) to hold we need $N(T + N + 1) / (\omega_T^2 T^2 N) \to 0$. If T and N are proportional, the latter convergence holds as long as ω_T goes to zero slower than $1/\sqrt{T}$.

Condition (10) is harder to satisfy than (9). In the first of the above examples, it requires that ω_T goes to zero slower than $1/\sqrt{T}$ (assuming that T and N are proportional). For the second example, it fails for ω_T that converges to zero at any rate when T and N are proportional, but holds for ω_T going to zero slower than $\sqrt{N/T}$ when T grows faster than N.

3.2 Local-to-unit roots

Now consider data having local-to-unit roots,

$$X_t - \mu_X = \rho (X_{t-1} - \mu_X) + \Psi(L) \varepsilon_t \tag{11}$$

with initial values X_0, where $\rho = \operatorname{diag}\{\rho_1, ..., \rho_N\}$ and $\rho_j = \exp\{-\phi_j/T\}$, $\phi_j \geq 0$, are local to unity. As above, we do allow $\Psi(1)$ to be of deficient rank, which may be interpreted as an analogue of the standard cointegration setting. We do not put any restrictions on the N-dimensional vector μ_X.

Literature on near integrated systems (e.g. Phillips (1988), Elliott (1998)) usually considers a triangular form of the system, where the data generating process for

[6] When $\omega_T = w/T$ with fixed $w > 0$, the eigenvalues of the sample covariance matrix still decay very fast, although the probability limits described by Theorem 2 (ii) should be altered. Similarly, the eigenvectors become imperfectly collinear with the cosine waves described by (i). The interested reader can find a partial analysis of the situation $\omega_T = w/T$ with fixed $w > 0$ in the SM's Section 3.1.2.

near integrated stochastic trends and the "cointegrating" relationships are modelled explicitly as two sub-systems. We work with (11) because this form is amenable to the analysis similar to that of (1).

As is well known (e.g. Phillips (1988), Stock (1994)), as $T \to \infty$, the step functions corresponding to normalized components of (11) weakly converge to the Ornstein-Uhlenbeck (OU) processes with decay rates ϕ_j. Had these decay rates been the same for all j, the situation would have been analogous to the unit root case with the Wiener process replaced by the OU process. Then we would have expected that the PCA "factors" correspond to the eigenfunctions of the covariance kernel of the demeaned OU process. As we show below, when ϕ_j are different, the spurious factors correspond to eigenfunctions of a weighted average of the covariance kernels of the demeaned OU processes with different decay rates.

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

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

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

Let $k_\phi\left(s,t\right)$ be the covariance kernel of the demeaned stationary OU process with decay rate ϕ. Precisely,

$$k_\phi\left(s,t\right) = a_\phi\left(s,t\right) - \int_0^1 a_\phi\left(s,t\right) \mathrm{d}s - \int_0^1 a_\phi\left(s,t\right) \mathrm{d}t + \int_0^1 \int_0^1 a_\phi\left(s,t\right) \mathrm{d}s\mathrm{d}t,$$

where $a_\phi\left(s,t\right) = e^{-\phi|t-s|}/\left(2\phi\right)$ is the covariance kernel of the stationary OU process before the demeaning (e.g. Karatzas and Shreve (1991, p. 358)). Further, let $k_0\left(s,t\right)$ be the covariance kernel of the demeaned standard Wiener process. Define the weighted average kernel as

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

where $\mathcal{F}$ is a probability distribution on $[0,\infty)^2$. Let $K_{\mathcal{F}}$ be the integral operator, acting in the space $C[0,1]$ of continuous functions on $[0,1]$, with kernel $k_{\mathcal{F}}(s,t)$.

Let $\mathcal{F}_{N,T}$ be the empirical joint distribution of Ω_{jj} and ϕ_j, $j = 1, ..., N$, where Ω_{jj} is the j-th diagonal element of $\Omega = \Psi(1)\Psi(1)'$. We will make the following assumptions.

Assumption A5. *$\mathcal{F}_{N,T}$ weakly converges to $\mathcal{F}$ as $N, T \to \infty$. The supports of $\mathcal{F}_{N,T}$ and $\mathcal{F}$ belong to $[0,\bar{\omega}] \times [0,\bar{\phi}]$ for some $0 < \bar{\omega}, \bar{\phi} < \infty$. The eigenvalues $\mu_1 > \mu_2 > ...$ of $K_{\mathcal{F}}$ are simple.*

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

The restriction on the supports of $\mathcal{F}_{N,T}$ and $\mathcal{F}$ implies that $\Omega_{jj} \leq \bar{\omega}$ for all j. Note that $\Omega_{jj}/(2\pi)$ equals the spectral density at frequency zero of the quasi-difference $X_{jt} - \rho_j X_{j,t-1}$. Hence, A5 requires that such spectral densities are bounded. Furthermore, the assumption $\mu_1 > \mu_2 > ...$ implies that the distribution $\mathcal{F}$ cannot be concentrated at $\omega = 0$. In other words, a nontrivial fraction of the series have spectral densities at frequency zero that are bounded away from zero, and hence, $\operatorname{tr}\Omega$ diverges to infinity at the same rate as N.

Assumption A2a. *In addition to A2, as $N \to \infty$, $\sum_{k=0}^{\infty}(1+k)\|\Psi_k\|_F = O(N^\alpha + \min\{N^{1/2}, N_\varepsilon^{1/2}\})$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.*

This assumption adds the Frobenius norm constraint to the constraint imposed on the spectral norm by A2. In cases where $\alpha = 0$, A2a is equivalent to A2 because $\|\Psi_k\|_F \leq \min\{N^{1/2}, N_\varepsilon^{1/2}\}\|\Psi_k\|$ for all $k \geq 0$. For $\alpha > 0$, a sufficient condition for A2a to hold is the existence of a decomposition $\Psi_k = \Psi_{1k} + \Psi_{2k}$, where Ψ_{1k} has a fixed rank and satisfies $\sum_{k=0}^{\infty}(1+k)\|\Psi_{1k}\| = O(N^\alpha)$, whereas Ψ_{2k} may have an unbounded rank, but satisfies $\sum_{k=0}^{\infty}(1+k)\|\Psi_{2k}\| = O(1)$. Such a decomposition would arise, for example, in situations where the data contain a fixed number of factors, represented by linear combinations $F_t = \sum_{k=0}^{\infty}\psi_k'\varepsilon_{t-k}$ with bounded $\|\psi_k\|$, and $\Psi_{1k} = \Lambda\psi_k'$ with $\|\Lambda\| = N^{\alpha}$.

Theorem 3 *Under A1, A2a, A3, and A5, if* $(T+N_\varepsilon)\, N^{2\alpha-1}/T^2 = o(1)$*, then for any fixed positive integer* k*,*
(i) $\left|\hat{F}_k' d_k\right| \xrightarrow{\mathrm{P}} 1$*, where* $d_k = (\varphi_k(1/T), ..., \varphi_k(T/T))/\sqrt{T}$ *and* $\varphi_k(s)$ *is the* k*-th principal eigenfunction of* $K_{\mathcal{F}}$*.*
(ii) $\hat{\lambda}_k/T^2 \xrightarrow{\mathrm{P}} \mu_k$*, where* μ_k *is the* k*-th principal eigenvalue of* $K_{\mathcal{F}}$*.*
(iii) If $\min\{N,T\}(T+N_\varepsilon)\, N^{2\alpha-1}/T^2 = o(1)$*, then* $\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma} \xrightarrow{\mathrm{P}} \mu_k/\sum_{j=1}^{\infty}\mu_j$*.*

Although Theorem 3 does not give us closed form expressions for the limits of the normalized principal eigenvalues and eigenvectors of $\hat{\Sigma}$, its message is similar to that of previous theorems. First, the PCA may be spurious in the sense that the estimated factors do not reflect cross-sectional linkages in the data. Second, the principal eigenvalues of the sample covariance matrix decay fast (μ_k, $k = 1, 2, ...,$ being summable and thus, fast decreasing), creating an impression of high "explanatory" content of the "factors".

To illustrate Theorem 3, consider a simple scenario where $\Psi(L) = I_N$ so that $\mathcal{F}$ is concentrated at $\omega = 1$, and where $\mathcal{F}$ is uniform on $\left[0, \bar{\phi}\right]$ with respect to the local-to-unity parameter ϕ. Figure 4 plots the principal eigenfunctions φ_1, φ_2, and φ_3, which we compute numerically for the case, where $\bar{\phi} = 10$. It can be interpreted as an analogue of Figure 2 for the local-to-unity case and is qualitatively similar to that figure.

Figure 5 shows the proportions of variation "explained" by the first three spurious factors as functions of $0 \leq \bar{\phi} \leq 10$. For $\bar{\phi} = 0$ (the unit root case), these proportions equal $6/(k\pi)^2$, $k = 1, 2, 3$, as in Theorem 2 (iii). As $\bar{\phi}$ increases so that the local-to-unity roots may deviate from the unity further, the proportion of variation "explained" by the first factor decreases. For $\bar{\phi} = 10$, it equals 38%, which brings it closer to the explanatory power of Uhlig's first "factor" (see Figure 1) extracted from factorless persistent, but stationary, data.

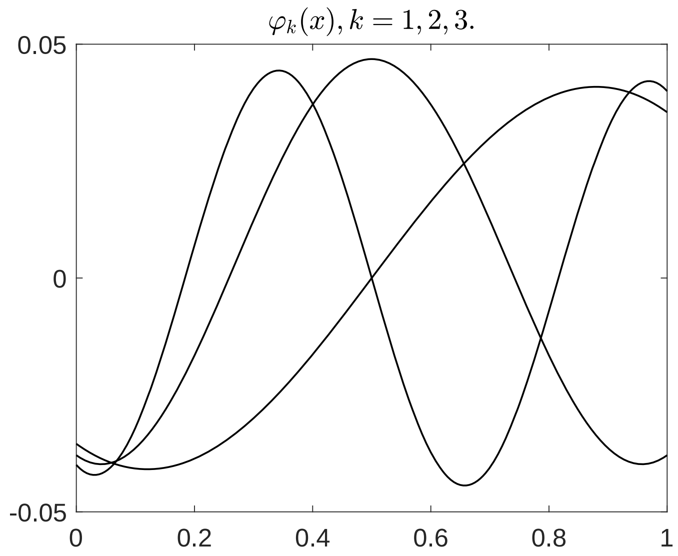

Figure 4: The probability "limits" of the first three spurious factor estimates. Local-to-unity parameter uniformly distributed on $[0, 10]$.

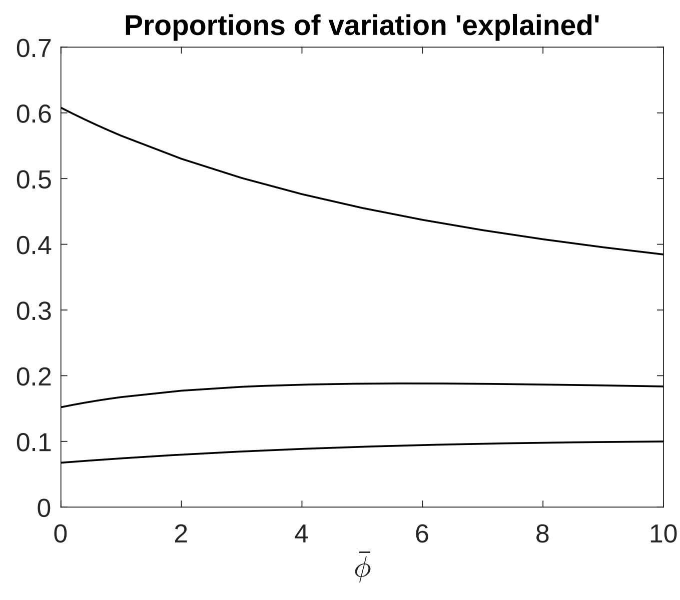

Figure 5: Proportions of the data variation "explained" by the first three spurious factors as functions of $\bar{\phi}$.

3.3 Demeaned and standardized data

In PCA applications, the data are often not only demeaned, but also standardized. As we show below, the spurious factor phenomenon is still present after the standardization.

We consider data generated by equation $X_t = X_{t-1} + \Psi(L)\varepsilon_t$, as in our basic setting. However, this time matrix $\hat{\Sigma}$ is defined as

$$\hat{\Sigma} = \left(X - \bar{X}\right)' D^{-1} \left(X - \bar{X}\right)/N,$$

where $D = \operatorname{diag}\left\{\left(X - \bar{X}\right)\left(X - \bar{X}\right)'/T\right\}$. This change substantially complicates our technical analysis. It requires us working with high-dimensional matrices whose entries are ratios of quadratic forms instead of just quadratic forms. As a result, our proofs for the demeaned case do not go through.

To overcome the technical challenge we simplify our setting.

Assumption A2b. *Matrix lag polynomial $\Psi(L)$ is diagonal. There exist absolute constants $B > 0$ and $b > 0$ such that $\max_i \sum_{k=0}^{\infty}(1+k)\left|(\Psi_k)_{ii}\right| \leq B$ and $\min_i \left|\sum_{k=0}^{\infty}(\Psi_k)_{ii}\right| \geq b$ for all N.*

Most important, we now require $\Psi(L)$ be diagonal, so our data are cross-sectionally independent. Although cross-sectionally independent data are rare in PCA applications, they are clearly factorless. Our point is to show that the PCA of such factorless data yields spurious factors even after the data are standardized. We leave analysis of cross-sectionally dependent standardized data for future research.

The existence of B, described in A2b, would follow from the diagonality of $\Psi(L)$ and A2 with $\alpha = 0$. The existence of b, described in A2b, is assumed to further simplify our proofs. It implies that all data series are integrated, so that no series, when first differenced, have zero spectral density at zero frequency.

Theorem 4 *Suppose that assumptions A1, A2b, and A3 hold. In addition, suppose that ε_{jt} are identically distributed. Then, for any fixed positive integer k,*
(i) $\left|\hat{F}_k' d_k\right| \xrightarrow{\mathrm{P}} 1$, *where* $d_k = (d_{k1}, ..., d_{kT})'$ *with* $d_{kt} = \sqrt{2/T}\cos(\pi k t/T)$.
(ii) $\hat{\lambda}_k/T \xrightarrow{\mathrm{P}} \nu_k$, *where* $\nu_k = \mathbb{E}\left(x_k^2/\sum_{j=1}^{\infty}(kx_j/j)^2\right)$ *with* x_j, $j = 1, 2, 3, ...$ *being i.i.d. standard normal random variables.*
(iii) $\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma} \xrightarrow{\mathrm{P}} \nu_k$.

Part (i) of the theorem shows that the standardization does not affect the asymptotic behavior of the spurious factors. They still converge to the cosine waves. However, the standardization does affect the form of the normalization of $\hat{\lambda}_k$ in (ii), as well as the form of the limits in (ii) and (iii).

The standardization removes the need for normalizing $\hat{\lambda}_k$ by the average long run variance parameter $\gamma_N = \operatorname{tr} \Omega / N$, as in Theorem 1 (ii). Further, since the conditional variance of an integrated process is of order T, the standardization leads to the situation where $\hat{\lambda}_k$ in Theorem 4 (ii) is divided by T as opposed to T^2 in Theorem 1 (ii).

Note that the limit $6/(k\pi)^2$ in Theorem 1 (iii) can be written in the form $1/\sum_{j=1}^{\infty}(k/j)^2$. Therefore, this limit can be obtained from the limit ν_k in Theorem 4 (iii) by replacing the chi-square variables x_i^2 by their expectation (unity).

Values of ν_k for different k can be obtained numerically. Our calculations show that $\nu_1 \approx 0.44$, $\nu_2 \approx 0.18$, and $\nu_3 \approx 0.095$. Hence, the "explanatory power" of the first spurious factor in the standardized setting is substantially lower than that in the non-standardized one, $6/\pi^2 \approx 0.61$. However, the "explanatory power" of the second and third spurious factors somewhat increase relative to the non-standardized $6/(2\pi)^2 \approx 0.15$ and $6/(3\pi)^2 \approx 0.068$. Overall, the first three spurious factors still "explain" an amazing 71.5% portion of variation in the factorless standardized data.

4 The "number of factors"

Now we return to the basic setup of Section 2 and ask the following question. What is the number of "factors" in factorless persistent data detected by information criteria? Bai (2004) proposes to estimate the number of factors in nonstationary panels by minimizing function

$$IPC(k) = V(k) + k\hat{\sigma}^2 p(N, T)$$

over $k = 0, 1, ..., k_{\max}$, where $V(k) = \operatorname{tr} \hat{\Sigma}/T - \sum_{j=1}^{k} \hat{\lambda}_j/T$, $\hat{\sigma}^2 = V(k_{\max})$, and $p(N,T)$ is one of the following three penalty functions

$$\begin{aligned} p_1(N,T) &= \alpha_T \frac{N+T}{NT} \log \frac{NT}{N+T}, \\ p_2(N,T) &= \alpha_T \frac{N+T}{NT} \log \delta_{NT}, \text{ or} \\ p_3(N,T) &= \alpha_T \frac{N+T-k}{NT} \log NT. \end{aligned}$$

Here $\alpha_T = T/(4 \log \log T)$, and $\delta_{NT} = \min\{N,T\}$.

Let us denote the value k that delivers the minimum of $IPC(k)$ based on penalty $p_j(N,T)$ as $\hat{k}_j$. Bai's (2004) Theorem 1 gives conditions under which $\hat{k}_j$ is consistent for the true number of factors. One of the theorem's assumptions is the weak temporary dependence of the idiosyncratic terms. Of course, it does not generally hold for data generated by N-dimensional integrated system (1). However, in actual empirical research, one would not know the validity of the assumptions. If the data are nonstationary, it would be natural to apply an IPC criterion.

As the following proposition shows, the asymptotic behavior of $\hat{k}_j$ is sensitive to the choice of $k_{\max}$. We consider the following two rules for choosing $k_{\max}$. One rule is fixing $k_{\max}$ independent of the data size. The other sets $k_{\max}$ at some small fraction of $\delta_{NT} = \min\{N,T\}$, say $k_{\max} = [\gamma \delta_{NT}]$.

Proposition 5 *Suppose A1-A3 and condition (6) of Theorem 1 hold. Further, let*

$$m_{NT} = \frac{1}{\delta_{NT}} + \frac{(T + N_\varepsilon) N^{2\alpha} \delta_{NT}}{T^2 \operatorname{tr} \Omega}.$$

(i) if $k_{\max}$ is fixed, then $\hat{k}_j \xrightarrow{\text{P}} 0$ as $N,T \to \infty$, for $j = 1,2,3$;
(ii) if $k_{\max} = [\gamma \delta_{NT}]$ with fixed $\gamma > 0$ and $m_{NT} p_j(N,T) \to 0$, then $\hat{k}_j \xrightarrow{\text{P}} \infty$ as $N,T \to \infty$, for $j = 1,2,3$.

For cases where N, N_ε, and T are of the same order of magnitude and $\alpha = 0$, the convergence $m_{NT} p_j(N,T) \to 0$ required by Proposition 5 (ii) is guaranteed if $p_j(N,T) / \operatorname{tr} \Omega \to 0$. The latter convergence holds whenever $\log N$ is asymptotically dominated by $\operatorname{tr} \Omega \log \log N$. This would happen, for example, for data that consist of i.i.d. random walks. Moreover, $\log N$ would be asymptotically dominated by

$\operatorname{tr}\Omega \log\log N$ even if the number of the random walks is $\log N$ while the rest $N - \log N$ series are white noises.

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

$$\log V(k) + kg(N,T).$$

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

In the Appendix we perform a Monte Carlo analysis of the finite sample behavior of $\hat{k}_j$ when data do not have any factors in them. We find that for empirically relevant data sizes and standard choices of $k_{\max}$, the estimated number of "factors" often equals two or three.

5 Problem detection

As we have seen above, factor analysis applied directly to large nonstationary panels may be spurious. This raises a question: how to detect spurious results? A simple, although inexact, check is to compare the time series plots of the estimated factors to the cosine functions. A similarity should raise the alarm.

As an example, consider Bai's (2004) analysis of sectoral employment in the US. Figure 6 replicates Figure 3 in Bai (2004). It shows the Bureau of Economic Analysis data (NIPA, Tables 6.5b and 6.5c) for the logarithm of employment across 58 sectors[7] in the US for the period from 1948 to 2000. The series are very persistent, and Bai (2004) identifies two nonstationary and one stationary factors in the data.

Figure 7 shows the time series plots of the PCA estimates of the three factors. Their resemblance to cosine functions is striking. It suggests that an extra caution should be exercised before structural interpretation of these factor estimates is at-

[7] Bai (2004) has data on 60 sectors. However, the data on two out of 60 sectors in the current versions of NIPA tables is incomplete. Therefore, we use 58 sectors.

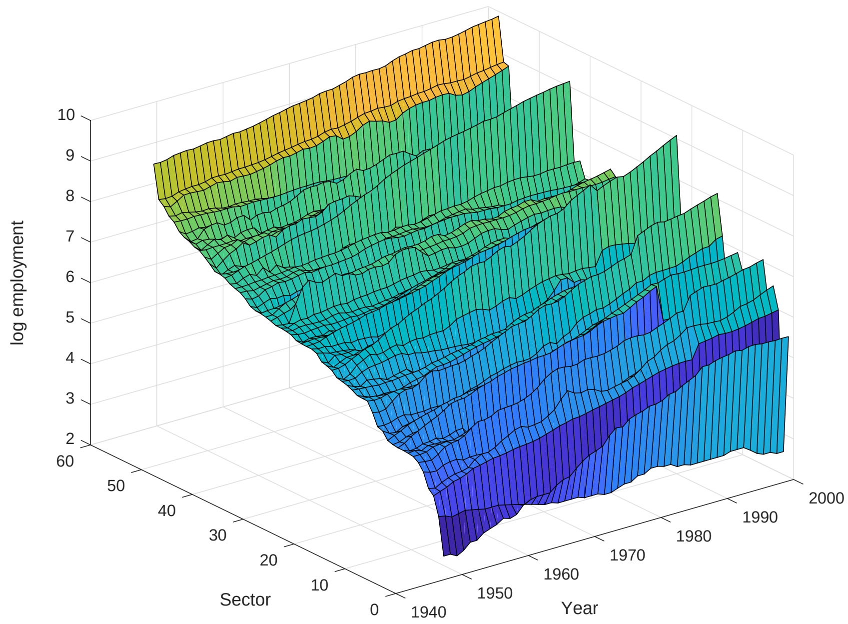

Figure 6: The number of full-time equivalent employees across 58 sectors. The sectors are arranged in ascending order according to their 1948 values.

tempted.

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

This strategy can be implemented as follows. Let P_V be the projection on the space spanned by the columns of matrix V. The quality of alignment between spaces spanned by the columns of $\Delta\hat{F}$ and of $\hat{f}$ can be measured by the eigenvalues of $P_{\Delta\hat{F}}P_{\hat{f}}$, which we denote as $\rho_1^2 \geq ... \geq \rho_r^2$. They may be interpreted as squared cosines of the principal angles between the spaces, or alternatively, as the squared sample canonical correlations between $\Delta\hat{F}$ and $\hat{f}$ (e.g. Hotelling, 1936). Observing

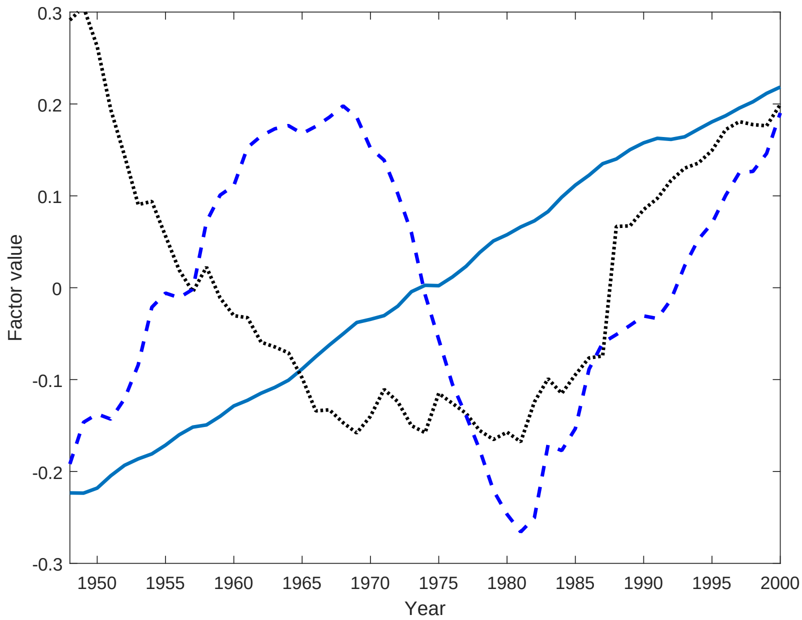

Figure 7: The principal components estimates of three factors in the employment data. The estimates are normalized to have unit Euclidean norms.

the squared canonical correlations substantially below unity indicates a problem.[8] Below, we derive a theory-based asymptotic threshold for $R_r = \sum_{i=1}^{r} \rho_i^2$.

Suppose that the N-dimensional data X_t, $t = 0, 1, .., T$ are truly generated by a factor model

$$X_t = \Lambda F_t + e_t \tag{12}$$

with r nonstationary factors and stationary idiosyncratic terms.

Assumption B1 (i) *The reciprocal of the smallest eigenvalue of* $\sum_{t=0}^{T} F_t F_t'/T^2$ *remains bounded in probability as* $T \to \infty$;
(ii) $E \|F_0\|^4 < C$ *for some positive constant* C;
(iii) *Let* $F_{dt} = F_t - F_{t-1}$. *Then* $\sum_{t=1}^{T} F_{dt} F_{dt}'/T \xrightarrow{\mathrm{P}} \Sigma_f > 0$ *and* $F_{dt} = \sum_{j=0}^{\infty} \Theta_j \zeta_{t-j}$, *where the components of* ζ_s, ζ_{is}, *are i.i.d.,* $E\zeta_{is} = 0$, $E\zeta_{is}^2 = 1$, $E\zeta_{is}^4 < C$, *and* $\sum_{j=0}^{\infty} (1 + j) \|\Theta_j\| < C$.

Assumption B2. *Loadings* Λ *are either deterministic or random, such that* $\Lambda'\Lambda/N \xrightarrow{\mathrm{P}}$

[8] The squared canonical correlations as a measure of discrepancy between subspaces related to factor estimates are used in recent work by Stock and Watson (2016) and Andreou et al (2019).

$\Sigma_\Lambda > 0.$

Assumption B3. *The matrix of idiosyncratic terms* $e = [e_0, ..., e_T]$ *satisfy* $\|e\| = O_{\mathrm{P}}\left(\sqrt{N} + \sqrt{T}\right).$

Assumption B4. Λ, $\{F_t\}$, *and* $\{e_t\}$ *are mutually independent.*

These assumptions are similar to those used in Bai (2004). The requirement in B1 that F_{dt} follows a linear process provides a convenient structure for our proofs. Assumption B3 puts mild restrictions on the serial and cross-sectional dependence of the idiosyncratic terms. For example, B3 holds if $e_{it} = \Psi(L)\varepsilon_{it}$, where ε_{it} satisfy A1 and $\Psi(L)$ satisfy A2 with $\alpha = 0$ (see Lemma 4 in the SM).

Slightly abusing previous notation, let $X = [X_0, ..., X_T]$ and $F = [F_0, ..., F_T]'$. Further let $\hat{F} = \left[\hat{F}_0, ..., \hat{F}_T\right]'$ be the $(T+1) \times r$ matrix whose i-th column equals the normalized i-th principal eigenvector of $X'X$, and let Δ be the $T \times (T+1)$ "differencing matrix" with all elements zero except the diagonal ones $\Delta_{ii} = -1$, and the super-diagonal ones $\Delta_{i,i+1} = 1$. Let $\hat{f}$ be the $T \times r$ matrix whose columns are the normalized principal eigenvectors of $\Delta X' X \Delta'$.

Note that

$$\frac{\Delta X' X \Delta'}{NT} = \Omega_0 + \Omega_1/\delta_{NT},$$

where $\delta_{NT} = \min\{N, T\}$, $\Omega_0 = \Delta\tilde{F}\Lambda'\Lambda\tilde{F}'\Delta'/NT$ with $\tilde{F} = F + e'\Lambda(\Lambda'\Lambda)^{-1}$, and $\Omega_1 = \delta_{NT}\Delta e' M_\Lambda e \Delta'/TN$ with $M_\Lambda = I_N - P_\Lambda$.

Lemma 6 *Under assumptions B1-B4,*

$$R_r = r - \mathcal{S}_r/\delta_{NT}^2 + O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right),$$

with $\mathcal{S}_r = \left\|\Omega_0^+ \Omega_1 M_{\Delta\tilde{F}}\right\|_F^2$, *where* Ω_0^+ *is the Moore-Penrose pseudoinverse of* Ω_0, *and* $\|\cdot\|_F$ *denotes the Frobenius norm.*

An immediate consequence of Lemma 6 is that R_r approaches r at the rate at least as fast as δ_{NT}^{-2}. Indeed, B1-B3 imply that the eigenvalues of Ω_0 converge to those of $\Sigma_\Lambda\Sigma_f$. Hence $\left\|\Omega_0^+\right\| = O_{\mathrm{P}}(1)$. On the other hand, B3 yields $\|\Omega_1\| = O_{\mathrm{P}}(1)$. Therefore $\mathcal{S}_r = O_{\mathrm{P}}(1)$, which implies the claimed convergence rate.

Of course, designing a practical threshold on R_r would require knowledge of the "scale" $\mathcal{S}_r$, which is a latent population parameter. The following lemma provides a consistent estimate of the scale, which is then used in Theorem 8 to construct the threshold.

Let $\hat{\lambda}$ be the $r \times r$ diagonal matrix with r principal eigenvalues of $X'X/NT^2$ on the diagonal. We estimate $M_{\Delta\tilde{F}}$, Ω_0, and Ω_1 by $M_{\Delta\hat{F}}$,

$$\begin{aligned}\hat{\Omega}_0 &= T\Delta\hat{F}\hat{\lambda}\hat{F}'\Delta', \text{ and} \\ \hat{\Omega}_1 &= \delta_{NT}\Delta X'X\Delta'/TN - \delta_{NT}\hat{\Omega}_0.\end{aligned}$$

Combining these estimates yield

$$\widehat{\mathcal{S}}_r = \left\| \hat{\Omega}_0^+ \hat{\Omega}_1 M_{\Delta\hat{F}} \right\|_F^2 .$$

Lemma 7 *Under assumptions B1-B4,* $\mathcal{S}_r - \widehat{\mathcal{S}}_r = O_{\mathrm{P}}\left(\delta_{NT}^{-1}\right)$. *As a consequence,*

$$R_r = r - \widehat{\mathcal{S}}_r/\delta_{NT}^2 + O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right).$$

Lemmas 6 and 7 imply that $\delta_{NT}^2\left(r - R_r\right) - \widehat{\mathcal{S}}_r$ converges in probability to zero as $N, T \to \infty$. This allows us to use $r - a\widehat{\mathcal{S}}_r/\delta_{NT}^2$ with any $a > 1$ as an asymptotically conservative threshold for R_r. To further ensure the conservativeness of the threshold one may use $a\left(\widehat{\mathcal{S}}_r + b\right)$ with $b > 0$ instead of $a\widehat{\mathcal{S}}_r$. This adjustment takes care of a possibility that $\widehat{\mathcal{S}}_r$ converges to zero, which may, theoretically, arise under assumptions B1-B4. For example, such a convergence would happen in a degenerate situation where the idiosyncratic components of the data, e_t, identically equal zero.

Theorem 8 *Under assumptions B1-B4,* $\delta_{NT}^2\left(r - R_r\right) - \widehat{\mathcal{S}}_r \xrightarrow{\mathrm{P}} 0$ *and, for any constants* $a > 1$ *and* $b > 0$,

$$\Pr\left(R_r > r - a\left(\widehat{\mathcal{S}}_r + b\right)/\delta_{NT}^2\right) \to 1.$$

A simple concrete choice of a, b would be $a = 2$ and $b = 1$. Then, the spurious results are flagged as soon as R_r is smaller than $r - 2\left(\widehat{\mathcal{S}}_r + 1\right)/\delta_{NT}^2$, or equivalently, as soon as $\delta_{NT}^2\left(r - R_r\right)$ is larger than $2\left(\widehat{\mathcal{S}}_r + 1\right)$. In the Appendix, we perform a Monte Carlo analysis that provides some evidence in support of this choice of a and b.

The value of $\delta_{NT}^2\left(r - R_r\right)$ with $r = 3$ for the sectoral employment data discussed above is 3,233.1. Assuming that data satisfy factor model (12) with three nonstationary factors, we would expect this value to be smaller than the threshold $2\left(\widehat{\mathcal{S}}_3 + 1\right)$. However, the value of this threshold is only 5.1, which indicates a potential spurious factor problem.

On the other hand, according to Bai (2004), there are both nonstationary and stationary factors in the sectoral employment data. Such a situation is not covered by our Theorem 8. It would be interesting and important to extend the theorem to cases where both nonstationary and stationary factors are present. We leave such an extension for future research.

6 Conclusion

This paper warns empirical researchers that a very high explanatory power of a few principal components of nonstationary data does not necessarily indicate the presence of factors. Even if such data are cross-sectionally independent, the first k principal components must explain $\sum_{j=1}^{k} 6/(j\pi)^2 \times 100\%$ of the variation, asymptotically. The extracted spurious factors correspond to the eigenfunctions of the auto-covariance kernel of the Wiener process and do not represent any cross-sectional common shocks driving the data's dynamics.

Unfortunately, the standard criteria for the determination of the number of factors are sensitive to the choice of the maximum number of factors $k_{\max}$. For empirically relevant data sizes and standard choices of $k_{\max}$, such criteria would often suggest two or three factors, when in fact, none are present. Moreover, checking the stationarity of the PCA residuals using the Dickey-Fuller tests may spuriously favour the stationarity hypothesis. This may mislead a researcher to conclude that all the non-stationarity in the data is captured by a few common factors, which are consistently estimated by the PCA.

To detect these potential problems, we propose to always look at the time series plots of the extracted factors. Their resemblance to cosine waves should raise the alarm. A more formal detection strategy would compare the factor estimates obtained from the data in levels and in first differences. We derive a theory-based threshold for the sum of the squared canonical correlations between the spaces spanned by the differenced factors extracted from the level data and the factors extracted from the differenced data. The sum of the squared canonical correlations going below the threshold signals a problem that necessitates a further analysis.

Mis-interpreting spurious factors as common shocks driving economic data may be devastating for structural economic analysis. Less obvious, using such factors in forecasting exercises may lead to forecast sub-optimality.[9] This can be clearly seen in the extreme situation where all data are independent random walks. For such data, optimal forecasts equal the most recent observations. They would be different from the forecasts based on the cosine waves that represent the spurious factors asymptotically.

In conclusion, we would like to stress that our critique does not apply to *all* PC analysis in economics. Most of this analysis is careful with respect to the assumptions made and is, therefore, immune to our critique. Furthermore, we would be very disappointed if some readers conclude from our analysis that there are no common economic forces affecting various economic data series. In the literature, there is ample evidence that such common forces are often present, which gives an indisputable value to careful economic research based on high-dimensional factor analysis.

7 Appendix

This Appendix uses Monte Carlo (MC) analysis to address three questions. First, what is the "number of factors" in factorless persistent data detected by information criteria proposed in Bai (2004)? Second, how oversized are the standard Dickey-Fuller tests of unit root in the "idiosyncratic" component of the factorless persistent data? Third, how conservative is the threshold for the squared canonical correlations proposed in Section 5?

7.1 The "number of factors" MC

We simulate data on N i.i.d. Gaussian random walks of length T, where the (N,T)-pairs correspond to the dimensions of four actual datasets described in Table 1. The number of MC replications is set to 10,000.

Table 2 reports the obtained MC distributions of $\hat{k} = \hat{k}_1$, the estimate of the number of factors produced by IPC with penalty $p_1(N,T)$. Results for $\hat{k}_2$ and $\hat{k}_3$ are similar and not reported. The columns of the table correspond to different choices of $k_{\max} = 6, ..., 15$. The entries of the table are the empirical probabilities (in percent rounded to the nearest integer) of observing a particular value of $\hat{k}$, which is given in the first column.

[9] We are grateful to James Stock for pointing out this fact to us.

(N,T)	Content	Source
$(60,52)$	US annual industry-level employment.	Bai (2004)
$(243,83)$	European quarterly macroeconomic data	Boivin et al. (2009)
$(128,710)$	Current version of FRED-MD monthly macroeconomic dataset	McCracken and Ng (2015)
$(58,220)$	US quarterly "real activity dataset"	Stock and Watson (2016)

Table 1: The dimensionalities of datasets used in the analysis below.

We see that the MC distributions of $\hat{k}$ concentrate at $\hat{k}=2$ or $\hat{k}=3$ for most of the settings. For example, when $k_{\max}=10$ and $(N,T)=(60,52)$, the MC probability of observing $\hat{k}=3$ equals 91%. For the same $k_{\max}$ and $(N,T)=(243,83)$, this probability becomes 100%. For $(N,T)=(128,710)$ and $(N,T)=(58,220)$, the mode of the MC distributions of $\hat{k}$ shifts to $\hat{k}=1$ (probability 100%) and $\hat{k}=2$ (probability 86%), respectively. Overall we see that, for empirically relevant data sizes, IPC criteria would typically estimate a small non-zero number of factors in the factorless persistent data.

7.2 Dickey-Fuller tests for the "idiosyncratic" series

One of the arguments in favour of doing factor analysis in levels discussed in Banerjee et al. (2017) is that the estimated idiosyncratic part of typical macroeconomic data looks stationary in applications. The hypothesis of a unit root in the estimated idiosyncratic components can often be easily rejected. As pointed out in the Introduction, such a rejection may be due to the standard unit root tests being seriously oversized.

To support this claim, we perform the following MC experiment. For each of the empirically relevant sample sizes (N,T) reported in Table 1, we simulate N i.i.d. Gaussian random walks of length T. Then, we extract $0,1,...,6$ "factors" from the simulated data and run the Dickey-Fuller regression (intercept only) on the remaining "idiosyncratic" series.

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

k_{max}	6	7	8	9	10	11	12	13	14	15
$(N,T) = (60,52)$ as in Bai (2004)										
$\hat{k}=0$	0	0	0	0	0	0	0	0	0	0
$\hat{k}=1$	0	0	0	0	0	0	0	0	0	0
$\hat{k}=2$	96	75	43	15	4	1	0	0	0	0
$\hat{k}=3$	4	25	57	84	91	79	54	27	9	2
$\hat{k}=4$	0	0	0	1	5	20	46	72	85	77
$\hat{k}=5$	0	0	0	0	0	0	0	1	6	21
$(N,T) = (243,83)$ as in Boivin et al. (2009)										
$\hat{k}=0$	0	0	0	0	0	0	0	0	0	0
$\hat{k}=1$	0	0	0	0	0	0	0	0	0	0
$\hat{k}=2$	99	76	23	2	0	0	0	0	0	0
$\hat{k}=3$	1	24	77	98	100	98	84	50	19	4
$\hat{k}=4$	0	0	0	0	0	2	16	50	81	96
$(N,T) = (128,710)$ as in FRED-MD dataset, McCracken and Ng (2015)										
$\hat{k}=0$	0	0	0	0	0	0	0	0	0	0
$\hat{k}=1$	100	100	100	100	100	97	88	67	41	20
$\hat{k}=2$	0	0	0	0	0	3	12	33	59	80
$(N,T) = (58,220)$ as in "real activity dataset", Stock and Watson (2016)										
$\hat{k}=0$	0	0	0	0	0	0	0	0	0	0
$\hat{k}=1$	98	87	62	33	14	4	1	0	0	0
$\hat{k}=2$	2	13	38	67	86	96	98	95	84	67
$\hat{k}=3$	0	0	0	0	0	0	1	5	16	33

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

becoming close to 100%.

	Number of "factors" extracted						
(N,T)	0	1	2	3	4	5	6
$(60,52)$	5	18.9	43.0	68.4	87.0	95.9	99.1
$(243,83)$	5	17.9	40.7	65.3	84.5	94.8	98.6
$(128,710)$	5	18.1	40.7	65.0	83.8	93.9	97.9
$(58,220)$	5	18.7	41.1	66.2	85.1	94.7	98.5

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

7.3 Threshold for squared canonical correlations

In this subsection, we perform an MC analysis to assess the quality of the choice $a = 2$ and $b = 1$ in the threshold $a\left(\widehat{\mathcal{S}}_r + b\right)$ proposed in Section 5. We would like to know how often this choice leads to the false and correct alarms for the spurious factor analysis.

To investigate the rate of the false alarms, we use the MC setting in Bai (2004), equations (21-23). That is, the data contain two genuine strong factors represented by two independent random walks with $N(0,1)$ increments. The entries of the loadings matrix Λ are i.i.d. $N(0,1)$. The idiosyncratic terms are generated by ARMA(1,1) so that

$$e_{it} = 0.5e_{i,t-1} + v_{it} + 0.5v_{i,t-1},$$

where v_{it} are i.i.d. $N(0,1)$. The factors, loadings, and idiosyncratic terms are mutually independent. Thirteen different choices of (N,T)-pairs are the same as in Bai (2004).

To investigate the rate of the correct alarms, we simulate i.i.d. Gaussian random walk data with the same dimensionality. The extracted factors for such data must be spurious, which ideally should be detected by the proposed method.

Table 4 reports the MC mean of R_r (with $r = 2$); the 1, 5, 95, and 99-th percentiles of the MC distribution of the ratio of $\delta^2_{NT}(r - R_r)$ to $\widehat{\mathcal{S}}_r$; and the percent of MC cases where $\delta^2_{NT}(r - R_r)$ is larger than $2\left(\widehat{\mathcal{S}}_r + 1\right)$, so that the spurious PCA alarm is triggered. The upper panel of the table correspond to MC settings where the alarm is undesirable, whereas the lower panel correspond to the MC settings where the alarm is wanted.

For the upper panel, the MC average value of R_2 is extremely close to 2 for all

N	T	MC mean of R_r	Percentiles of $\frac{\delta^2_{NT}(r-R_r)}{\hat{\mathcal{S}}_r}$				$\frac{\delta^2_{NT}(r-R_r)}{2(\hat{\mathcal{S}}_r+1)} > 1$
			1	5	95	99	(% of MC cases)
			Two genuine factors in the data, $r = 2$.				
100	40	1.9969	0.90	0.95	1.07	1.11	0
100	60	1.9985	0.97	0.99	1.06	1.09	0
200	60	1.9990	0.96	0.98	1.03	1.04	0
500	60	1.9993	0.94	0.96	1.01	1.02	0
1000	60	1.9994	0.93	0.95	1.00	1.01	0
40	100	1.9984	1.04	1.05	1.15	1.19	0
60	100	1.9989	1.02	1.03	1.09	1.12	0
60	200	1.9996	1.03	1.04	1.08	1.10	0
60	500	1.9999	1.04	1.04	1.07	1.08	0
60	1000	1.9999	1.04	1.04	1.07	1.08	0
50	50	1.9964	0.97	1.00	1.14	1.21	0
100	100	1.9993	1.01	1.02	1.06	1.07	0
200	200	1.9999	1.01	1.01	1.02	1.03	0
			Spurious factors, $r = 2$. The data are i.i.d. random walks.				
100	40	0.3731	2.42	2.96	8.06	10.0	99.83
100	60	0.3185	2.45	2.97	7.28	8.80	99.89
200	60	0.2383	3.42	4.04	8.64	10.0	100
500	60	0.1699	6.57	7.50	14.1	16.1	100
1000	60	0.1375	12.0	13.5	23.8	26.5	100
40	100	0.3734	2.55	3.20	8.59	10.6	99.94
60	100	0.3172	2.56	3.08	7.49	9.07	99.92
60	200	0.2389	3.63	4.24	9.15	10.8	100
60	500	0.1716	7.10	8.05	15.1	17.1	100
60	1000	0.1368	12.8	14.5	25.2	28.4	100
50	50	0.4454	1.82	2.37	7.26	9.16	98.23
100	100	0.2556	2.65	3.15	6.92	8.16	99.98
200	200	0.1378	3.38	3.80	6.52	7.34	100

Table 4: Results of the MC experiment that compares factors extracted from the differenced data with differenced factors extracted from the level data. Upper panel: MC setting is as in Bai (2004, eqs. (21-23)). Lower panel: i.i.d. random walk data. The number of MC replications is 10,000. Third column: MC mean of the sum of the two squared sample canonical correlations. Columns 4-7: Percentiles of the MC distribution of the ratio $\delta^2_{NT}(r - R_r)/\hat{\mathcal{S}}_r$. Last column: percent of MC replications where the proposed threshold is violated.

considered combinations of N and T. This accords well with our Lemma 7, which shows that $2 - R_2$ must converge to zero at the rate $\max\{N^{-2}, T^{-2}\}$.

The reported percentiles of the ratio of $\delta_{NT}^2(r - R_r)$ to $\widehat{\mathcal{S}}_r$ accord well with another corollary of Lemma 7 that the difference $\delta_{NT}^2(r - R_r) - \widehat{\mathcal{S}}_r$ is of asymptotic order $\max\{N^{-1}, T^{-1}\}$. All the percentiles are close to unity, although for relatively small N, $\widehat{\mathcal{S}}_r$ tends to be slightly smaller than $\delta_{NT}^2(r - R_r)$. Finally, as shown in the last column of the table, none of our MC replications resulted in the false alarm for spurious PCA.

For the lower panel, the MC average value of R_2 is an order of magnitude smaller than 2. The distribution of the ratio $\delta_{NT}^2(r - R_r)$ to $\widehat{\mathcal{S}}_r$ has all mass substantially above unity. Practically all MC replications resulted in the spurious factor analysis alarm triggered correctly.

References

[1] Andreou, E., Gagliardini, P., Ghysels, E., and Rubin, M. (2019) "Inference in Group Factor Models With an Application to Mixed-Frequency Data," *Econometrica* 87, 1267–1305.

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

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

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

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

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

[7] Böttcher, A. and Silbermann, B. (1999) *Introduction to Large Truncated Toeplitz Matrices*, Springer.

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

[9] Corielli, F. and Marcellino, M. (2006) "Factor based index tracking," *Journal of Banking & Finance* 30, 2215–2233.

[10] Eickmeier, S. (2009) "Comovements and Heterogeneity in the Euro Area Analyzed in a Non-stationary Dynamic Factor Model," *Journal of Applied Econometrics* 24, 933–959.

[11] Elliott, G. (1998) "The Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica* 66, 149–158.

[12] Elliott, G. (1999) "Efficient Tests for a Unit Root When the Initial Observation is Drawn from its Unconditional Distribution," *International Economic Review* 40, 767–783.

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

[14] Ghate, C. and S. Wright (2012) "The "V-factor": Distribution, timing and correlates of the great Indian growth turnaround," *Journal of Development Economics* 99, 58–67.

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

[16] Karatzas, I., and Shreve, S. E. (1991) *Brownian Motion and Stochastic Calculus*, Second edition, Springer.

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

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

[19] Moon, H. R. and Perron, B. (2007) "An Empirical Analysis of Nonstationarity in a Panel of Interest Rates with Factors," *Journal of Applied Econometrics* 22, 383–400.

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

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

[22] Phillips, P.C.B. (1988) "Regression Theory for Near-integrated Time Series," *Econometrica* 56, 1021–1043.

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

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

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

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

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

[28] Vershynin, R. (2012) "Introduction to the non-asymptotic analysis of random matrices," in Eldar, Y. and Kutyniok, G. *Compressed Sensing, Theory and Applications.* Cambridge University Press, 210–268.

[29] von Borstel, J., Eickmeier, S., and Krippner, L. (2016) "The interest rate pass-through in the euro area during the sovereign debt crisis," *Journal of International Money and Finance* 68, 386–402.

[30] Wang, Y. C. and Wu J. L. (2015) "Fundamentals and Exchange Rate Prediction Revisited," *Journal of Money, Credit and Banking* 47, 1651–1671.

[31] West, K. D. and Wong, K. F. (2014) "A factor model for co-movements of commodity prices," *Journal of International Money and Finance* 42, 289–309.

Supplementary Material for "Spurious Factor Analysis."

Alexei Onatski* and Chen Wang†
*Faculty of Economics, University of Cambridge.
†Department of Statistics and Actuarial Science,
University of Hong Kong.

January 13, 2020

Abstract

This note contains supplementary material for Onatski and Wang (2019) (OW in what follows). It is lined up with sections in the main text to make it easy to locate the required proofs.

Contents

1 Introduction

1.1 There is no supplementary material for this section of OW.

2 Basic setup and main results

2.1 Proof of Theorem OW1

Consider the multivariate Beveridge-Nelson decomposition of the demeaned X_t

$$X_t - \bar{X} = \Psi(1)\left(\xi_t - \bar{\xi}\right) + \Psi^*(L)\left(\varepsilon_t - \bar{\varepsilon}\right),$$

where $\Psi^*(L) = \sum_{k=0}^{\infty} \Psi_k^* L^k$ with $\Psi_k^* = -\sum_{i=k+1}^{\infty} \Psi_i$, and $\xi_t = \sum_{j=1}^{t} \varepsilon_j$. In matrix notations,

$$XM = \Psi(1)\,\varepsilon UM + \Psi^*(L)\,\varepsilon M, \tag{1}$$

where M is the projection matrix on the space orthogonal to the T-dimensional vector of ones, ε is the $N_\varepsilon \times T$ matrix with columns ε_t, and U is the upper triangular matrix with ones above and on the main diagonal.

Recall that $\hat{\Sigma}$ is the sample covariance matrix of the demeaned data XM. Let $\tilde{\Sigma}$ be the sample covariance of the I(1) term in the Beveridge-Nelson decomposition (1) of XM, that is

$$\tilde{\Sigma} = MU'\varepsilon' W \varepsilon UM/N, \tag{2}$$

where $W = \Psi(1)' \Psi(1)$. Denote the eigenvalues of $\tilde{\Sigma}$ as $\tilde{\lambda}_1 \geq ... \geq \tilde{\lambda}_T$ and corresponding eigenvectors as $\tilde{F}_1, ..., \tilde{F}_T$. Since variation of I(1) series dominates that of I(0) series, it is reasonable to expect that $\hat{\Sigma}$ and $\tilde{\Sigma}$ are close in some sense. Therefore our proof strategy is, first, show that statements (i)-(iii) of Theorem OW1 hold when $\hat{\lambda}_k, \hat{F}_k, \hat{\Sigma}$ are replaced by $\tilde{\lambda}_k, \tilde{F}_k, \tilde{\Sigma}$ and then, prove that replacing back "tildes" by "hats" does not affect the theorem's validity.

2.1.1 Proof of Theorem OW1 for $\tilde{\lambda}_k, \tilde{F}_k, \tilde{\Sigma}$

First, we will prove the theorem for $k = 1$. Then, we handle general k by mathematical induction. The following lemma is established in Subsection 2.2 of this note.

Lemma 1 *Matrix MU' has the following singular value decomposition $MU' = \sum_{q=1}^{T} \sigma_q w_q v_q'$, where for $q < T$, $\sigma_q = (2\sin(\pi q/(2T)))^{-1}$ and the s-th coordinates of vectors w_q and v_q equal*

$$w_{qs} = -\sqrt{2/T}\cos((s-1/2)\pi q/T) \;\; \text{and} \;\; v_{qs} = \sqrt{2/T}\sin((s-1)\pi q/T).$$

For $q = T$ we have $\sigma_T = 0$, $w_T = l_T/\sqrt{T}$, and $v_T = e_1$, where l_T is the T-dimensional vector of ones and e_1 is the first coordinate vector of $\mathbb{R}^T$.

Since w_q, $q = 1, ..., T-1$, form an orthonormal basis in the space orthogonal to l_T and $\tilde{F}_1$ belongs to this space, we have a representation

$$\tilde{F}_1 = \sum_{q=1}^{T-1} \alpha_q w_q. \tag{3}$$

Let us show that $\alpha_1^2 \xrightarrow{\text{P}} 1$. This would establish part (i) of the theorem because $(w_1' d_1)^2 \to 1$.

Representation (3) and Lemma 1 yield

$$N\tilde{\lambda}_1 = N\sum_{r,q=1}^{T-1} \alpha_r\alpha_q w_r'\tilde{\Sigma}w_q = \gamma' W \gamma,$$

where $\gamma = \sum_{r=1}^{T-1} \alpha_r \sigma_r \varepsilon v_r$. The idea of the proof consists of, first, showing that the sum in the latter display is dominated by the terms $\alpha_r^2 w_r'\tilde{\Sigma}w_r$, and, then, demonstrating that $w_r'\tilde{\Sigma}w_r$ is quickly decreasing in r so that the maximum of the sum with respect to α's is achieved when α_1^2 is close to unity whereas α_r^2 with $r > 1$ are close to zero.

The following lemma is established in Subsection 2.3 of this note.

Lemma 2 *Suppose assumption A1 of OW holds. Let a, b, c, d and A, B be any deterministic T-dimensional vectors and $N_\varepsilon \times N_\varepsilon$ matrices, respectively. Then*

$$\mathbb{E}\left(a'\varepsilon' A\varepsilon b\right) = a'b \operatorname{tr} A, \;\; \text{and} \tag{4}$$

$$\begin{aligned}&\left|Cov\left(a'\varepsilon' A\varepsilon b, c'\varepsilon' B\varepsilon d\right) - \left(a'c\right)\left(b'd\right)\operatorname{tr}\left(A'B\right) - \left(a'd\right)\left(b'c\right)\operatorname{tr}\left(AB\right)\right| \\ &\quad \le 2\varkappa_4 \sum_{i=1}^{N_\varepsilon}\sum_{t=1}^{T}\left|A_{ii}B_{ii}a_t b_t c_t d_t\right|,\end{aligned} \tag{5}$$

where a_t, b_t, c_t, and d_t are the t-th components of vectors a, b, c, and d.

Corollary 3 *Suppose assumptions A1 and A3 of OW hold. Then, for any positive integers i, j such that $i \leq j \leq T$,*

$$v_i' \varepsilon' W \varepsilon v_j = \operatorname{tr} W \left(\delta_{ij} + o_{\mathrm{P}}(1) \right), \tag{6}$$

where δ_{ij} is the Kronecker delta, and

$$\begin{aligned} \sum\nolimits_{r=i}^{j} \sigma_r^2 v_r' \varepsilon' W \varepsilon v_r &= \operatorname{tr} W \sum\nolimits_{r=i}^{j} \sigma_r^2 \left(1 + o_{\mathrm{P}}(1)\right) \\ &= \operatorname{tr} W \sum\nolimits_{r=i}^{j} \sigma_r^2 + \operatorname{tr} W o_{\mathrm{P}}(T^2). \end{aligned} \tag{7}$$

Proof: Since W is positive semi-definite, we have

$$\left(\|W\| / \operatorname{tr} W \right)^2 \leq \operatorname{tr}\left(W^2\right) / \left(\operatorname{tr} W\right)^2 \leq \|W\| / \operatorname{tr} W.$$

Further, $\|W\| = \|\Omega\|$ and $\operatorname{tr} W = \operatorname{tr} \Omega$. Therefore, assumption A3 of OW is equivalent to the requirement

$$\operatorname{tr}\left(W^2\right) = o(1) \left(\operatorname{tr} W\right)^2. \tag{8}$$

The first and second equalities of the corollary follows from Lemma 2, Chebyshev's inequality, and (8). The last equality follows from the fact that $\sum\nolimits_{r=1}^{T-1} \sigma_r^2 = O(T^2)$. $\square$

Let K be a fixed non-negative integer. Consider a decomposition $\gamma = \gamma_1 + \gamma_2$, where

$$\gamma_1 = \sum\nolimits_{r=1}^{K} \alpha_r \sigma_r \varepsilon v_r \text{ and } \gamma_2 = \sum\nolimits_{r=K+1}^{T-1} \alpha_r \sigma_r \varepsilon v_r.$$

We have, by the Cauchy-Schwarz inequality,

$$\gamma' W \gamma \leq \left(\left(\gamma_1' W \gamma_1\right)^{1/2} + \left(\gamma_2' W \gamma_2\right)^{1/2} \right)^2. \tag{9}$$

Since K is fixed and $\sigma_r^2 = O\left(T^2\right)$, equation (6) of Corollary 3 yields

$$\begin{aligned} \gamma_1' W \gamma_1 &= \operatorname{tr} W \sum\nolimits_{r=1}^{K} \alpha_r^2 \sigma_r^2 + \operatorname{tr} W \, o_{\mathrm{P}}\left(T^2\right) \\ &\leq \operatorname{tr} W \sum\nolimits_{r=1}^{T-1} \alpha_r^2 \sigma_r^2 + \operatorname{tr} W \, o_{\mathrm{P}}\left(T^2\right). \end{aligned} \tag{10}$$

Further, we have

$$\gamma_2' W \gamma_2 = \sum\nolimits_{j=1}^{N_\varepsilon} \left(\sum\nolimits_{r=K+1}^{T-1} \alpha_r \sigma_r \left[W^{1/2} \varepsilon v_r \right]_j \right)^2,$$

where $\left[W^{1/2} \varepsilon v_r\right]_j$ is the j-th component of vector $W^{1/2} \varepsilon v_r$. By the Cauchy-Schwarz inequality,

$$\left(\sum\nolimits_{r=K+1}^{T-1} \alpha_r \sigma_r \left[W^{1/2} \varepsilon v_r \right]_j \right)^2 \leq \sum\nolimits_{r=K+1}^{T-1} \alpha_r^2 \sum\nolimits_{r=K+1}^{T-1} \left(\sigma_r \left[W^{1/2} \varepsilon v_r \right]_j \right)^2.$$

But $\sum_{r=1}^{T}\alpha_r^2=1$. Therefore,

$$\gamma_2'W\gamma_2\leq\sum_{j=1}^{N_\varepsilon}\sum_{r=K+1}^{T-1}\left(\sigma_r\left[W^{1/2}\varepsilon v_r\right]_j\right)^2=\sum_{r=K+1}^{T-1}\sigma_r^2v_r'\varepsilon'W\varepsilon v_r.$$

This inequality and equations (7) of Corollary 3 yield

$$\gamma_2'W\gamma_2\leq\operatorname{tr}W\sum_{r=K+1}^{T-1}\sigma_r^2+\operatorname{tr}W o_{\mathrm{P}}(T^2). \tag{11}$$

Using (10) and (11) in (9), we obtain

$$\begin{aligned}\gamma'W\gamma\quad&\leq\quad\operatorname{tr}W\left(\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2+\sum_{r=K+1}^{T-1}\sigma_r^2\right.\\&\qquad\left.+2\left(\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2\sum_{r=K+1}^{T-1}\sigma_r^2\right)^{1/2}+o_{\mathrm{P}}\left(T^2\right)\right).\end{aligned} \tag{12}$$

Note that

$$\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2\leq\sigma_1^2=\left(4\sin^2\left(\pi/\left(2T\right)\right)\right)^{-1}.$$

Since $\sin x\geq 2x/\pi$ for $x\in[0,\pi/2]$, we have $\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2\leq T^2/4$.

Similarly,

$$\sum_{r=1}^{T-1}\sigma_r^2=\sum_{r=1}^{T-1}\left(4\sin^2\left(\pi r/\left(2T\right)\right)\right)^{-1}\leq(T^2/4)\sum_{r=1}^{T-1}r^{-2}.$$

Let us choose K so that $\sum_{r=K+1}^{T-1}\sigma_r^2\leq\delta^2T^2/4$, where $\delta<1$ is an arbitrarily small positive number. Then, from (12),

$$\begin{aligned}\gamma'W\gamma\quad&\leq\quad\operatorname{tr}W\left(\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2+\left(\frac{1}{4}\delta^2+\frac{1}{2}\delta\right)T^2+o_{\mathrm{P}}\left(T^2\right)\right)\\&\leq\quad\operatorname{tr}W\left(\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2+\delta T^2+o_{\mathrm{P}}\left(T^2\right)\right).\end{aligned}$$

Since δ can be made arbitrarily small,

$$\gamma'W\gamma\leq\operatorname{tr}W\left(\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2+o_{\mathrm{P}}\left(T^2\right)\right).$$

Now recall that $\gamma'W\gamma=N\tilde{\lambda}_1$. Since

$$\sum_{r=1}^{T-1}\alpha_r^2\sigma_r^2\operatorname{tr}W\leq\alpha_1^2\sigma_1^2\operatorname{tr}W+(1-\alpha_1^2)\sigma_2^2\operatorname{tr}W,$$

we have

$$N\tilde{\lambda}_1\leq\alpha_1^2\sigma_1^2\operatorname{tr}W+(1-\alpha_1^2)\sigma_2^2\operatorname{tr}W+o_{\mathrm{P}}\left(1\right)T^2\operatorname{tr}W. \tag{13}$$

On the other hand, $N\tilde{\lambda}_1$ must be no smaller than $Nw_1'\tilde{\Sigma}w_1 = \sigma_1^2 v_1'\varepsilon' W \varepsilon v_1$. By Corollary 3,

$$\sigma_1^2 v_1'\varepsilon' W \varepsilon v_1 = \sigma_1^2 \operatorname{tr} W + o_{\mathrm{P}}(1)\, T^2 \operatorname{tr} W. \tag{14}$$

Therefore,

$$N\tilde{\lambda}_1 \geq \sigma_1^2 \operatorname{tr} W + o_{\mathrm{P}}(1)\, T^2 \operatorname{tr} W. \tag{15}$$

Combining this with (13), we obtain

$$\sigma_1^2 \operatorname{tr} W + o_{\mathrm{P}}(1)\, T^2 \operatorname{tr} W \leq \alpha_1^2 \sigma_1^2 \operatorname{tr} W + (1-\alpha_1^2)\sigma_2^2 \operatorname{tr} W + o_{\mathrm{P}}(1)\, T^2 \operatorname{tr} W,$$

which implies

$$1-\alpha_1^2 \leq o_{\mathrm{P}}(1)\, T^2 / \left(\sigma_1^2 - \sigma_2^2\right) = o_{\mathrm{P}}(1). \tag{16}$$

Hence,

$$\alpha_1^2 = \left(\tilde{F}_1' w_1\right)^2 \xrightarrow{\mathrm{P}} 1, \tag{17}$$

which completes our proof of statement (i) for $k=1$.

To establish (ii), note that inequalities (13) and (15) yield

$$\left| N\tilde{\lambda}_1 - \sigma_1^2 \operatorname{tr} W \right| \leq \left|1-\alpha_1^2\right| \left(\sigma_1^2 + \sigma_2^2\right) \operatorname{tr} W + o_{\mathrm{P}}(1)\, T^2 \operatorname{tr} W.$$

Combining this with the facts that $\alpha_1^2 = 1 + o_{\mathrm{P}}(1)$ and $\sigma_1^2 = T^2/\pi^2 + o(T^2)$, we obtain

$$\tilde{\lambda}_1 = \frac{T^2 \operatorname{tr} W}{\pi^2 N}\left(1 + o_{\mathrm{P}}(1)\right) = \frac{T^2 \operatorname{tr} \Omega}{\pi^2 N}\left(1 + o_{\mathrm{P}}(1)\right), \tag{18}$$

as claimed by statement (ii).

Further, by Lemma 1,

$$N \operatorname{tr} \tilde{\Sigma} = \operatorname{tr}\left(\sum\nolimits_{r=1}^{T} \sigma_r w_r v_r' \varepsilon' W \varepsilon \sum\nolimits_{q=1}^{T} \sigma_q v_q w_q'\right) = \sum\nolimits_{r=1}^{T} \sigma_r^2 v_r' \varepsilon' W \varepsilon v_r,$$

where the last equality follows from the orthonormality of the basis $\{w_r, r = 1, ..., T\}$. Hence, by Corollary 3,

$$N \operatorname{tr} \tilde{\Sigma} = \operatorname{tr} W \sum_{r=1}^{T} \sigma_r^2 \left(1 + o_{\mathrm{P}}(1)\right). \tag{19}$$

On the other hand, for any fixed K,

$$\sum_{r=1}^{K} \sigma_r^2 / T^2 = \sum_{r=1}^{K} 1/\left(4T^2 \sin^2\left(\pi r/(2T)\right)\right) \to \sum_{r=1}^{K} 1/(\pi r)^2$$

as $T \to \infty$. Furthermore, $\sum\nolimits_{r=K+1}^{T} \sigma_r^2/T^2$ can be made arbitrarily small by choosing sufficiently large K. Hence, the Euler formula $\sum_{r=1}^{\infty} r^{-2} = \pi^2/6$ yields $\sum_{r=1}^{T} \sigma_r^2/T^2 \to 1/6$.

The latter convergence and (19) give us

$$\operatorname{tr}\tilde{\Sigma} = \frac{T^2}{6N} \operatorname{tr} W \left(1 + o_{\mathrm{P}}\left(1\right)\right) = \frac{T^2}{6N} \operatorname{tr} \Omega \left(1 + o_{\mathrm{P}}\left(1\right)\right).$$

Combining this with (18), we obtain

$$\tilde{\lambda}_1 / \operatorname{tr}\tilde{\Sigma} = \left(6/\pi^2\right)\left(1 + o_{\mathrm{P}}(1)\right), \tag{20}$$

which concludes the proof of the theorem for $k = 1$.

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

$$N\tilde{F}_m' \tilde{\Sigma} \tilde{F}_m = \sum_{q,r=m}^{T-1} \alpha_q \alpha_r \sigma_q \sigma_r v_q' \varepsilon' W \varepsilon v_r + o_{\mathrm{P}}\left(T^2\right) \operatorname{tr} W. \tag{21}$$

To see that (21) holds, it is sufficient to establish equalities $N\alpha_j w_j' \tilde{\Sigma} \sum_{r=m}^{T-1} \alpha_r w_r = o_{\mathrm{P}}\left(T^2\right) \operatorname{tr} W$ for any $j < m$, and equalities $N\alpha_j \alpha_r w_j' \tilde{\Sigma} w_r = o_{\mathrm{P}}\left(T^2\right) \operatorname{tr} W$ for any $j, r < m$. Such equalities easily follow from the facts that $\alpha_j = o_{\mathrm{P}}\left(1\right)$ for all $j < m$ and $N\left\|\tilde{\Sigma}\right\| = N\tilde{\lambda}_1 = \left(T^2/\pi^2\right) \operatorname{tr} W \left(1 + o_{\mathrm{P}}(1)\right)$.

In addition to (21), we must have

$$N\sum_{i=1}^{m-1} \tilde{\lambda}_i + N\tilde{F}_m' \tilde{\Sigma} \tilde{F}_m \geq \sum_{i=1}^{m} w_i' M U' \varepsilon' W \varepsilon U M w_i = \left(\sum_{i=1}^{m} \sigma_i^2 + o_{\mathrm{P}}\left(T^2\right)\right) \operatorname{tr} W,$$

where the latter equality is obtained similarly to (14). Combining the above two displays, and using the induction hypothesis, this time regarding the validity of the identities $N\tilde{\lambda}_i = \left(\sigma_i^2 + o_{\mathrm{P}}\left(T^2\right)\right) \operatorname{tr} W$ for all $i < m$, we obtain

$$\sum_{q,r=m}^{T-1} \alpha_q \alpha_r \sigma_q \sigma_r v_q' \varepsilon' W \varepsilon v_r \geq \sigma_m^2 \operatorname{tr} W + o_{\mathrm{P}}\left(T^2\right) \operatorname{tr} W. \tag{22}$$

Statements (i), (ii), and (iii) for $k = m$ now follow by arguments that are very similar to those used above for the case $k = 1$.

That is, we represent the sum on the left hand side of (22) in the form $\gamma' W \gamma$, where $\gamma = \sum_{r=m}^{T-1} \alpha_r \sigma_r \varepsilon v_r$. Then proceed along the lines of the above proof to obtain an upper bound on $\gamma' W \gamma$, similar to the right hand side of (13). Then, combining this upper bound with the lower bound (22), we prove the convergence $\alpha_m^2 \xrightarrow{\mathrm{P}} 1$. Finally, we proceed to establishing parts (ii) and (iii) using part (i). We omit details to save space.

2.1.2 Proof of Theorem OW1 for $\hat{\lambda}_k, \hat{F}_k, \hat{\Sigma}$

We need to show that the theorem's validity for $\tilde{F}_k$, $\tilde{\lambda}_k$ and $\tilde{\Sigma}$ implies its validity for $\hat{F}_k$, $\hat{\lambda}_k$ and $\hat{\Sigma}$. By standard perturbation theory (e.g. Kato (1980), ch.2), such an implication for statements (i) and (ii) would follow if we are able to show that $\left\|\hat{\Sigma} - \tilde{\Sigma}\right\| = \frac{T^2}{N} \operatorname{tr} W\, o_{\mathrm{P}}(1)$. That is, the norm of $\hat{\Sigma} - \tilde{\Sigma}$ is asymptotically dominated by the sizes of the gaps between adjacent eigenvalues, $\tilde{\lambda}_k - \tilde{\lambda}_{k+1}$ and $\tilde{\lambda}_{k-1} - \tilde{\lambda}_k$. The Beveridge-Nelson decomposition (1) implies that it is sufficient to show that $\|\Psi^*(L)\varepsilon M\|^2 = T^2 \operatorname{tr} W\, o_{\mathrm{P}}(1)$. To establish this equality, we need the following lemma.

Lemma 4 *Suppose that assumption A1 of OW holds. Let $Z_t = \Pi(L)\varepsilon_t$ and $Z = [Z_1, ..., Z_T]$, where $\Pi(L) = \sum_{k=0}^{\infty} \Pi_k L^k$ is an $N \times N_\varepsilon$ matrix lag polynomial that may depend on N, N_ε, and T. If $\sum_{k=0}^{T} \|\Pi_k\| = O(N^\alpha)$ and $T \sum_{k=T+1}^{\infty} \|\Pi_k\|_F^2 = O\left(N_\varepsilon N^{2\alpha}\right)$ for an $\alpha \geq 0$, where $\|\cdot\|_F$ denotes the Frobenius norm, then*

$$\|Z\| = O_{\mathrm{P}}\left(T^{1/2}N^\alpha + N_\varepsilon^{1/2}N^\alpha\right). \tag{23}$$

Proof: This is a modification of Proposition 1 from Onatski (2015), where a proportional asymptotic regime with N/T converging to a nonzero constant is considered. The triangle inequality yields

$$\|Z\| \leq \sum_{k=0}^{T} \|\Pi_k\| \, \|\varepsilon_{-k}\| + \|r_T\|,$$

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

$$\|Z\| \leq O_{\mathrm{P}}\left(T^{1/2} + N_\varepsilon^{1/2}\right) \sum_{k=0}^{T} \|\Pi_k\| + \|r_T\| = O_{\mathrm{P}}\left(T^{1/2}N^\alpha + N_\varepsilon^{1/2}N^\alpha\right) + \|r_T\|. \tag{24}$$

On the other hand,

$$\begin{aligned} \mathbb{E}\|r_T\|^2 &\leq \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left[(r_T)_{it}^2\right] = \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=T+1}^{\infty}\sum_{s=1}^{N_\varepsilon} (\Pi_k)_{is}\, \varepsilon_{s,t-k}\right]^2 \\ &\leq T\sum_{k=T+1}^{\infty} \|\Pi_k\|_F^2 = O\left(N_\varepsilon N^{2\alpha}\right). \end{aligned}$$

Hence, $\|r_T\| = O_{\mathrm{P}}\left(N_\varepsilon^{1/2}N^\alpha\right)$. Combining this with (24) yields (23). □

Remark 5 *The lemma holds under following simple but stronger assumptions: $\sum_{k=0}^{\infty} \|\Pi_k\| = O(N^\alpha)$ and $\sum_{k=0}^{\infty} k\|\Pi_k\|^2 = O\left(N^{2\alpha}\right)$. This follows from the inequalities $\|\Pi_k\|_F^2 \leq \min\{N, N_\varepsilon\}\|\Pi_k\|^2$ and $T\sum_{k=T+1}^{\infty} \|\Pi_k\|^2 \leq \sum_{k=0}^{\infty} k\|\Pi_k\|^2$.*

By definition of Ψ_k^*, we have

$$\sum_{k=0}^{\infty} \|\Psi_k^*\| \leq \sum_{k=0}^{\infty} k\|\Psi_k\| = O(N^\alpha),$$

where the latter equality holds by A2. Further,

$$k\left\|\Psi_k^*\right\| \leq k \sum\nolimits_{j=k+1}^{\infty}\left\|\Psi_j\right\| \leq \sum\nolimits_{j=k+1}^{\infty} j\left\|\Psi_j\right\| = O\left(N^{\alpha}\right).$$

Therefore, $\sum\nolimits_{k=0}^{\infty} k\left\|\Psi_k^*\right\|^2 = O\left(N^{2\alpha}\right)$. Hence, by Remark 5,

$$\left\|\Psi^*(L)\varepsilon M\right\|^2 \leq \left\|\Psi^*(L)\varepsilon\right\|^2 = O_{\mathrm{P}}\left(TN^{2\alpha}+N_{\varepsilon}N^{2\alpha}\right). \tag{25}$$

By assumption of Theorem OW1, the right hand side of (25) is dominated by $T^2 \operatorname{tr} W = T^2 \operatorname{tr} \Omega$, which implies that statements (i) and (ii) of the theorem remain valid when $\tilde{\lambda}_k$ and $\tilde{F}_k$ are replaced by $\hat{\lambda}_k$ and $\hat{F}_k$.

To show that (iii) holds for $\hat{\lambda}_k$ and $\hat{\Sigma}$ if it holds for $\tilde{\lambda}_k$ and $\tilde{\Sigma}$, we need to establish asymptotic equivalence of $\operatorname{tr}\hat{\Sigma} = \sum\nolimits_{i=1}^{T}\hat{\lambda}_i$ and $\operatorname{tr}\tilde{\Sigma} = \sum\nolimits_{i=1}^{T}\tilde{\lambda}_i$. From (1),

$$\left|\hat{\lambda}_i^{1/2}-\tilde{\lambda}_i^{1/2}\right| \leq \left\|\Psi^*(L)\varepsilon M\right\|/\sqrt{N} \text{ and } \hat{\lambda}_i = \tilde{\lambda}_i = 0 \text{ for } i > \min\{N,T\}.$$

Therefore, by Minkowski's inequality,

$$\left|\left(\operatorname{tr}\hat{\Sigma}\right)^{1/2}-\left(\operatorname{tr}\tilde{\Sigma}\right)^{1/2}\right| \leq \left\|\Psi^*(L)\varepsilon M\right\| \min\left\{1,\sqrt{T/N}\right\}, \tag{26}$$

and

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

Using (19) and (25), we conclude that

$$\begin{aligned}\left|\operatorname{tr}\hat{\Sigma}-\operatorname{tr}\tilde{\Sigma}\right| \quad \leq \quad & T\min\left\{1,\sqrt{T/N}\right\} O_{\mathrm{P}}\left(T^{1/2}N^{\alpha}+N_{\varepsilon}^{1/2}N^{\alpha}\right)\left(\operatorname{tr}W/N\right)^{1/2} \\ & +O_{\mathrm{P}}\left(TN^{2\alpha}+N_{\varepsilon}N^{2\alpha}\right)\min\{1,T/N\}.\end{aligned}$$

It remains to show that, under the assumption made in (iii), the right hand side of the latter equality is asymptotically dominated by $\operatorname{tr}\tilde{\Sigma}$. By (19), such an asymptotic domination takes place if

$$\left(\operatorname{tr}W\right)^{-1} = o\left(\frac{T^2}{\min\{N,T\}\left(T+N_{\varepsilon}\right)N^{2\alpha}}\right).$$

But this is equivalent to the assumption made in (iii) because $\operatorname{tr}W = \operatorname{tr}\Omega$.

2.2 Proof of Lemma 1

Note that

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

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

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

We have

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

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

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

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

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

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

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

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

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

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

Hence,

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

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

Remark. From (OW7), we see that w_{ks} with $s = 1, ..., T$ and $k < T$ are proportional to the values at $(s-1/2)/T$ of the k-th principal eigenfunction of the covariance operator of the demeaned Wiener process.

2.3 Proof of Lemma 2

We have

$$\begin{aligned}
\mathbb{E}\left(a'\varepsilon' A\varepsilon b\right) &= \sum\nolimits_{t,s=1}^{T}\sum\nolimits_{i,j=1}^{N_\varepsilon}\mathbb{E}\left(a_t\varepsilon_{it}A_{ij}\varepsilon_{js}b_s\right)\\
&= \sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N_\varepsilon}a_tA_{ii}b_t = a'b\operatorname{tr}A.
\end{aligned}$$

Further, denoting the i-th row of ε as $\varepsilon_{i\cdot}$, we have

$$\begin{aligned}
&\mathbb{E}\left(a'\varepsilon' A\varepsilon bc'\varepsilon' B\varepsilon d\right)\\
=\ &\sum\nolimits_{i,j=1}^{N_\varepsilon}\sum\nolimits_{p,l=1}^{N_\varepsilon}\mathbb{E}\left(\varepsilon_{i\cdot}aA_{ij}\varepsilon_{j\cdot}b\varepsilon_{p\cdot}cB_{pl}\varepsilon_{l\cdot}d\right)\\
=\ &\sum\nolimits_{i=1}^{N_\varepsilon}\sum\nolimits_{j\neq i}^{N_\varepsilon}\mathbb{E}\left(\left(\varepsilon_{i\cdot}a\right)\left(\varepsilon_{i\cdot}b\right)\left(\varepsilon_{j\cdot}c\right)\left(\varepsilon_{j\cdot}d\right)A_{ii}B_{jj}\right)\\
&+\sum\nolimits_{i=1}^{N_\varepsilon}\sum\nolimits_{j\neq i}^{N_\varepsilon}\mathbb{E}\left(\left(\varepsilon_{i\cdot}a\right)\left(\varepsilon_{i\cdot}c\right)\left(\varepsilon_{j\cdot}b\right)\left(\varepsilon_{j\cdot}d\right)A_{ij}B_{ij}\right)\\
&+\sum\nolimits_{i=1}^{N_\varepsilon}\sum\nolimits_{j\neq i}^{N_\varepsilon}\mathbb{E}\left(\left(\varepsilon_{i\cdot}a\right)\left(\varepsilon_{i\cdot}d\right)\left(\varepsilon_{j\cdot}b\right)\left(\varepsilon_{j\cdot}c\right)A_{ij}B_{ji}\right)\\
&+\sum\nolimits_{i=1}^{N_\varepsilon}\mathbb{E}\left(\left(\varepsilon_{i\cdot}a\right)\left(\varepsilon_{i\cdot}b\right)\left(\varepsilon_{i\cdot}c\right)\left(\varepsilon_{i\cdot}d\right)A_{ii}B_{ii}\right).
\end{aligned}$$

We have, first,

$$\begin{aligned}
&\sum\nolimits_{i=1}^{N_\varepsilon}\sum\nolimits_{j\neq i}^{N_\varepsilon}\mathbb{E}\left(\left(\varepsilon_{i\cdot}a\right)\left(\varepsilon_{i\cdot}b\right)\left(\varepsilon_{j\cdot}c\right)\left(\varepsilon_{j\cdot}d\right)A_{ii}B_{jj}\right)\\
=\ &\sum\nolimits_{i=1}^{N_\varepsilon}\sum\nolimits_{j\neq i}^{N_\varepsilon}\left(a'b\right)\left(c'd\right)A_{ii}B_{jj}\\
=\ &\left(a'b\right)\left(c'd\right)\left[\left(\operatorname{tr}A\right)\left(\operatorname{tr}B\right)-\sum\nolimits_{i=1}^{N_\varepsilon}A_{ii}B_{ii}\right],
\end{aligned}$$

second,

$$\begin{aligned}
& \sum_{i=1}^{N_\varepsilon} \sum_{j\neq i}^{N_\varepsilon} \mathbb{E}\left(\left(\varepsilon_{i\cdot} a\right)\left(\varepsilon_{i\cdot} c\right)\left(\varepsilon_{j\cdot} b\right)\left(\varepsilon_{j\cdot} d\right) A_{ij} B_{ij} \right) \\
= & \sum_{i=1}^{N_\varepsilon} \sum_{j\neq i}^{N_\varepsilon} \left(a'c\right)\left(b'd\right) A_{ij} B_{ij} \\
= & \left(a'c\right)\left(b'd\right)\left[\operatorname{tr}\left(A'B\right) - \sum_{i=1}^{N_\varepsilon} A_{ii} B_{ii}\right],
\end{aligned}$$

third,

$$\begin{aligned}
& \sum_{i=1}^{N_\varepsilon} \sum_{j\neq i}^{N_\varepsilon} \mathbb{E}\left(\left(\varepsilon_{i\cdot} a\right)\left(\varepsilon_{i\cdot} d\right)\left(\varepsilon_{j\cdot} b\right)\left(\varepsilon_{j\cdot} c\right) A_{ij} B_{ji} \right) \\
= & \sum_{i=1}^{N_\varepsilon} \sum_{j\neq i}^{N_\varepsilon} \left(a'd\right)\left(b'c\right) A_{ij} B_{ji} \\
= & \left(a'd\right)\left(b'c\right)\left[\operatorname{tr}\left(AB\right) - \sum_{i=1}^{N_\varepsilon} A_{ii} B_{ii}\right],
\end{aligned}$$

and finally,

$$\begin{aligned}
& \sum_{i=1}^{N_\varepsilon} \mathbb{E}\left(\left(\varepsilon_{i\cdot} a\right)\left(\varepsilon_{i\cdot} b\right)\left(\varepsilon_{i\cdot} c\right)\left(\varepsilon_{i\cdot} d\right) A_{ii} B_{ii} \right) \\
= & \sum_{i=1}^{N_\varepsilon} \mathbb{E}\left(\sum_{t=1}^{T} \varepsilon_{it} a_t \sum_{t=1}^{T} \varepsilon_{it} b_t \sum_{t=1}^{T} \varepsilon_{it} c_t \sum_{t=1}^{T} \varepsilon_{it} d_t A_{ii} B_{ii} \right) \\
= & \sum_{i=1}^{N_\varepsilon} A_{ii} B_{ii} \left(\sum_{t,s:t\neq s}^{T} a_t b_t c_s d_s + \sum_{t,s:t\neq s}^{T} a_t b_s c_t d_s + \sum_{t,s:t\neq s}^{T} a_t b_s c_s d_t \right. \\
& \left. + \sum_{t=1}^{T} \mathbb{E}\varepsilon_{it}^4 a_t b_t c_t d_t \right) \\
= & \sum_{i=1}^{N_\varepsilon} A_{ii} B_{ii} \left(\left(a'b\right)\left(c'd\right) + \left(a'c\right)\left(b'd\right) + \left(a'd\right)\left(b'c\right) \right. \\
& \left. + \sum_{t=1}^{T} \left(\mathbb{E}\varepsilon_{it}^4 - 3\right) a_t b_t c_t d_t \right)
\end{aligned}$$

Summing up,

$$\begin{aligned}
& \mathbb{E}\left(a'\varepsilon' A \varepsilon b c' \varepsilon' B \varepsilon d\right) \\
= & \left(a'b\right)\left(c'd\right)\left(\operatorname{tr} A\right)\left(\operatorname{tr} B\right) + \left(a'c\right)\left(b'd\right)\operatorname{tr}\left(A'B\right) + \left(a'd\right)\left(b'c\right)\operatorname{tr}\left(AB\right) \\
& + \sum_{i=1}^{N_\varepsilon} A_{ii} B_{ii} \sum_{t=1}^{T} \left(\mathbb{E}\varepsilon_{it}^4 - 3\right) a_t b_t c_t d_t .
\end{aligned}$$

Recall that $\mathbb{E}\left(a'\varepsilon' A \varepsilon b\right) = a'b \operatorname{tr} A$ and $\mathbb{E}\left(c'\varepsilon' B \varepsilon d\right) = c'd \operatorname{tr} B$. These equalities and the last display yield

$$\begin{aligned}
Cov\left(a'\varepsilon' A \varepsilon b, c'\varepsilon' B \varepsilon d\right) = & \left(a'c\right)\left(b'd\right)\operatorname{tr}\left(A'B\right) + \left(a'd\right)\left(b'c\right)\operatorname{tr}\left(AB\right) \\
& + \sum_{i=1}^{N_\varepsilon} A_{ii} B_{ii} \sum_{t=1}^{T} \left(\mathbb{E}\varepsilon_{it}^4 - 3\right) a_t b_t c_t d_t .
\end{aligned}$$

The inequality (5) follows because $\left|\mathbb{E}\varepsilon_{it}^4 - 3\right|$ is bounded by $2\varkappa_4$ uniformly over i and t. Indeed, by assumption A1, $\mathbb{E}\varepsilon_{it}^4 \leq \varkappa_4$, and $\mathbb{E}\varepsilon_{it}^4 - 3 \leq \varkappa_4$. On the other hand, $\mathbb{E}\varepsilon_{it}^4 \geq \left(\mathbb{E}\varepsilon_{it}^2\right)^2 = 1$, and thus, $\varkappa_4 \geq 1$ and $\mathbb{E}\varepsilon_{it}^4 - 3 \geq -2 \geq -2\varkappa_4$.

3 Extensions

3.1 Local level model

3.1.1 Proof of Theorem OW2

Consider the decomposition

$$YM = \omega_T XM + ZM, \tag{27}$$

where $X = [X_1, ..., X_T]$ and $Z = [Z_1, ..., Z_T]$. Note that the principal eigenvalues and eigenvectors of $MX'XM/N$ satisfy Theorem OW1 (i-ii) as long as condition (OW5) of that theorem holds. Statements (i) and (ii) of Theorem OW2 would follow from this fact and the standard perturbation theory (e.g. Kato (1980), ch.2) if we are able to show that $\|MZ'ZM\| = \omega_T^2 T^2 \operatorname{tr} \Omega\, o_{\mathrm{P}}(1)$.

Assumption A4 of OW yields $\sum_{k=0}^{\infty} \|\Pi_k\| = O\left(N^{\beta}\right)$ and $k\|\Pi_k\| \leq \sum_{k=0}^{\infty} (1+k)\|\Pi_k\| = O\left(N^{\beta}\right)$. Therefore, $\sum_{k=0}^{\infty} k\|\Pi_k\|^2 = O\left(N^{2\beta}\right)$ and, as explained in Remark 5, we can apply Lemma 4 to obtain

$$\|ZM\| \leq \|Z\| = O_{\mathrm{P}}\left(T^{1/2}N^{\beta} + N_{\eta}^{1/2}N^{\beta}\right).$$

Hence, Theorem OW2 (i-ii) holds as long as (OW5) and (OW9) hold. But these are the assumptions of Theorem OW2 (i-ii).

For (iii) to hold, it is sufficient that (OW6) is satisfied and $\left|\operatorname{tr}\check{\Sigma} - \omega_T^2 \operatorname{tr}(MX'XM)/N\right|$ is asymptotically dominated by $\omega_T^2 \operatorname{tr}(MX'XM)/N$. Using arguments very similar to those employed in the proof of Theorem OW1 (iii) after equation (25), we see that such an asymptotic domination takes place if $\omega_T^2 T^2 \operatorname{tr}\Omega/N$ asymptotically dominates $N^{2\beta}(T + N_\eta)\min\{1, T/N\}$, which is implied by the assumptions of Theorem OW2 (iii).

3.1.2 The case of the I(1) weight proportional to 1/T

In this subsection we would like to revisit the example given in the main text immediately after the formulation of Theorem OW2. We would like to show how, in that example, the theorem would be violated if ω_T converges to zero faster than allowed by condition (OW9).

Consider X_t and Z_t that follow a pure multivariate random walk and white noise processes, respectively. For simplicity, we assume that X_t and Z_t are independent and Gaussian, and that $T/N = c \in (0, \infty)$. In this setting, A1-A4 are satisfied with $\alpha = \beta = 0$ and $\operatorname{tr}\Omega = N_\varepsilon = N_\eta = N$, so that condition (OW5) of Theorem OW1 is trivially satisfied while condition (OW9) of Theorem OW2 is violated if and only if ω_T converges to zero as fast or faster than $1/T$. We will assume that $\omega_T = w/T$ for some positive fixed w.

Consider a singular value decomposition $\omega_T XM/\sqrt{N} = USV$. Here U and V are orthonormal matrices and S is a diagonal matrix of the singular values of $\omega_T XM/\sqrt{N}$. By Theorem OW1 (i-ii), the k-th row of V becomes asymptotically collinear with a cosine wave (represented by vector d_k), and the k-th diagonal element of S converges to $w/(k\pi)$ as $T \to \infty$. We would like to know whether and how the principal eigenvectors of $\check{\Sigma} = MY'YM/N$ differ from the cosine waves.

From (27), we have

$$U'YMV'/\sqrt{N} = S + U'ZMV'/\sqrt{N}.$$

Note that the last diagonal element of S is zero (because M has deficient rank), and the last row of V belongs to the null space of M. Denote matrix $U'YMV'$ with the last (zero) column removed as $\tilde{Y}$. Similarly, denote matrices S and $U'ZMV'$ with last (zero) columns removed as $\tilde{S}$ and $\tilde{\varepsilon}$, respectively. With this notation,we have $\tilde{Y}/\sqrt{N} = \tilde{S} + \tilde{\varepsilon}/\sqrt{N}$.

By definition, the entries of the k-th principal eigenvector of $\tilde{Y}'\tilde{Y}/N$ equal the scalar products of the k-th principal eigenvector of $\check{\Sigma}$ with the rows of V (which become asymptotically collinear with the cosine waves). Further, since we have assumed that X_t and Z_t are independent Gaussian, $\tilde{\varepsilon}$ has i.i.d. (standard) Gaussian entries.

Now, let $\bar{S}$ be an $N \times (T-1)$ matrix with all elements zero, except the first K diagonal elements. For $k \le K$, let $\bar{S}_{kk} = w/(k\pi)$. Obviously,

$$\tilde{Y}/\sqrt{N} = \bar{S} + \tilde{\varepsilon}/\sqrt{N} + \left(\tilde{S} - \bar{S}\right).$$

For *arbitrarily* small $\delta > 0$, we can choose K so large that $\left\|\tilde{S} - \bar{S}\right\| < \delta$ with probability at least $1-\delta$, for all sufficiently large N, T. Therefore, the asymptotic behavior of the k-th principal eigenvectors (and eigenvalues) of $\tilde{Y}'\tilde{Y}/N$ and of $\left(\sqrt{N}\bar{S} + \tilde{\varepsilon}\right)'\left(\sqrt{N}\bar{S} + \tilde{\varepsilon}\right)/N$ is the same. In particular, the k-th components of these two principal eigenvectors converge to the same limit.

By Theorem 1 of Onatski (2018), if $w^2 > (k\pi)^2\sqrt{c}$, the k-th component of the k-th principal eigenvector of $\left(\sqrt{N}\bar{S} + \tilde{\varepsilon}\right)'\left(\sqrt{N}\bar{S} + \tilde{\varepsilon}\right)/N$ converges to

$$l_{kk} := \sqrt{\frac{w^4 - c(k\pi)^4}{w^2\left(w^2 + (k\pi)^2\right)}}.$$

If $w^2 \le (k\pi)^2\sqrt{c}$, then the k-th component converges to zero. Furthermore, by Theorem 5 of Onatski (2018), if $w^2 > (k\pi)^2\sqrt{c}$, the k-th principal eigenvalue of $\left(\sqrt{N}\bar{S} + \tilde{\varepsilon}\right)'\left(\sqrt{N}\bar{S} + \tilde{\varepsilon}\right)/N$ converges to

$$\mu_k := \left(w^2/(k\pi)^2 + c\right)\left(1 + (k\pi)^2/w^2\right).$$

If $w^2 \le (k\pi)^2\sqrt{c}$, then the k-th eigenvalue converges to $(1+\sqrt{c})^2$.

In our setting, these results show that Theorem OW2 does not hold when $\omega_T = w/T$. Specifically, the scalar product of the k-th principal eigenvector of $\check{\Sigma}$ with the "k-th cosine wave" does not converge to one, and the k-th principal eigenvalue of $\check{\Sigma}$ does not converge to $w^2/(k\pi)^2$. Instead, if $w^2 > (k\pi)^2 \sqrt{c}$, the scalar product converges to $l_{kk} < 1$ and the eigenvalue converge to $\mu_k > w^2/(k\pi)^2$. If $w^2 \leq (k\pi)^2 \sqrt{c}$, the k-th principal eigenvector of $\check{\Sigma}$ is asymptotically orthogonal to the "k-th cosine wave", and the k-th eigenvalue asymptotically depend only on c, but not on w or k.

Interestingly, even though Theorem OW2 becomes violated, the principal eigenvalues of $\check{\Sigma}$ still decay very fast, for relatively large w. Hence, the scree plot for matrix $\check{\Sigma}$ still can be wrongfully interpreted as showing the existence of factors in the data. This phenomenon gradually disapperas as w becomes smaller and smaller. Similarly, for large w, the k-th principal eigenvector of $\check{\Sigma}$ is "almost collinear" with the "k-th cosine wave", but the quality of the alignment deteriorates as w decreases.

3.2 Local-to-unit roots

Similarly to the proof of Theorem OW1 in Section 2, we analyze the eigenvalues and eigenvectors of $\hat{\Sigma} = MX'XM/N$ in two steps. First, we study a matrix $\tilde{\Sigma}$ with simpler structure, and then show that the results still hold when $\tilde{\Sigma}$ is replaced by $\hat{\Sigma}$. To define $\tilde{\Sigma}$, consider the following extension of the Beveridge-Nelson (BN) decomposition to nearly integrated series (OW11),

$$X_t = Z_t + \Psi^{**}(L)\varepsilon_t, \tag{28}$$

where

$$Z_t - \mu_X = \rho \left(Z_{t-1} - \mu_X\right) + \Psi\left(1\right)\varepsilon_t \tag{29}$$

with

$$Z_0 = X_0 - \Psi^{**}(L)\varepsilon_0, \tag{30}$$

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

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

The series Z_t can be interpreted as the "long run component" of X_t. When $\rho = I$, $\Psi_k^{**} = \Psi_k^*$ and the decomposition reduces to the standard BN one.

To see the validity of (28), use a standard recursive substitution in (OW11) and (29) to obtain

$$X_t - \mu_X = \sum_{j=0}^{t-1} \rho^j \Psi(L)\varepsilon_{t-j} + \rho^t \left(X_0 - \mu_X\right) \text{ and} \tag{31}$$

$$Z_t - \mu_X = \sum_{j=0}^{t-1} \rho^j \Psi(1)\varepsilon_{t-j} + \rho^t \left(Z_0 - \mu_X\right). \tag{32}$$

Subtract (32) from (31), substitute $\rho^t (X_0 - Z_0)$ by $\rho^t \Psi^{**}(L)\varepsilon_0$, and verify that the right hand side of the so obtained equality has form $\Psi^{**}(L)\varepsilon_t$ by matching the coefficients on different lags of ε_t.

We will show that the first-order asymptotic behavior of principal eigenvalues and eigenvectors of $\hat{\Sigma} = MX'XM/N$ is not affected when X is replaced by its long run component $Z = [Z_1, ..., Z_T]$. This is similar to the unit root case. In contrast to the unit root case, $MZ'ZM/N$ is not invariant with respect to the initial values Z_0, which are not eliminated by time avearging $Z \longmapsto ZM$.

To handle the effect of the initial values, we will treat components of Z_t having unit root (first N_1 components) and local-to-unity roots with positive local parameters (last $N - N_1$ components) separately. Denote the j-th rows of $\Psi(L)$ and Z as $\Psi_{j\cdot}(L)$ and $Z_{j\cdot}$, respectively. By assumption, for any $j > N_1$ we have $X_{j0} - \mu_{Xj} = \sum_{i=0}^{\infty} \rho_j^i \Psi_{j\cdot}(L) \varepsilon_{-i}$. Using this in (30) yields $Z_{j0} - \mu_{Xj} = \sum_{i=0}^{\infty} \rho_j^i \Psi_{j\cdot}(1) \varepsilon_{-i}$. Combining this with (32), we see that for any $j > N_1$, Z_{jt} is a stationary process with the initial value Z_{j0} distributed according to its unconditional distribution.

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

$$Z_{jt} - \mu_{Xj} = \Psi_{j\cdot}(1) \sum_{i=0}^{\tau+t-1} \rho_j^i \varepsilon_{t-i} + \rho_j^{\tau+t} \left(Z_{j,-\tau} - \mu_{Xj}\right). \tag{33}$$

for any $\tau \geq 0$ and $j > N_1$. For $\tau \geq 0$ and $j \leq N_1$, let us define $Z_{j,-\tau}$ as $Z_{j0} - \Psi_{j\cdot}(1) \sum_{i=0}^{\tau-1} \varepsilon_{-i}$. With this definition, representation (33) holds for any $\tau \geq 0$ and all $j = 1, ..., N$, not only for $j > N_1$.

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

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

With this notation, we have

$$Z_{j\cdot} = \Psi_{j\cdot}(1) \xi U_j + \rho_j^{T^3} \left[\rho_j^1, \rho_j^2, ..., \rho_j^T\right] \left(Z_{j,-T^3} - \mu_{Xj}\right) + \mu_{Xj}$$

for all $j = 1, ..., N$. Using this representation together with (28), we obtain

$$XM = \begin{bmatrix} \Psi_{1\cdot}(1) \xi U_1 \\ \vdots \\ \Psi_{N\cdot}(1) \xi U_N \end{bmatrix} M + X_{\text{ini}} M + \Psi^{**}(L) \varepsilon M, \tag{34}$$

where

$$X_{\text{ini}} = \rho^{T^3} \left[\rho^1 \left(Z_{-T^3} - \mu_X\right), ..., \rho^T \left(Z_{-T^3} - \mu_X\right)\right] + \mu_X.$$

Similarly to the unit root case, we will show that, under the assumptions of Theorem OW3, the behavior of a few of the largest eigenvalues and the corresponding eigenvectors of $\hat{\Sigma}$ is asymptotically equivalent to that of a few of the largest eigenvalues and corresponding eigenvectors of

$$\tilde{\Sigma} = \frac{1}{N} M \begin{bmatrix} \Psi_{1\cdot}(1)\,\xi U_1 \\ \vdots \\ \Psi_{N\cdot}(1)\,\xi U_N \end{bmatrix}' \begin{bmatrix} \Psi_{1\cdot}(1)\,\xi U_1 \\ \vdots \\ \Psi_{N\cdot}(1)\,\xi U_N \end{bmatrix} M. \tag{35}$$

Therefore our proof strategy is as follows. First, establish statements (i)-(iii) of Theorem OW3 for $\tilde{\lambda}_k, \tilde{F}_k, \tilde{\Sigma}$ instead of $\hat{\lambda}_k, \hat{F}_k, \hat{\Sigma}$ and then, prove that replacing "tildes" by "hats" does not affect the theorem's validity. Here, $\tilde{\lambda}_k$ and $\tilde{F}_k$ denote the k-th principal eigenvalue and eigenvector of $\tilde{\Sigma}$ defined by (35).

3.2.1 Proof of Theorem OW3 for $\tilde{\lambda}_k, \tilde{F}_k, \tilde{\Sigma}$

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

$$\tilde{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} M U_i' \xi' \Psi_{i\cdot}'(1)\, \Psi_{i\cdot}(1)\, \xi U_i M. \tag{36}$$

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

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

where $\Omega = \Psi(1)\,\Psi'(1)$. As will be seen below, the asymptotic behavior of a few of the largest eigenvalues and corresponding eigenvectors of $\tilde{\Sigma}$ and $\mathbb{E}\tilde{\Sigma}$ coincide.

Let us denote the k-th principal eigenvalue and eigenvector of $\mathbb{E}\tilde{\Sigma}$ as $\tilde{\mu}_k$ and $\tilde{\varphi}_k$, respectively. In Section 3.2.4, we prove that under OW's assumptions A1, A2 (which is weaker than A2a), A3 and A5, for any fixed positive integer k, $\tilde{\mu}_k/T^2 \to \mu_k$, where μ_k is the k-th principal eigenvalue of the integral operator $K_{\mathcal{F}}$, defined in the main text. Furthermore, $|\tilde{\varphi}_k' d_k| \to 1$, where $d_k = (\varphi_k(1/T), ..., \varphi_k(T/T))/\sqrt{T}$ and $\varphi_k(s)$ is the k-th principal eigenfunction of $K_{\mathcal{F}}$, and $\tilde{\mu}_k/\operatorname{tr}\mathbb{E}\tilde{\Sigma} \to \mu_k/\sum_{j=1}^{\infty}\mu_j$. In other words, statements (i), (ii), and (iii) of Theorem OW3 hold when $\hat{\lambda}_k$, $\hat{F}_k$, and $\hat{\Sigma}$ are replaced by $\tilde{\mu}_k$, $\tilde{\varphi}_k$, and $\mathbb{E}\tilde{\Sigma}$, respectively. The convergences after replacement are the usual ones rather than in probability because $\mathbb{E}\tilde{\Sigma}$ is a nonrandom matrix.

Given the results announced in the previous paragraph, showing that

$$\left|\tilde{F}_k'\tilde{\varphi}_k\right| \xrightarrow{\mathrm{P}} 1,\ \tilde{\lambda}_k/\tilde{\mu}_k - 1 \xrightarrow{\mathrm{P}} 0,\ \text{and}\ \tilde{\lambda}_k/\operatorname{tr}\tilde{\Sigma} - \tilde{\mu}_k/\operatorname{tr}\mathbb{E}\tilde{\Sigma} \xrightarrow{\mathrm{P}} 0 \tag{37}$$

would establish Theorem OW3 with $\hat{\lambda}_k$, $\hat{F}_k$, and $\hat{\Sigma}$ replaced by $\tilde{\lambda}_k$, $\tilde{F}_k$, and $\tilde{\Sigma}$. Let us now prove the convergencies in (37).

We start from the case $k=1$. Let us represent $\tilde{F}_1$ in the form

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

where the latter equality holds because $\tilde{F}_1$ must be orthogonal to $l_T/\sqrt{T} = \tilde{\varphi}_T$, which is an eigenvector of $\tilde{\Sigma}$ and of $\mathbb{E}\tilde{\Sigma}$ corresponding to the zero eigenvalue (we remind the reader that l_T denotes the T-dimensional vector of ones). The above representation and the definition (36) of $\tilde{\Sigma}$ yield

$$\tilde{\lambda}_1 = \sum\nolimits_{r,q=1}^{T-1} \alpha_r \alpha_q \frac{1}{N} \sum\nolimits_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \tag{38}$$

where

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

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

$$\tilde{\lambda}_{11} = \sum\nolimits_{r,q=1}^{K} \alpha_r \alpha_q \frac{1}{N} \sum\nolimits_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \tag{39}$$

$$\tilde{\lambda}_{12} = \sum\nolimits_{r,q=K+1}^{T-1} \alpha_r \alpha_q \frac{1}{N} \sum\nolimits_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q, \tag{40}$$

and

$$\tilde{\lambda}_{13} = 2\sum\nolimits_{r=1}^{K} \sum\nolimits_{q=K+1}^{T-1} \alpha_r \alpha_q \frac{1}{N} \sum\nolimits_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q. \tag{41}$$

Note that

$$\tilde{\lambda}_1 \leq \left(\tilde{\lambda}_{11}^{1/2} + \tilde{\lambda}_{12}^{1/2}\right)^2. \tag{42}$$

Consider the inner sum in the expression (39) for $\tilde{\lambda}_{11}$. Equation (4) of Lemma 2 yields

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

Further,

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

Equation (5) of Lemma 2 yields

$$\begin{aligned}
& Cov\left(\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q, \tilde{\varphi}_r' A^{(j)}\tilde{\varphi}_q\right) \\
\leq\ & \tilde{\varphi}_r' M U_i' U_j M \tilde{\varphi}_r \tilde{\varphi}_q' M U_i' U_j M \tilde{\varphi}_q \operatorname{tr}\left(\Psi_{i\cdot}'(1)\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\Psi_{j\cdot}(1)\right) \\
& + \tilde{\varphi}_r' M U_i' U_j M \tilde{\varphi}_q \tilde{\varphi}_q' M U_i' U_j M \tilde{\varphi}_r \operatorname{tr}\left(\Psi_{i\cdot}'(1)\Psi_{i\cdot}(1)\Psi_{j\cdot}'(1)\Psi_{j\cdot}(1)\right) \\
& + 2\varkappa_4 \|U_i M\tilde{\varphi}_r\| \left\|U_i M\tilde{\varphi}_q\right\| \|U_j M\tilde{\varphi}_r\| \left\|U_j M\tilde{\varphi}_q\right\| \sum\nolimits_{s=1}^{N_\varepsilon}\left(\Psi_{is}(1)\Psi_{js}(1)\right)^2.
\end{aligned}$$

Section 3.2.3 below proves the following inequality

$$\sup_{\rho_i\in[0,1]} \|U_i M\| \leq \sqrt{2}T. \tag{43}$$

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

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

and

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

We have

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

Therefore,

$$\begin{aligned}
Var\left(\frac{1}{N}\sum\nolimits_{i=1}^{N}\tilde{\varphi}_r' A^{(i)}\tilde{\varphi}_q\right) & \leq \frac{8T^4}{N^2}(1+\varkappa_4)\operatorname{tr}\left[\left(\Psi(1)\Psi'(1)\right)^2\right] \\
& = \frac{8T^4}{N^2}(1+\varkappa_4)\operatorname{tr}\left[\Omega^2\right] = o(1)\frac{T^4}{N^2}\left(\operatorname{tr}\Omega\right)^2,
\end{aligned} \tag{44}$$

where the last equality follows from A3 (as explained in the proof of Corollary 3 above). By Chebyshev's inequality,

$$\tilde{\lambda}_{11} = \sum\nolimits_{r=1}^{K}\alpha_r^2\tilde{\mu}_r + o_{\mathrm{P}}(1)T^2\operatorname{tr}\Omega/N. \tag{45}$$

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

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

By the Cauchy-Schwarz inequality,

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

Lemma 2 yields

$$\mathbb{E}\frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\,\xi U_i M\tilde{\varphi}_r\right)^2 = \sum\nolimits_{r=K+1}^{T-1}\tilde{\mu}_r \tag{46}$$

and

$$\begin{aligned}&Var\left(\frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\,\xi U_i M\tilde{\varphi}_r\right)^2\right)\\ =\;& \frac{1}{N^2}\sum\nolimits_{i,j=1}^{N}\sum\nolimits_{r,q=K+1}^{T-1}Cov\left(\left(\Psi_{i\cdot}(1)\,\xi U_i M\tilde{\varphi}_r\right)^2,\left(\Psi_{j\cdot}(1)\,\xi U_j M\tilde{\varphi}_q\right)^2\right)\\ \leq\;& \frac{2}{N^2}\sum\nolimits_{i,j=1}^{N}\sum\nolimits_{r,q=K+1}^{T-1}\left\|U_iM\tilde{\varphi}_r\right\|^2\left\|U_jM\tilde{\varphi}_q\right\|^2\left[\left(\Psi_{i\cdot}(1)\,\Psi'_{j\cdot}(1)\right)^2+\varkappa_4\sum\nolimits_{s=1}^{T}\Psi_{is}^2(1)\,\Psi_{js}^2(1)\right].\end{aligned}$$

Note that

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

where the latter inequality follows from the fact, established in Section 3.2.3, that $\operatorname{tr}\left(MU_i'U_iM\right) \leq 2T^2$ for any i.

Therefore,

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

Following the steps of the above analysis leading to (44), we obtain

$$Var\left(\frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{r=K+1}^{T-1}\left(\Psi_{i\cdot}(1)\,\xi U_i M\tilde{\varphi}_r\right)^2\right) \leq o(1)\frac{T^4}{N^2}\left(\operatorname{tr}\Omega\right)^2. \tag{47}$$

Chebyshev's inequality together with (46) and (47) yields

$$\tilde{\lambda}_{12} \leq \sum\nolimits_{r=K+1}^{T-1}\tilde{\mu}_r + o_{\mathrm{P}}(1)T^2\operatorname{tr}\Omega/N. \tag{48}$$

The following lemma is proven in Section 3.2.5.

Lemma 6 *For any fixed positive integer J,*

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

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

$$\tilde{\mu}_k \geq C_k T^2 \operatorname{tr} \Omega / N. \tag{49}$$

for all sufficiently large T.

Using the first inequality of the lemma in (48), we obtain

$$\tilde{\lambda}_{12} \leq (1 + o_{\mathrm{P}}(1)) \frac{T^2}{9K} \operatorname{tr} \Omega / N. \tag{50}$$

Now, use (50) and (45) in (42), noting the following two facts. First, by Lemma 6, $\sum_{r=1}^{K} \tilde{\mu}_r / (\operatorname{tr} \Omega / N)$ is of order T^2 for large T. Second, $1/K$ in (50) can be chosen arbitrarily close to zero. Hence, (42) yields

$$\begin{aligned} \tilde{\lambda}_1 & \leq \sum_{r=1}^{T} \alpha_r^2 \tilde{\mu}_r + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N \\ & \leq \alpha_1^2 \tilde{\mu}_1 + (1 - \alpha_1^2) \tilde{\mu}_2 + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N. \end{aligned} \tag{51}$$

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

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

and, by (44),

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

we have by Chebyshev's inequality

$$\tilde{\varphi}_1' \tilde{\Sigma} \tilde{\varphi}_1 = \tilde{\mu}_1 + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N. \tag{52}$$

Therefore,

$$\tilde{\lambda}_1 \geq \tilde{\mu}_1 + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N. \tag{53}$$

Combining this with (51), we obtain

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

which implies

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

But, as is proven in Section 3.2.4, $\tilde{\mu}_1 / T^2 \to \mu_1$ and $\tilde{\mu}_2 / T^2 \to \mu_2$. Since by A5, $\mu_1 > \mu_2$, we have

$$\left(\tilde{F}_1' \tilde{\varphi}_1 \right)^2 = \alpha_1^2 \xrightarrow{\mathrm{P}} 1. \tag{54}$$

This establishes the first convergence in (37) for $k = 1$.

Next, inequalities (51) and (53) yield

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

Combining this with the facts that $\alpha_1^2 = 1 + o_{\mathrm{P}}(1)$ and, by Lemma 6, $\tilde{\mu}_1 \geq CT^2 \operatorname{tr} \Omega / N$ for some $C > 0$, we obtain

$$\tilde{\lambda}_1 = \tilde{\mu}_1 (1 + o_{\mathrm{P}}(1)), \tag{55}$$

which gives us the second convergence in (37) for $k = 1$.

Further,

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

Hence,

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

and, by (47) which holds for all fixed K, including $K = 0$,

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

Hence, by Chebyshev's inequality

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

and

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

where the latter equality is a consequence of Lemma 6. Thus,

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

which establishes the last convergence in (37) for $k = 1$. Note that, by Lemma 6, $\tilde{\mu}_1 / \operatorname{tr} \mathbb{E}\tilde{\Sigma}$ remains bounded away from zero as $N, T \to \infty$.

For $k = m > 1$, the statements of (37) follow by mathematical induction. Indeed, suppose they hold for $k < m$. Consider a representation $\tilde{F}_m = \sum_{q=1}^{T-1} \alpha_q \tilde{\varphi}_q$. Since $\tilde{F}_m' \tilde{F}_j = 0$ for all $j < m$, and since $\left|\tilde{F}_j' \tilde{\varphi}_j\right| = 1 + o_{\mathrm{P}}(1)$ by the induction hypothesis, we must have $\alpha_j = o_{\mathrm{P}}(1)$ for all $j < m$. In particular,

$$\tilde{F}_m' \tilde{\Sigma} \tilde{F}_m = \sum_{q,r=m}^{T-1} \alpha_q \alpha_r \frac{1}{N} \sum\nolimits_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N. \tag{56}$$

Indeed, to see that (56) holds, it is sufficient to establish equalities

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

for any $j < m$, and equalities

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

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

In addition to (56), we must have

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

where the latter equality is obtained similarly to (52). Combining the above two displays, and using the induction hypothesis, this time regarding the validity of the identities

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

for all $i < m$, we obtain

$$\sum_{q,r=m}^{T-1} \alpha_q \alpha_r \frac{1}{N} \sum\nolimits_{i=1}^{N} \tilde{\varphi}_r' A^{(i)} \tilde{\varphi}_q \geq \tilde{\mu}_m + o_{\mathrm{P}}(1) T^2 \operatorname{tr} \Omega / N \tag{57}$$

Statements of (37) for $k = m$ now follow by arguments that are very similar to those used above for the case $k = 1$.

That is, we represent the sum on the left hand side of (57) in the form $\tilde{\lambda}_{m1} + \tilde{\lambda}_{m2} + \tilde{\lambda}_{m3}$, defined similarly to (39-41). Then proceed along the lines of the above proof to obtain an upper bound on $\tilde{\lambda}_{m1} + \tilde{\lambda}_{m2} + \tilde{\lambda}_{m3}$, similar to the right hand side of (51). Then, combining this upper bound with the lower bound (57), we prove the convergence $\alpha_m^2 \xrightarrow{\mathrm{P}} 1$. Finally, we proceed to establishing the other statements of (37) using this convergence.

3.2.2 Proof of Theorem OW3 for $\hat{\lambda}_k, \hat{F}_k, \hat{\Sigma}$

We need to show that the theorem's validity for $\tilde{F}_k$, $\tilde{\lambda}_k$ and $\tilde{\Sigma}$ implies its validity for $\hat{F}_k$, $\hat{\lambda}_k$ and $\hat{\Sigma}$. By standard perturbation theory (e.g. Kato (1980), ch. 2), such an implication for statements (i) and (ii) would follow if we are able to show that $\left\|\hat{\Sigma} - \tilde{\Sigma}\right\| = \frac{T^2}{N} \operatorname{tr} \Omega \, o_{\mathrm{P}}(1)$. Equation (34) implies that it is sufficient to establish two facts. First, $\|X_{\text{ini}}M\|^2 = T^2 \operatorname{tr}\Omega\, o_{\text{P}}(1)$, and second, $\|\Psi^{**}(L)\,\varepsilon M\|^2 = T^2 \operatorname{tr}\Omega\, o_{\text{P}}(1)$.

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

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

and

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

where $h(\rho_i)$ is as defined in (67) below. As shown there, $h(\rho_i)$ is non-negative, continuous, $|h(\rho_i)| \leq 1$ for all T, and $h(1-x/T) \leq x/4$ for $x \in [0,1)$. This implies that

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

and

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

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

$$\mathbb{E}\,\|X_{\text{ini}}M\|_F^2 \leq \frac{1}{2}\sum\nolimits_{i=N_1+1}^{N} \Psi_{i\cdot}(1)\,\Psi_{i\cdot}'(1) \leq \frac{1}{2}\operatorname{tr}\Omega. \tag{58}$$

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

$$\|X_{\text{ini}}M\|^2 = \operatorname{tr}\Omega\, O_{\text{P}}(1) = T^2 \operatorname{tr}\Omega\, o_{\text{P}}(1), \tag{59}$$

as required.

It remains to show that $\|\Psi^{**}(L)\,\varepsilon M\|^2 = T^2 \operatorname{tr}\Omega\, o_{\text{P}}(1)$. Note that

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

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

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

We have

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

Further,

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

Combining the latter two displays, we obtain

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

Hence, by Lemma 4 and Remark 5,

$$\|\Pi^{**}(L)\varepsilon\|^2=O_{\mathrm{P}}\left(TN^{2\alpha}+N_{\varepsilon}N^{2\alpha}\right). \tag{60}$$

This equality, the assumption of the theorem that $(T+N_{\varepsilon})N^{2\alpha-1}/T^2=o(1)$, and the fact that, under A5, $N/\operatorname{tr}\Omega=O(1)$, yield

$$\|\Pi^{**}(L)\varepsilon\|^2=T^2\operatorname{tr}\Omega\, o_{\mathrm{P}}(1).$$

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

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

Therefore,

$$\begin{aligned}k\|\Theta_k^{**}\|_F &\leq k\sum_{j=1}^{k}\left\|\frac{\rho^{k-j}-\rho^k}{j}\right\|j\|\Psi_j\|_F\\ &\leq \left\|I-\rho^k\right\|\sum_{j=1}^{k}j\|\Psi_j\|_F=O\left(N^{\alpha}+\min\left\{N^{1/2},N_{\varepsilon}^{1/2}\right\}\right),\end{aligned}$$

where the last equality follows by assumption A2a. Therefore,

$$\begin{aligned} T\sum\nolimits_{k=T+1}^{\infty}\|\Theta_k^{**}\|_F^2 &\leq O\left(N^{2\alpha}+\min\{N,N_\varepsilon\}\right)T\sum\nolimits_{k=T+1}^{\infty}\frac{1}{k^2} \\ &= O\left(N^{2\alpha}+\min\{N,N_\varepsilon\}\right). \end{aligned}$$

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

$$\begin{aligned} \sum\nolimits_{k=1}^{T}\|\Theta_k^{**}\| &\leq \sum\nolimits_{k=1}^{T}\sum_{j=1}^{k}\left\|\rho^{k-j}-\rho^k\right\|\|\Psi_j\| \\ &= \sum_{j=1}^{T}\|\Psi_j\|\sum\nolimits_{k=j}^{T}\left\|\rho^{k-j}-\rho^k\right\|. \end{aligned} \tag{61}$$

By assumption A5, there exists $\bar{\phi}>0$ such that $\phi_j\leq\bar{\phi}$ for all $j\in\mathbb{N}$. Note that the maximum of $r^{k-j}-r^k$ on $r\in[0,1]$ is achieved at $r=(1-j/k)^{1/j}$. On the other hand, the smallest possible diagonal element of ρ equals $e^{-\bar{\phi}/T}$, and

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

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

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

and

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

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

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

and for all sufficiently large T,

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

Hence,

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

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

$$\left\|\rho^{k-j} - \rho^k\right\| \leq \left(\frac{k-j}{k}\right)^{k/j-1} \frac{j}{k} \leq \frac{j}{k} \tag{62}$$

and

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

Hence, overall,

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

and thus,

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

In particular, the assumptions of statement (ii) of lemma 4 are satisfied and

$$\left\|\Theta^{**}(L)\,\varepsilon\right\| = O_{\mathrm{P}}\left(T^{1/2} + N_{\varepsilon}^{1/2}\right) N^{\alpha}. \tag{63}$$

Since by assumption $(T + N_{\varepsilon})\, N^{2\alpha-1}/T^2 = o(1)$, we have

$$\left\|\Theta^{**}(L)\,\varepsilon\right\|^2 = T^2 \operatorname{tr}\Omega\, o_{\mathrm{P}}(1),$$

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

Part (iii) of the theorem can be established similarly to part (iii) of Theorem OW1, using the fact that, by Lemma 6, there exist positive constants C_1 and C_2 such that

$$C_1 \frac{T^2}{N} \operatorname{tr}\Omega \leq \operatorname{tr}\tilde{\Sigma} \leq C_2 \frac{T^2}{N} \operatorname{tr}\Omega. \tag{64}$$

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

From (34),

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

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

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

and

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

By (64), $\left(\operatorname{tr}\tilde{\Sigma}\right)^{1/2}=O_{\mathrm{P}}\left(T\right)$. Therefore, to establish the asymptotic dominance of $\left|\operatorname{tr}\hat{\Sigma}-\operatorname{tr}\tilde{\Sigma}\right|$ by T^{2} it is sufficient to show that

$$
\begin{aligned}
\left\|\Psi^{**}\left(L\right)\varepsilon M\right\|^{2}\min\left\{1,T/N\right\} \quad &= \quad o_{\mathrm{P}}\left(T^{2}\right)\ \text{ and} && (65)\\
\left\|X_{\text{ini}}M\right\|^{2}\min\left\{1,T/N\right\} \quad &= \quad o_{\mathrm{P}}\left(T^{2}\right). && (66)
\end{aligned}
$$

Since $\left\|\Psi^{**}\left(L\right)\varepsilon M\right\|\leq\left\|\Theta^{**}\left(L\right)\varepsilon\right\|+\left\|\Pi^{**}\left(L\right)\varepsilon\right\|$, equalities (60) and (63) yield

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

Hence,

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

But, by assumption of (iii), $\left(T+N_{\varepsilon}\right)N^{2\alpha}/\left(T\max\left\{N,T\right\}\right)\to0$. Therefore (65) holds.

Finally, by (59),

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

Therefore,

$$
\left\|X_{\text{ini}}M\right\|^{2}\min\left\{1,T/N\right\}=\frac{T\operatorname{tr}\Omega}{\max\left\{T,N\right\}}O_{\mathrm{P}}\left(1\right).
$$

Since $\operatorname{tr}\Omega/N=O(1)$, we have

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

and (66) holds.

3.2.3 Bound on the norm of U_iM

Since $\left\|U_{i}M\right\|^{2}\leq\operatorname{tr}\left(MU_{i}'U_{i}M\right)$, it is sufficient to prove that $\sup_{\rho_{i}\in[0,1]}\operatorname{tr}\left(MU_{i}'U_{i}M\right)\leq2T^{2}$. Let $U_{i}^{(1)}$ be the upper $T^{3}\times T$ block of U_{i} and $U_{i}^{(2)}$ be the lower $T\times T$ block. Then,

$$
\begin{aligned}
\operatorname{tr}\left(MU_{i}'U_{i}M\right) \quad &= \quad \operatorname{tr}\left(MU_{i}^{(1)\prime}U_{i}^{(1)}M\right)+\operatorname{tr}\left(MU_{i}^{(2)\prime}U_{i}^{(2)}M\right)\\
&\leq \quad \operatorname{tr}\left(MU_{i}^{(1)\prime}U_{i}^{(1)}M\right)+\operatorname{tr}\left(U_{i}^{(2)\prime}U_{i}^{(2)}\right)\\
&\leq \quad \operatorname{tr}\left(MU_{i}^{(1)\prime}U_{i}^{(1)}M\right)+T^{2},
\end{aligned}
$$

where the last inequality follows from the fact that tr $\left(U_i^{(2)\prime}U_i^{(2)}\right)$ equals the sum of squared elements of the $T\times T$ matrix $U_i^{(2)}$ and all these elements are non-negative and no larger than 1. Hence, it is sufficient to prove that $\sup_{\rho_i\in[0,1]}\operatorname{tr}\left(MU_i^{(1)\prime}U_i^{(1)}M\right)\leq T^2$.

Note that

$$U_i^{(1)}=\left(\begin{array}{cccc}\rho_i^{T^3} & \rho_i^{T^3-1} & \cdots & \rho_i\end{array}\right)'\left(\begin{array}{cccc}1 & \rho_i & \cdots & \rho_i^{T-1}\end{array}\right).$$

Therefore, for $\rho_i=1$, $U_i^{(1)}M=0$ and tr $\left(MU_i^{(1)\prime}U_i^{(1)}M\right)\leq T^2$ trivially holds. For $\rho_i<1$, an elementary calculation yields

$$\begin{aligned}
\operatorname{tr}\left(MU_i^{(1)\prime}U_i^{(1)}M\right) &= \rho_i^2\frac{1-\rho_i^{2T^3}}{1-\rho_i^2}\left(\frac{1-\rho_i^{2T}}{1-\rho_i^2}-\frac{1}{T}\left(\frac{1-\rho_i^T}{1-\rho_i}\right)^2\right)\\
&\leq \frac{1}{1-\rho_i}\left(\frac{1-\rho_i^{2T}}{1-\rho_i^2}-\frac{1}{T}\left(\frac{1-\rho_i^T}{1-\rho_i}\right)^2\right)\\
&= \frac{1-\rho_i^T}{(1-\rho_i)^2}\left(\frac{1+\rho_i^T}{1+\rho_i}-\frac{1}{T}\frac{1-\rho_i^T}{1-\rho_i}\right)\\
&\leq \frac{T}{1-\rho_i}\left(\frac{1+\rho_i^T}{1+\rho_i}-\frac{1}{T}\frac{1-\rho_i^T}{1-\rho_i}\right).
\end{aligned}$$

Since the term in the final bracket is no larger than unity, the obtained bound on tr $\left(MU_i^{(1)\prime}U_i^{(1)}M\right)$ is no larger than T^2 for all non-negative $\rho_i\leq 1-1/T$. Hence, it is sufficient to show that

$$\sup_{\rho_i\in(1-1/T,1)}\frac{1}{1-\rho_i}\left(\frac{1+\rho_i^T}{1+\rho_i}-\frac{1}{T}\frac{1-\rho_i^T}{1-\rho_i}\right)\leq T.$$

Let us reparametrize the problem using $\rho_i=1-x/T$, where $x\in(0,1)$. It is sufficient to show that

$$\sup_{x\in(0,1)}\frac{1}{x}\left(\frac{1+(1-x/T)^T}{2-x/T}-\frac{1-(1-x/T)^T}{x}\right)\leq 1.$$

The Taylor expansion of $(1-x/T)^T$ at zero yields

$$(1-x/T)^T=1-x+\frac{T-1}{2T}\left(1-\frac{x^*}{T}\right)^{T-2}x^2,$$

where $x^*\in[0,x]$. Therefore, for all $T\geq 2$ and $x\in(0,1)$ we have

$$(1-x/T)^T=1-x+R_{x,T}x^2 \text{ with } |R_{x,T}|\leq 1/2.$$

This yields

$$\frac{1}{x}\left(\frac{1+(1-x/T)^T}{2-x/T}-\frac{1-(1-x/T)^T}{x}\right)=\frac{1}{x}\left(\frac{x/T-x+R_{x,T}x^2}{2-x/T}+R_{x,T}x\right).$$

But for $T \geq 2$ and $x \in (0,1)$, we have $x/T - x + R_{x,T}x^2 \leq 0$. Therefore, the right hand side of the displayed equality is no larger than $R_{x,T}$. Thus,

$$\sup_{x\in(0,1)} \frac{1}{x}\left(\frac{1+(1-x/T)^T}{2-x/T} - \frac{1-(1-x/T)^T}{x}\right) \leq 1/2 < 1.$$

This completes the proof of inequality (43) for all $T \geq 2$. A direct verification shows that the inequality also holds for $T = 1$.

As a bi-product, we established the fact that function

$$h(\rho_i) = \begin{cases} \frac{1}{T}\frac{1}{1-\rho_i^2}\left(\frac{1+\rho_i^T}{1+\rho_i} - \frac{1}{T}\frac{1-\rho_i^T}{1-\rho_i}\right) & \text{for } \rho_i \in [0,1) \\ 0 & \text{for } \rho_i = 1 \end{cases} \tag{67}$$

is non-negative, continuous, uniformly in T bounded, and such that, for all T, $h(\rho_i) \leq 1$ and $h(1-x/T) \leq x/4$ for $x \in [0,1)$. We refer to equation (67) in Section 3.2.2 above.

3.2.4 Asymptotic analysis of the eigenstructure of $\mathbb{E}\tilde{\Sigma}$

The final goal of this section is to establish the results claimed in the paragraph immediately preceding equation (37), which formed the basis of our proof of Theorem OW3 in Section 3.2.1. Recall that the results are formulated as follows. Under A1, A2, A3 and A5, for any fixed positive integer k

$$\tilde{\mu}_k/T^2 \rightarrow \mu_k, \tag{68}$$

$$\left|\tilde{\varphi}_k' d_k\right| \rightarrow 1, \text{and} \tag{69}$$

$$\tilde{\mu}_k/\operatorname{tr}\mathbb{E}\tilde{\Sigma} \rightarrow \mu_k/\sum\nolimits_{j=1}^{\infty}\mu_j. \tag{70}$$

Here $\tilde{\mu}_k$ and $\tilde{\varphi}_k$ are the k-th principal eigenvalue and eigenvector of $\mathbb{E}\tilde{\Sigma}$, μ_k is the k-th principal eigenvalue of $K_{\mathcal{F}}$, and $d_k = (\varphi_k(1/T), ..., \varphi_k(T/T))/\sqrt{T}$, where φ_k is the k-th principal eigenfunction of $K_{\mathcal{F}}$.

To establish (68-70), we will prove that there exist approximating integral operators $K_{N,T}$ acting on the space of continuous functions on $[0,1]$ equipped with the supremum norm, $\|\cdot\|_{\sup}$, such that, on one hand, their principal eigenvalues and eigenfunctions converge to those of $K_{\mathcal{F}}$, and on the other hand, the nonzero eigenvalues of $K_{N,T}$ coincide with those of $\mathbb{E}\tilde{\Sigma}/T^2$, and the corresponding eigenfunctions evaluated on the grid $1/T, 2/T, ..., T/T$ are eigenvectors of $\mathbb{E}\tilde{\Sigma}/T^2$. Convergences (68-69) immediately follow from the existence of such approximating operators. Convergence (70) follows from such an existence, Lemma 6, and the fact that, by assumption A5, $\operatorname{tr}\Omega/N \leq \bar{\omega} < \infty$.

In the rest of this section, we establish the existence of $K_{N,T}$ with the above described properties. Consider the stationary Ornstein-Uhlenbeck process $x_\phi(s)$, generated by stochastic differential equation

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

with the standard Wiener process $W(s)$ and $\phi > 0$. The initial observation $x_{\phi}(0)$ is drawn from the unconditional distribution of $x_{\phi}(s)$. As is well known (e.g. Karatzas and Shreve (1998, p. 358)), the covariance kernel of $x_{\phi}(s)$ is given by $e^{-\phi|t-s|}/\left(2\phi\right)$. It is straightforward to verify that the covariance kernel of the *demeaned* Ornstein-Uhlenbeck process equals

$$\begin{aligned} k_{\phi}(s,t) &= a_{\phi}(s,t) - \int_0^1 a_{\phi}(s,t)\,\mathrm{d}s - \int_0^1 a_{\phi}(s,t)\,\mathrm{d}t + \int_0^1\int_0^1 a_{\phi}(s,t)\,\mathrm{d}s\mathrm{d}t \\ &= a_{\phi}(s,t) - b_{\phi}(t) - b_{\phi}(s) + c_{\phi}, \end{aligned}$$

where

$$\begin{aligned} a_{\phi}(s,t) &= \left(e^{-\phi|t-s|} - 1\right)/\left(2\phi\right), \\ b_{\phi}(t) &= \left(2 - \phi - e^{-\phi t} - e^{-\phi(1-t)}\right)/\left(2\phi^2\right), \text{ and} \\ c_{\phi} &= (e^{-\phi} - 1 + \phi - \phi^2/2)/\phi^3. \end{aligned}$$

The Taylor expansion of the numerators of $a_{\phi}(s,t)$, $b_{\phi}(t)$, $b_{\phi}(s)$, and c_{ϕ} at $\phi = 0$ reveals that as $\phi \to 0$, $k_{\phi}(s,t)$ converges to

$$\begin{aligned} k_0(s,t) &= -\frac{|t-s|}{2} + \int_0^1 \frac{|t-s|}{2}\mathrm{d}s + \int_0^1 \frac{|t-s|}{2}\mathrm{d}t - \int_0^1\int_0^1 \frac{|t-s|}{2}\mathrm{d}s\mathrm{d}t \\ &= \min\{t,s\} + s^2/2 - s + t^2/2 - t + 1/3, \end{aligned}$$

which is the covariance kernel of the *demeaned* Wiener process. We have

$$k_0(s,t) = a_0(s,t) - b_0(t) - b_0(s) + c_0$$

with

$$\begin{aligned} a_0(s,t) &= -|t-s|/2, \\ b_0(t) &= -t^2/2 + t/2 - 1/4, \text{ and } c_0 = -1/6. \end{aligned}$$

Let

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

A direct derivation yields, for $\phi > 0$,

$$\begin{aligned} \frac{1}{T}\left(MU_\phi' U_\phi M\right)_{ij} &= \frac{1}{T}\left(U_\phi' U_\phi\right)_{ij} - \frac{1}{T^2}\left(U_\phi' U_\phi l\right)_i - \frac{1}{T^2}\left(U_\phi' U_\phi l\right)_j + \frac{1}{T^3} l' U_\phi' U_\phi l \\ &= \frac{e^{-\phi|j-i|/T} - e^{-2\phi T^2} e^{-\phi(j+i)/T}}{T\left(1-e^{-2\phi/T}\right)} \\ &\quad - \frac{e^{\phi/T}\left(1-e^{-\phi i/T}\right) + 1 - e^{-\phi} e^{\phi i/T} - e^{-2\phi T^2} e^{-\phi i/T}\left(1-e^{-\phi}\right)}{T^2\left(1-e^{-2\phi/T}\right)\left(e^{\phi/T}-1\right)} \\ &\quad - \frac{e^{\phi/T}\left(1-e^{-\phi j/T}\right) + 1 - e^{-\phi} e^{\phi j/T} - e^{-2\phi T^2} e^{-\phi j/T}\left(1-e^{-\phi}\right)}{T^2\left(1-e^{-2\phi/T}\right)\left(e^{\phi/T}-1\right)} \\ &\quad + \frac{2e^{\phi/T}\left(e^{-\phi}-1\right) + T\left(e^{2\phi/T}-1\right) - e^{-2\phi T^2}\left(1-e^{-\phi}\right)^2}{T^3\left(1-e^{-2\phi/T}\right)\left(e^{\phi/T}-1\right)^2}. \end{aligned}$$

For $\phi = 0$, we have

$$\begin{aligned} \frac{1}{T}\left(MU_0' U_0 M\right)_{ij} &= \min\{i/T, j/T\} + \frac{(i/T)^2}{2} - i/T - (i/T)/(2T) \\ &\quad + \frac{(j/T)^2}{2} - j/T - (j/T)/(2T) + T(T+1)(2T+1)/\left(6T^3\right) \\ &= -|i/T - j/T|/2 + \frac{(i/T)^2}{2} - \frac{i/T}{2} + \frac{1}{4} + \frac{(j/T)^2}{2} - \frac{j/T}{2} + \frac{1}{4} \\ &\quad - (i/T)/(2T) - (j/T)/(2T) + T(T+1)(2T+1)/\left(6T^3\right) - 1/2. \end{aligned}$$

For $\phi \geq 0$, we have the following representation

$$\frac{1}{T}\left(MU_\phi' U_\phi M\right)_{ij} = \omega_{\phi 1,T} a_\phi(s_i, t_j) - \omega_{\phi 2,T}\left(b_\phi(t_j) + b_\phi(s_i)\right) + d_{\phi,T} - e_{\phi,T}(s_i, t_j),$$

where $s_i = i/T$, $t_j = j/T$, a_ϕ and b_ϕ are as defined above, whereas $\omega_{\phi 1,T}$, $\omega_{\phi 2,T}$, $d_{\phi,T}$, and $e_{\phi,T}(s_i, t_j)$ are as follows. For $\phi > 0$,

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

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

and

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

For $\phi=0$,

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

For $\phi \geq 0$, define

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

Then

$$\left(MU_\phi' U_\phi M\right)_{ij}/T=k_{\phi,T}\left(s_i,t_j\right). \tag{71}$$

Now, consider integrated kernels

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

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

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

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

Identity (71) implies that the eigenvalues of $\mathbb{E}\tilde{\Sigma}/T^2$ are also eigenvalues of $K_{N,T}$. Moreover, if $x(t)$ is an eigenfunction of $K_{N,T}$, then $\left(x\left(t_1\right),...,x\left(t_T\right)\right)'$ is an eigenvector of $\mathbb{E}\tilde{\Sigma}/T^2$. Vice versa, if $\left(x_1,...,x_T\right)'$ is an eigenvector of $\mathbb{E}\tilde{\Sigma}/T^2$, then there exists $x\in C[0,1]$ with $x\left(t_j\right)=x_j$ such that x is an eigenfunction of $K_{N,T}$. In other words, the spectral properties of $K_{N,T}$ and $\mathbb{E}\tilde{\Sigma}/T^2$ are essentially the same, even though the first is an operator in $C[0,1]$ while the second is a $T\times T$ matrix. Anselone

(1967) traces the technique of approximating integral operator by matrices back to Fredholm, and the idea of mapping matrices to operators with essentially same spectral properties to Nystrom.

It remains to prove that the principal eigenvalues and eigenfunctions of $K_{N,T}$ converge to those of $K_{\mathcal{F}}$. Our proof is based on the ideas of Anselone (1967). The key facts to establish are: the pointwise convergence $K_{N,T} \to K_{\mathcal{F}}$ and the collective compactness of the sequence of operators $\{K_{N,T} : N, T = 1, 2, ...\}$ (see Anselone (1967) and the discussion below for the definition of collective compactness). After establishing these facts, we show how they imply the convergence of the principal eigenvalues and eigenfunctions.

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

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

$$\int \mathbf{1}\{\phi \geq \phi_\epsilon\} \mathrm{d}\mathcal{F}(z,\phi) < \epsilon/(3\bar{\omega}) \text{ and } \int \mathbf{1}\{\phi \geq \phi_\epsilon\} \mathrm{d}\mathcal{F}_N(z,\phi) < \epsilon/(21\bar{\omega})$$

for all $N > N_2$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function. For any $\epsilon > 0$, the displayed inequalities can be satisfied by choosing ϕ_ϵ sufficiently large because $\mathcal{F}(z,\phi)$ is a cumulative distribution function of a proper probability distribution and $\mathcal{F}_N$ weakly converges to $\mathcal{F}$ as $N \to \infty$. In fact, by A5, any ϕ from the supports of $\mathcal{F}(z,\phi)$ and $\mathcal{F}_N(z,\phi)$ satisfies $\phi \leq \bar{\phi}$. In particular, we can set $\phi_\epsilon = \bar{\phi}$. However, in this subsection, we do not need to (and will not) assume the boundedness of the supports of $\mathcal{F}(z,\phi)$ and $\mathcal{F}_N(z,\phi)$ with respect to ϕ.

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

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

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

and $P_3 = K_{N,T}x - K_{\mathcal{F}}x - P_1 - P_2$ is the remainder. To analyze P_1 and P_2, we need the following lemma.

Lemma 7 *Kernels $k_\phi(s,t)$ and $k_{\phi,T}(s,t)$ are bounded by absolute value uniformly in $\phi \geq 0$. Specifically,*

$$\sup_{\phi \geq 0} \max_{s,t \in [0,1]^2} |k_\phi(s,t)| \leq 1 \text{ and } \sup_{\phi \geq 0} \sup_{T \geq 1} \max_{s,t \in [0,1]^2} |k_{\phi,T}(s,t)| \leq 7.$$

Proof: The uniform boundedness of $|k_\phi(s,t)|$ follows from that of $|a_\phi(s,t)| = -a_\phi(s,t)$ and the definitions $b_\phi(s) = \int_0^1 a_\phi(s,t) \mathrm{d}t$ and $c_\phi = \int_0^1 \int_0^1 a_\phi(s,t) \mathrm{d}t\mathrm{d}s$. The uniform boundedness of

$|a_\phi(s,t)|$ follows from the inequality $e^{-\phi x} \geq 1-\phi$. This inequality implies that the maximum of $|a_\phi(s,t)|$ over $s,t \in [0,1]^2$ is no larger than $1/2$. The uniform bound on $|k_\phi(s,t)|$ equals 1 because $k_\phi(s,t) \leq -b_\phi(s) - b_\phi(t) \leq 1$ and $-k_\phi(s,t) \leq -a_\phi(s,t) - c_\phi \leq 1$.

To establish the uniform boundedness of $|k_{\phi,T}(s,t)|$, we will prove that $|\omega_{\phi 1,T} a_\phi(s,t)|$, $|\omega_{\phi 2,T}(b_\phi(t)+b_\phi(s))|$, $|d_{\phi,T}|$, and $|e_{\phi,T}(s,t)|$ are uniformly bounded. For $\phi > 0$, we have

$$|\omega_{\phi 1,T} a_\phi(s,t)| = \frac{1-e^{-\phi|t-s|}}{T\left(1-e^{-2\phi/T}\right)} = \frac{1-\rho^{T|t-s|}}{T\left(1-\rho^2\right)},$$

where $\rho = e^{-\phi/T}$. This yields

$$|\omega_{\phi 1,T} a_\phi(s,t)| \leq \frac{1-\rho^T}{T\left(1-\rho^2\right)} \leq 1.$$

Clearly, $|\omega_{01,T} a_0(s,t)| = |t-s|/2 < 1$. Hence, $|\omega_{\phi 1,T} a_\phi(s,t)| \leq 1$ for all $\phi \geq 0$ and all positive integers T.

Note that

$$\omega_{\phi 2,T} b_\phi(t) = \frac{\phi/T}{e^{\phi/T}-1}\omega_{\phi 1,T} b_\phi(t)$$

for $\phi > 0$ and $\omega_{02,T} b_0(t) = \omega_{01,T} b_0(t)$. Since $\omega_{\phi 1,T} b_\phi(t) = \int_0^1 \omega_{\phi 1,T} a_\phi(s,t)\,\mathrm{d}s$ and $|\omega_{\phi 1,T} a_\phi(s,t)| \leq 1$ for all $\phi \geq 0$ and T, we have $|\omega_{\phi 1,T} b_\phi(t)| \leq 1$. But $\left|\frac{\phi/T}{e^{\phi/T}-1}\right| \leq 1$. Therefore, $|\omega_{\phi 2,T} b_\phi(t)| \leq 1$ for all $\phi \geq 0$ and T. Hence, $|\omega_{\phi 2,T}(b_\phi(s)+b_\phi(t))| \leq 2$ for all $\phi \geq 0$ and T.

Next, by definition, for $\phi > 0$,

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

where $\rho = e^{-\phi/T}$. This yields, after some algebra,

$$d_{\phi,T} = -\frac{\sum_{j=0}^{T-2}(T-j)(T-j-1)\rho^j}{T^3(1+\rho)}.$$

Therefore, $|d_{\phi,T}| \leq 1$ for all $\phi > 0$ and T. For $\phi = 0$, $d_{0,T} = (T+1)(2T+1)/\left(6T^2\right) - 1/2$, and hence, $|d_{0,T}| \leq 1/2$ for all T. To summarize, $|d_{\phi,T}| \leq 1$ for all $\phi \geq 0$ and T.

Finally, for $\phi > 0$, we have

$$e_{\phi,T}(s,t) = e_{\phi 1,T}(s,t) + e^{-2\phi T^2} e_{\phi 2,T}(s,t)$$

with

$$e_{\phi 1,T}(s,t) = \frac{2-e^{-\phi s}-e^{-\phi t}}{T^2\left(1-e^{-2\phi/T}\right)} \text{ and } \tag{72}$$

$$\begin{aligned} e_{\phi 2,T}(s,t) &= \frac{e^{-\phi(t+s)}-1}{T\left(1-e^{-2\phi/T}\right)} + \frac{\left(2-e^{-\phi s}-e^{-\phi t}\right)e^{-\phi/T}\left(1-e^{-\phi}\right)}{T^2\left(1-e^{-2\phi/T}\right)\left(1-e^{-\phi/T}\right)} \\ &\quad + \frac{\left(T\left(1-e^{-\phi/T}\right)-\left(1-e^{-\phi}\right)e^{-\phi/T}\right)^2}{T^3\left(1-e^{-2\phi/T}\right)\left(1-e^{-\phi/T}\right)^2}. \end{aligned} \tag{73}$$

For term $e_{\phi 1,T}(s,t)$, we have

$$e_{\phi 1,T}(s,t) = \frac{2-\rho^{Ts}-\rho^{Tt}}{T^2\left(1-\rho^2\right)} \leq \frac{2\left(1-\rho^T\right)}{T^2\left(1-\rho^2\right)} \leq \frac{2}{T}. \tag{74}$$

For term $e_{\phi 2,T}(s,t)$, we have, after some algebra,

$$\begin{aligned} e_{\phi 2,T}(s,t) &= -\frac{1-\rho^{T(t+s)}}{T\left(1-\rho^2\right)} + \frac{\left(2-\rho^{Ts}-\rho^{Tt}\right)\rho\left(1-\rho^T\right)}{T^2\left(1-\rho^2\right)(1-\rho)} \\ &\quad - \frac{\sum_{j=0}^{T-1}(T-j)\rho^j}{T^3(1+\rho)}. \end{aligned} \tag{75}$$

On the other hand,

$$\frac{1-\rho^{T(t+s)}}{T\left(1-\rho^2\right)} \leq \frac{1-\rho^{2T}}{T\left(1-\rho^2\right)} \leq \frac{1+...+\rho^{2T-1}}{T(1+\rho)} \leq 2, \tag{76}$$

$$\frac{\left(2-\rho^{Ts}-\rho^{Tt}\right)\rho\left(1-\rho^T\right)}{T^2\left(1-\rho^2\right)(1-\rho)} \leq \frac{2\rho\left(1+...+\rho^{T-1}\right)^2}{T^2(1+\rho)} \leq 2, \tag{77}$$

and

$$\frac{\sum_{j=0}^{T-1}(T-j)\rho^j}{T^3(1+\rho)} \leq \frac{1}{T} \leq 1. \tag{78}$$

These bounds yield $e_{\phi 2,T}(s,t) \leq 2$ and $-e_{\phi 2,T}(s,t) \leq 3$. Combining this with the above bound for $e_{\phi 1,T}(s,t)$ yields $e_{\phi,T}(s,t) \leq 3$ and $-e_{\phi,T}(s,t) \leq 3$ so that

$$|e_{\phi,T}(s,t)| \leq 3$$

for all $\phi > 0$ and all T. For $\phi = 0$, we obviously have $|e_{0,T}(s,t)| = |s+t|/(2T) \leq 1$. Summing up the above results, we obtain

$$\sup_{\phi\geq 0}\sup_{T\geq 1}\max_{s,t\in[0,1]^2}|k_{\phi,T}(s,t)| \leq 1+2+1+3 = 7.\square$$

Lemma 7 implies that, for all $N > N_2$,

$$\begin{aligned} |P_1| &\leq \int_0^1 \int |\omega (1 - f_\epsilon(\phi)) k_\phi(s,t) x(t)| \, \mathrm{d}\mathcal{F}(\omega,\phi) \, \mathrm{d}t \\ &\leq \int \bar{\omega} \mathbf{1} \{\phi \geq \phi_\epsilon\} \, \mathrm{d}\mathcal{F}(\omega,\phi) < \bar{\omega}\epsilon / (3\bar{\omega}) = \epsilon/3. \end{aligned} \tag{79}$$

Similarly, for all $N > N_2$,

$$\begin{aligned} |P_2| &\leq \frac{1}{T} \sum_{j=1}^{T} \int |\omega (1 - f_\epsilon(\phi)) k_{\phi,T}(s,t_j) x(t_j)| \, \mathrm{d}\mathcal{F}_N(\omega,\phi) \\ &\leq \int 7\bar{\omega} \mathbf{1} \{\phi \geq \phi_\epsilon\} \, \mathrm{d}\mathcal{F}_N(\omega,\phi) < 7\bar{\omega}\epsilon / (21\bar{\omega}) = \epsilon/3. \end{aligned} \tag{80}$$

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

Consider the following decomposition

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

where

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

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

and

$$\begin{aligned} y_3(s) &= \frac{1}{T} \sum_{j=1}^{T} \int \omega f_\epsilon(\phi) k_{\phi,T}(s,t_j) x(t_j) \mathrm{d}\mathcal{F}_N(\omega,\phi) \\ &\quad - \int_0^1 \int \omega f_\epsilon(\phi) k_{\phi,T}(s,t) x(t) \mathrm{d}\mathcal{F}_N(\omega,\phi) \, \mathrm{d}t. \end{aligned}$$

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

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

$$\sup_{\phi \geq 0} \sup_{s,t \in [0,1]^2} |f_\epsilon(\phi) (k_\phi(s,t) - k_{\phi,T}(s,t))| \to 0 \tag{81}$$

as $T \to \infty$. To see that (81) holds, consider the decomposition

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

As follows from the proof of Lemma 7, $|a_\phi(s,t)|$ and $|b_\phi(t)+b_\phi(s)|$ are bounded uniformly in $\phi \geq 0$. On the other hand, $1-\omega_{\phi 1,T} \to 0$, $1-\omega_{\phi 2,T} \to 0$, and $c_\phi - d_{\phi,T} \to 0$ uniformly on $\phi \in [0, 2\phi_\epsilon]$ (the support of f_ϵ). Hence, the first three terms on the right hand side of the above display converge to zero uniformly in ϕ, s, and t.

For the last term, we have

$$\begin{aligned} & \left|f_\epsilon(\phi)\, e_{\phi,T}(s,t)\right| \leq \left|e_{\phi 1,T}(s,t)\right| + \left|e^{-2\phi T^2} e_{\phi 2,T}(s,t)\right| \\ \leq \; & \frac{3}{T} + \rho^{2T^3}\left|\frac{\left(2-\rho^{Ts}-\rho^{Tt}\right)\rho\left(1-\rho^T\right)}{T^2\left(1-\rho^2\right)\left(1-\rho\right)} - \frac{1-\rho^{T(s+t)}}{T\left(1-\rho^2\right)}\right|, \end{aligned}$$

where $\rho = e^{-\phi/T}$, $e_{\phi 1,T}(s,t)$ and $e_{\phi 2,T}(s,t)$ are as defined in (72) and (73), and we used (72), (75) and (78) for the last inequality. From (76) and (77), we see that the second term on the right hand side of the latter inequality is no larger than $2\rho^{2T^3}$. Therefore,

$$\left|f_\epsilon(\phi)\, e_{\phi,T}(s,t)\right| \leq \frac{3}{T} + 2\rho^{2T^3} \leq \frac{3}{T} + 2e^{-2T}$$

for $\rho \in \left[0, 1-1/T^2\right]$. On the other hand,

$$\begin{aligned} & \left|\frac{\left(2-\rho^{Ts}-\rho^{Tt}\right)\rho\left(1-\rho^T\right)}{T^2\left(1-\rho^2\right)\left(1-\rho\right)} - \frac{1-\rho^{T(s+t)}}{T\left(1-\rho^2\right)}\right| \\ = \; & \left|\frac{\left(1-\rho^{Ts}\right)\left(1-\rho^{Tt}\right)}{T\left(1-\rho^2\right)} - \frac{\left(2-\rho^{Ts}-\rho^{Tt}\right)\sum_{j=0}^{T}\left(T-j\right)\rho^j}{T^2\left(1+\rho\right)}\right| \\ \leq \; & \frac{\left(1-\rho^T\right)^2}{T\left(1-\rho^2\right)} + \frac{2\left(1-\rho^T\right)}{1+\rho} \leq 2T\left(1-\rho\right). \end{aligned}$$

Therefore, for $\rho \in \left(1-1/T^2, 1\right]$,

$$\left|f_\epsilon(\phi)\, e_{\phi,T}(s,t)\right| \leq \frac{3}{T} + 2T\left(1-\rho\right)\rho^{2T^3} \leq \frac{5}{T}.$$

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

Turning to the analysis of $y_3(s)$, let us define bounded linear functionals

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

similar to Anselone (1967, p.9). Functionals ψ_T converge to ψ uniformly on totally bounded subsets of $C[0,1]$. We have

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

where

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

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

To summarize, functions y_1, y_2, y_3 converge to zero as $N, T \to \infty$. Hence, there exists N_3, T_0 such that for all $N > N_3$ and $T > T_0$, $\|P_3\|_{\sup} < \epsilon/3$. Combining this with (79) and (80), and setting $N_0 = \max\{N_2, N_3\}$, we see that, for all $N > N_0$ and $T > T_0$, $\|K_{T,N}x - K_{\mathcal{F}}x\|_{\sup} < \epsilon$, which finishes the proof of the pointwise convergence $K_{T,N} \to K_{\mathcal{F}}$.

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

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

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

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

Convergence of the principal eigenvalues and eigenfunctions Recall that we denote the eigenvalues of $K_{\mathcal{F}}$ as $\mu_1, \mu_2, ...$ and corresponding eigenfucntions as $\varphi_1, \varphi_2, ...$. By assumption A5, these eigenvalues are simple so that $\mu_1 > \mu_2 > ...$ Denote the eigenvalues of $K_{N,T}$ as $\mu_{1,NT} \geq \mu_{2,NT} \geq ...$ and corresponding eigenfunctions as $\varphi_{1,NT}, \varphi_{2,NT}, ...$ Let us show that, for any fixed k, $\mu_{k,NT} \to \mu_k$ and $\varphi_{k,NT} \to \varphi_k$, the latter convergence being in $C[0,1]$.[1]

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

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

3.2.5 Proof of Lemma 6

First, let us prove the following lemma. Let $\mathbf{w}$ be the $T\times T$ orthogonal matrix with t-th column w_t, where w_t are as defined in Lemma 1. Namely, for $t<T$, w_t is a vector with s-th coordinate $w_{ts}=-\sqrt{2/T}\cos\left(\left(s-1/2\right)\pi t/T\right)$, while $w_T=l/\sqrt{T}$. Here l is the T-dimensional vector of ones.

Let U_ϕ be a $T\left(T^2+1\right)\times T$ matrix such that

$$U_\phi'=\begin{pmatrix} e^{-\phi T^2} & ... & e^{-\phi/T} & 1 & 0 & ... & 0\\ e^{-\phi(T^2+1/T)} & ... & e^{-2\phi/T} & e^{-\phi/T} & 1 & ... & 0\\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots\\ e^{-\phi(T^2+1-1/T)} & ... & e^{-\phi} & e^{-\phi(1-1/T)} & e^{-\phi(1-2/T)} & ... & 1\end{pmatrix}.$$

Lemma 8 *For any $\phi\geq 0$,*

$$\mathbf{w}'MU_\phi'U_\phi M\mathbf{w}=D_\phi-\Delta_\phi,$$

where D_ϕ is a diagonal matrix with t-th diagonal element equal to $\left|1-\exp\left\{\left(\mathrm{i}\pi t-\phi\right)/T\right\}\right|^{-2}$ if $t<T$ and zero if $t=T$; and Δ_ϕ is a positive semi-definite matrix of rank two with t,s-th entry

$$\begin{aligned}\Delta_{\phi,ts} &= \frac{2}{T}e^{-\phi/T}\left(1-e^{-\phi/T}\right)\frac{\cos\left(\pi t/2T\right)}{\left|e^{(-\phi+\mathrm{i}\pi t)/T}-1\right|^2}\frac{\cos\left(\pi s/2T\right)}{\left|e^{(-\phi+\mathrm{i}\pi s)/T}-1\right|^2}\\ &\quad\times\left(\frac{1+e^{-\phi/T}e^{-2\phi T^2}}{1+e^{-\phi/T}}\left(1-(-1)^t e^{-\phi}\right)\left(1-(-1)^s e^{-\phi}\right)+(-1)^{t+s}\frac{1-e^{-2\phi}}{1+e^{-\phi/T}}\right).\end{aligned}$$

Proof: Let us partition U_ϕ into the upper $T^3\times T$ submatrix $U_\phi^{(1)}$ and the lower $T\times T$ matrix $U_\phi^{(2)}$. We have

$$U_\phi^{(1)}=\left(e^{-\phi T^2},...,e^{-2\phi/T},e^{-\phi/T}\right)'v_1',$$

where v_1 is the T-dimensional vector with t-th coordinate $v_{1t}=e^{-\phi(t-1)/T}$. Obviously,

$$\mathbf{w}'MU_\phi^{(1)\prime}U_\phi^{(1)}M\mathbf{w}=0\text{ for }\phi=0. \tag{82}$$

For $\phi>0$,

$$\mathbf{w}'MU_\phi^{(1)\prime}U_\phi^{(1)}M\mathbf{w}=\frac{1-e^{-2\phi T^2}}{e^{2\phi/T}-1}x_1x_1', \tag{83}$$

where $x_1 = \mathbf{w}' M v_1$.

Next, note that

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

and therefore,

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

It is straightforward to verify that w_t, $t = 1, ..., T$ are eigenvectors of

$$\left(U_\phi^{(2)\prime} U_\phi^{(2)}\right)^{-1} - e^{-\phi/T} e_1 e_1' - e^{-\phi/T}\left(1 - e^{-\phi/T}\right) e_T e_T' - \mathbf{1}_{\phi=0} l l'/T$$

with corresponding eigenvalues equal to $d_{\phi t}^{-1} = \left|1 - \exp\left\{(-\phi + \mathrm{i}\pi t)/T\right\}\right|^2$ for $t < T$ and $d_{\phi T}^{-1} = \left|1 - \exp\left\{-\phi/T\right\}\right|^2 - \mathbf{1}_{\phi=0}$. Here e_t denotes the t-th column of the T-dimensional identity matrix, and $\mathbf{1}_{\phi=0}$ is the indicator of the event $\phi = 0$.

Let $\bar{D}_\phi = \operatorname{diag}\left\{d_{\phi 1}, ..., d_{\phi,T}\right\}$. Then

$$\left(\left(U_\phi^{(2)\prime} U_\phi^{(2)}\right)^{-1} - e^{-\phi/T} e_1 e_1' - e^{-\phi/T}\left(1 - e^{-\phi/T}\right) e_T e_T' - \mathbf{1}_{\phi=0} l l'/T\right)^{-1} = \mathbf{w}\bar{D}_\phi \mathbf{w}'$$

Applying the Sherman-Morrison formula for the inverse of a low rank perturbation of an invertible matrix to the left hand side of the above equality yields, for $\phi > 0$,

$$\mathbf{w}\bar{D}_\phi \mathbf{w}' = U_\phi^{(2)\prime} U_\phi^{(2)} + \frac{e^{-\phi/T}}{1 - e^{-\phi/T}} v_1 v_1' + \frac{e^{-2\phi}}{\left(1 - e^{-2\phi/T}\right)\left(1 - e^{-2\phi}\right)} v_2 v_2', \tag{84}$$

where v_1 is as defined above, and v_2 is the T-dimensional vector with t-th coordinate

$$v_{2t} = e^{\phi/2T} e^{\phi(t-1)/T} + e^{-\phi/2T} e^{-\phi(t-1)/T}.$$

Similarly, for $\phi = 0$, the Sherman-Morrison formula yields

$$\mathbf{w}\bar{D}_\phi \mathbf{w}' = U_\phi^{(2)\prime} U_\phi^{(2)} + lA + Bl', \tag{85}$$

where A and B are some matrices, exact form of which is of no consequence to what follows.

Mutiplying both sides of equation (84) by $\mathbf{w}'M$ from left and by $M\mathbf{w}$ from right and rearranging, we obtain

$$\begin{aligned}\mathbf{w}'MU_\phi^{(2)\prime}U_\phi^{(2)}M\mathbf{w} &= \mathbf{w}'M\mathbf{w}\bar{D}_\phi\mathbf{w}'M\mathbf{w} - \frac{e^{-\phi/T}}{1-e^{-\phi/T}}x_1x_1' \\ &\quad - \frac{e^{-2\phi}}{\left(1-e^{-2\phi/T}\right)\left(1-e^{-2\phi}\right)}x_2x_2',\end{aligned}$$

where $x_2 = \mathbf{w}'Mv_2$. Summing up with (83) yields

$$\begin{aligned}\mathbf{w}'MU_\phi'U_\phi M\mathbf{w} &= \mathbf{w}'M\mathbf{w}\bar{D}_\phi\mathbf{w}'M\mathbf{w} - \frac{e^{-2\phi(T^2+1/T)}+e^{-\phi/T}}{1-e^{-2\phi/T}}x_1x_1' \\ &\quad - \frac{e^{-2\phi}}{\left(1-e^{-2\phi/T}\right)\left(1-e^{-2\phi}\right)}x_2x_2'\end{aligned} \tag{86}$$

for $\phi > 0$.

Note that $\mathbf{w}'M\mathbf{w} = I_T - e_Te_T'$, where e_T denotes the last column of the T-dimensional identity matrix, so that $\mathbf{w}'M\mathbf{w}\bar{D}_\phi\mathbf{w}'M\mathbf{w} = D_\phi$. Further, a direct calculation shows that the t-th coordinates of x_1 and x_2 equal

$$\begin{aligned}x_{1t} &= -\sqrt{\frac{2}{T}}\frac{\left(1-e^{-\phi/T}\right)\left(1-(-1)^t e^{-\phi}\right)\cos\left(\pi t/2T\right)}{\left|e^{(-\phi+\mathrm{i}\pi t)/T}-1\right|^2}, \\ x_{2t} &= -\sqrt{\frac{2}{T}}\frac{\left(1-e^{-\phi/T}\right)(-1)^t\left(e^{\phi}-e^{-\phi}\right)e^{-\phi/2T}\cos\left(\pi t/2T\right)}{\left|e^{(-\phi+\mathrm{i}\pi t)/T}-1\right|^2}.\end{aligned}$$

For $\phi > 0$, the lemma now follows from (86) by verifying that

$$\frac{e^{-2\phi(T^2+1/T)}+e^{-\phi/T}}{1-e^{-2\phi/T}}x_1x_1' + \frac{e^{-2\phi}}{\left(1-e^{-2\phi/T}\right)\left(1-e^{-2\phi}\right)}x_2x_2' = \Delta_\phi.$$

For $\phi = 0$, mutiplying both sides of equation (85) by $\mathbf{w}'M$ from left and by $M\mathbf{w}$ from right and rearranging, we obtain $\mathbf{w}'MU_\phi^{(2)\prime}U_\phi^{(2)}M\mathbf{w} = \mathbf{w}'M\mathbf{w}\bar{D}_\phi\mathbf{w}'M\mathbf{w}$. Summing this up with (82) yields

$$\mathbf{w}'MU_\phi'U_\phi M\mathbf{w} = \mathbf{w}'M\mathbf{w}\bar{D}_\phi\mathbf{w}'M\mathbf{w} = D_\phi.$$

This establishes the lemma for $\phi = 0$ because, as is easy to see, $\Delta_\phi = 0$ for $\phi = 0$. □

Let us now turn to the proof of Lemma 6. By definition of $\mathbb{E}\tilde{\Sigma}$ and Lemma 8,

$$\sum\nolimits_{j=J+1}^{T}\tilde{\mu}_j \leq \frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=J+1}^{T}D_{\phi_i,jj}\Omega_{ii}.$$

On the other hand,

$$\begin{aligned}\sum\nolimits_{j=J+1}^{T} D_{\phi_i,jj} &= \sum\nolimits_{j=J+1}^{T-1} \left|1-\exp\left\{\left(-\phi_i+\mathrm{i}\pi j\right)/T\right\}\right|^{-2} \\ &\le \sum\nolimits_{j=J+1}^{\infty} \frac{T^2}{\phi_i^2+\pi^2 j^2} + o\left(T^2\right) \\ &\le \sum\nolimits_{j=J+1}^{\infty} \frac{T^2}{\pi^2 j^2} + o\left(T^2\right) \\ &\le \frac{T^2}{\pi^2 J} + o\left(T^2\right) \le \frac{T^2}{9J}\end{aligned}$$

for all sufficiently large T, uniformly over $\phi_i \geq 0$. Therefore,

$$\sum\nolimits_{j=J+1}^{T} \tilde{\mu}_j \le \frac{T^2}{N} \sum\nolimits_{i=1}^{N} \frac{1}{9J} \Omega_{ii} = \frac{T^2}{9JN} \operatorname{tr} \Omega.$$

The lemma's second inequality is a straightforward consequence of the convergence $\tilde{\mu}_k/T^2 \to \mu_k > 0$ and the fact that, as implied by A5, $\operatorname{tr} \Omega/N$ is converging to a positive value as $N \to \infty$.

3.3 Demeaned and standardized data

3.3.1 Proof of Theorem OW4

First, we prove the theorem for $k = 1$, and then establish it for general k using mathematical induction. For the demeaned and standardized case, $\hat{F}_1$ is defined as a normalized eigenvector of

$$\hat{\Sigma} = MX'D^{-1}XM/N,$$

corresponding to its largest eigenvalue $\hat{\lambda}_1$. Here $D = \operatorname{diag}\{XMX'/T\}$ and M is the projector on the space orthogonal to the T-dimensional vector of ones.

In contrast to the proof of Theorem OW1, we will not approximate $\hat{\Sigma}$ by $\tilde{\Sigma}$, where the latter matrix is derived from the Beveridge-Nelson decomposition

$$XM = \Psi(1)\varepsilon UM + \Psi^*(L)\,\varepsilon M.$$

In fact, we will not be using the BN decomposition at all. There are two reasons for this. First, Lemma 4 cannot be applied to the standardized version of $\Psi^*(L)\,\varepsilon M$, that is, $D^{-1/2}\Psi^*(L)\,\varepsilon M$. Second, even if we manage to reduce the analysis of $D^{-1/2}XM$ to that of $D^{-1/2}\Psi(1)\varepsilon UM$, our method of handling $\Psi(1)\varepsilon UM$ would not extend to $D^{-1/2}\Psi(1)\varepsilon UM$, because D and ε are not independent. To summarize, we are not going to use the BN decomposition, and will work directly with the demeaned and standardized data $D^{-1/2}XM = D^{-1/2}eUM$, where $e = [e_1, ..., e_T]$ with $e_t = \Psi(L)\,\varepsilon_t$.

Recall that by Lemma 1, $UM = \sum\nolimits_{q=1}^{T} \sigma_q v_q w_q'$. Consider a representation of $\hat{F}_1$ in the basis

$w_1, ..., w_T$

$$\hat{F}_1 = \sum_{q=1}^{T-1} \alpha_q w_q \tag{87}$$

Vector $\hat{F}_1$ is orthogonal to w_T, hence summation runs up to $q = T - 1$. Representation (87) yields

$$\hat{\lambda}_1 = \sum_{k,q=1}^{T-1} \alpha_k \alpha_q w_k' \hat{\Sigma} w_q = \sum_{k,q=1}^{T-1} \alpha_k \alpha_q \sigma_k \sigma_q v_k' e' D^{-1} e v_q / N. \tag{88}$$

It is convenient to represent $\hat{\lambda}_1$ in the form $\hat{\lambda}_1 = A'A$, where

$$A = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \alpha_k \sigma_k D^{-1/2} e v_k + \frac{1}{\sqrt{N}} \sum_{k=K+1}^{T-1} \alpha_k \sigma_k D^{-1/2} e v_k = A_1 + A_2$$

with K being a fixed positive integer. Let $e_{j\cdot}$ denote the j-th row of e. Then, we have the following explicit expressions for $\|A_1\|^2$ and $\|A_2\|^2$.

$$\|A_1\|^2 = T \sum_{k,q=1}^{K} \alpha_k \alpha_q \frac{1}{N} \sum\nolimits_{j=1}^{N} \frac{\sigma_k \sigma_q (e_{j\cdot} v_k)(e_{j\cdot} v_q)}{\sum\nolimits_{t=1}^{T-1} \sigma_t^2 (e_{j\cdot} v_t)^2} \text{ and} \tag{89}$$

$$\|A_2\|^2 = \frac{T}{N} \sum_{j=1}^{N} \frac{\left(\sum\nolimits_{k=K+1}^{T-1} \alpha_k \sigma_k e_{j\cdot} v_k \right)^2}{\sum\nolimits_{t=1}^{T-1} \sigma_t^2 (e_{j\cdot} v_t)^2}. \tag{90}$$

Let

$$M_{j,kq} = \sigma_k \sigma_q (e_{j\cdot} v_k)(e_{j\cdot} v_q) / \sum\nolimits_{t=1}^{T-1} \sigma_t^2 (e_{j\cdot} v_t)^2.$$

Then

$$\|A_1\|^2 = T \sum_{k,q=1}^{K} \alpha_k \alpha_q \frac{1}{N} \sum\nolimits_{j=1}^{N} M_{j,kq},$$

and by A2b, $M_{j,kq}$ are independent for different $j = 1, ..., N$. Moreover, since

$$\left| \sigma_k \sigma_q v_k' e_{j\cdot}' e_{j\cdot} v_q \right| \leq \frac{\sigma_k^2 v_k' e_{j\cdot}' e_{j\cdot} v_k + \sigma_q^2 v_q' e_{j\cdot}' e_{j\cdot} v_q}{2},$$

we have $|M_{j,kq}| \leq 1/2$. Therefore, the variance of $\frac{1}{N} \sum\nolimits_{j=1}^{N} M_{j,kq}$ is no larger than $1/(4N)$, and thus, the asymptotic behavior of $\|A_1\|^2$ is, to a large extent, determined by that of $\frac{1}{N} \sum\nolimits_{j=1}^{N} \mathbb{E} M_{j,kq}$.

Consider the finite Fourier transform of $e_{j\cdot}$ (e.g. Brillinger (2001, ch. 3.1))

$$d(\varpi) = \sum\nolimits_{t=1}^{T} e_{jt} \exp\{-\mathrm{i}(t-1)\varpi\}, \varpi \in [0, 2\pi].$$

Let us denote $d(\varpi_r/2)$ as d_r and $d(-\varpi_r/2)$ as d_{-r}, where $\varpi_r = 2\pi r/T$. By definition (see Lemma

1), the t-th entry of v_r for $r = 1, ..., T-1$ equals

$$v_{rt} = \sqrt{2/T}\left(\exp\left\{\mathrm{i}\left(t-1\right)\varpi_r/2\right\} - \exp\left\{-\mathrm{i}\left(t-1\right)\varpi_r/2\right\}\right)/\left(2\mathrm{i}\right).$$

Therefore,

$$e_{j\cdot}v_r = \sqrt{2/T}\left(d_{-r} - d_r\right)/\left(2\mathrm{i}\right). \tag{91}$$

Theorem 13, ch.4 of Hannan (1970) (one of the assumptions of this theorem requires that the spectral density of $e_{j\cdot}$ at zero is positive, which is ensured by A2b), identity (91), and the definition of σ_i imply that, for any fixed j, k, and q, as $T \to \infty$, $M_{j,kq} \xrightarrow{d} \mathcal{M}_{kq}$, where

$$\mathcal{M}_{kq} = \left(kq\right)^{-1}\eta_k\eta_q/\sum\nolimits_{t=1}^{\infty}\left(t\right)^{-2}\eta_t^2.$$

and $\{\eta_k\}_{k=1}^{\infty}$ is a sequence of i.i.d. $N(0,1)$ random variables. Since $M_{j,kq}$ is bounded, the convergence in disitribution implies the convergence of the moments of $M_{j,kq}$. In particular, as $T \to \infty$

$$\mathbb{E}M_{j,kq} = \mathbb{E}\mathcal{M}_{kq} + o_j(1). \tag{92}$$

To proceed further, we need to establish the uniformity of $o_j(1)$ in $j = 1, ..., N$.

In preparation for the proof of the uniformity, we establish some bounds on the spectral density of the series e_{jt}, $t \in \mathbb{Z}$ at frequency ϖ,

$$f_j\left(\varpi\right) = \frac{1}{2\pi}\left|\sum\nolimits_{k=0}^{\infty}\left(\Psi_k\right)_{jj}\exp\left\{\mathrm{i}k\varpi\right\}\right|^2. \tag{93}$$

By assumption A2b, for all j,

$$\max_{\varpi}\left|f_j\left(\varpi\right)\right| \le B^2/\left(2\pi\right). \tag{94}$$

Furthermore, differentiating both sides of (93) with respect to ϖ, we obtain

$$f_j\left(\varpi\right)' = \frac{1}{2\pi}\sum\nolimits_{k,r=0}^{\infty}\mathrm{i}\left(k-r\right)\left(\Psi_k\right)_{jj}\left(\Psi_r\right)_{jj}\exp\left\{\mathrm{i}k\varpi - \mathrm{i}r\varpi\right\}.$$

Since $|k-r| \le (k+1)(r+1)$, we conclude, using A2b, that for all j,

$$\max_{\varpi}\left|f_j'\left(\varpi\right)\right| \le B^2/\left(2\pi\right). \tag{95}$$

Finally, A2b also implies that, for all j,

$$f_j\left(0\right) \ge b^2/\left(2\pi\right). \tag{96}$$

We will need the following two lemmas. Their proofs can be found in Sections 3.3.2 and 3.3.3.

Lemma 9 *Under the assumptions of Theorem OW4, there exists an absolute constant C such that, for any $j = 1, ..., N$ and any $q, r, p, l = 1, ..., T-1$, we have*

(i) $\left|\mathbb{E}\left(v_q' e_{j\cdot}' e_{j\cdot} v_r\right) - 2\pi f_j\left(\varpi_q/2\right)\delta_{qr}\right| \leq CB^2/T$, *where* δ_{qr} *is the Kronecker delta and* $\varpi_q = 2\pi q/T$;
(ii) $\left|Cov\left(v_q' e_{j\cdot}' e_{j\cdot} v_r, v_p' e_{j\cdot}' e_{j\cdot} v_l\right)\right| \leq C\left(\delta_{qp}\delta_{rl} + \delta_{ql}\delta_{rp} + \left(1+\varkappa_4\right)/T\right)B^4$, *where* $\varkappa_4$ *is as defined in A1.*

Lemma 10 *Let* $\mathbf{X}$ *be an* R*-dimensional vector with the* k*-th coordinate* $e_{j\cdot} v_k$ *and let* $\mathbf{Y}$ *be an* R*-dimensional vector with i.i.d. normal coordinates with mean zero and variance* $2\pi f_j(0)$*. Further, let* $g : \mathbb{R}^R \to \mathbb{R}$ *be a thrice continuously differentiable function with all derivatives up to and including the third order are bounded by absolute value by a constant* M_g*. Then, under assumptions A1 and A2b, we have, for all sufficiently large* T,

$$|\mathbb{E}g(\mathbf{X}) - \mathbb{E}g(\mathbf{Y})| \leq M_g C/\sqrt{T}$$

where C *depends only on* $R, B,$ *and* $\varkappa_4$, *with* $\varkappa_4$ *and* B *as defined in A1 and A2b.*

Now we are ready to prove the uniformity of $o_j(1)$ in (92). By definition,

$$M_{j,kq} = \frac{\sigma_k \sigma_q \mathbf{X}_k \mathbf{X}_q}{\sum_{t=1}^{R} \sigma_t^2 \mathbf{X}_t^2 + \mathbf{Z}},$$

where $\mathbf{Z} = \sum_{t=R+1}^{T-1} \sigma_t^2 v_t' e_{j\cdot}' e_{j\cdot} v_t$. For $\max\{k,q\} \leq R$, denote $\sigma_k\sigma_q\mathbf{X}_k\mathbf{X}_q / \sum_{t=1}^{R}\sigma_t^2\mathbf{X}_t^2$ as $\bar{M}_{j,kq}$.

Consider the event $\mathcal{E} = \left\{\mathbf{X}_1^2 \leq \delta\right\}$ and let $\mathbf{1}_{\mathcal{E}}$ and $\mathbf{1}_{\mathcal{E}^c}$ be the indicators of this event and of its complement, respectively. Since $|M_{j,kq}| \leq 1/2$ and $\left|\bar{M}_{j,kq}\right| \leq 1/2$, we have

$$\mathbb{E}\left[\left|M_{j,kq} - \bar{M}_{j,kq}\right| \times \mathbf{1}_{\mathcal{E}}\right] \leq p_\delta = \Pr(\mathcal{E}).$$

By setting function g in Lemma 10 so that it approximates $\mathbf{1}_{\mathcal{E}}$, we see that p_δ can be made arbitrarily small for all sufficiently large T unifomly in j by choosing δ sufficiently small. On the other hand,

$$\mathbb{E}\left[\left|M_{j,kq} - \bar{M}_{j,kq}\right| \times \mathbf{1}_{\mathcal{E}^c}\right] = \mathbb{E}\left[\frac{|M_{j,kq}|\,\mathbf{Z}\mathbf{1}_{\mathcal{E}^c}}{\sum_{t=1}^{R}\sigma_t^2\mathbf{X}_t^2}\right] \leq \frac{\mathbb{E}\mathbf{Z}}{2\delta\sigma_1^2}.$$

By Lemma 9 (i) and by (94),

$$\mathbb{E}\mathbf{Z} \leq \sum_{t=R+1}^{T-1} \sigma_t^2\left(B + CB^2/T\right) \leq \tilde{C}\sigma_1^2/R$$

for some absolute constant $\tilde{C}$, where the latter inequality follows from the definition of σ_t^2. Therefore,

$$\mathbb{E}\left[\left|M_{j,kq} - \bar{M}_{j,kq}\right| \times \mathbf{1}_{\mathcal{E}^c}\right] \leq \tilde{C}/\left(2\delta R\right),$$

and

$$\begin{aligned}\mathbb{E}\left|M_{j,kq}-\bar{M}_{j,kq}\right| &= \mathbb{E}\left[\left|M_{j,kq}-\bar{M}_{j,kq}\right|\times \mathbf{1}_{\mathcal{E}}\right]+\mathbb{E}\left[\left|M_{j,kq}-\bar{M}_{j,kq}\right|\times \mathbf{1}_{\mathcal{E}^c}\right]\\ &\leq p_\delta+\tilde{C}/\left(2\delta R\right),\end{aligned}$$

which can be made arbitrarily small for all sufficiently large T uniformly in j by choosing sufficiently small δ and sufficiently large R.

Further, let $\bar{\mathcal{M}}_{kq}=\sigma_k\sigma_q\eta_k\eta_q/\sum_{t=1}^{R}\sigma_t^2\eta_t^2$. By choosing R sufficiently large, we can make $\mathbb{E}\left|\bar{\mathcal{M}}_{kq}-\mathcal{M}_{kq}\right|$ arbitrarily small for all sufficiently large T.

Now consider $\mathbb{E}\bar{M}_{j,kq}-\mathbb{E}\bar{\mathcal{M}}_{kq}$. To bound this expression uniformly in j, we would like to use Lemma 10 again. Unfortunately, $\bar{M}_{j,kq}$ does not have bounded derivatives as a function of $\mathbf{X}=(\mathbf{X}_1,...,\mathbf{X}_R)'$. The derivatives are unbounded in a neighborhood of $\mathbf{X}=0$.

To overcome this difficulty, let us introduce $\tau:[0,\infty)\to\mathbb{R}$, a thrice continuously differential function such that

$$\begin{aligned}\tau(z) &= z \text{ for } z>\delta,\\ \tau(z) &> \delta/2 \text{ for } z\geq 0,\end{aligned}$$

and the first three derivatives of $\tau(z)$ bounded for $z\in[0,\delta]$. Further, let

$$\tilde{M}_{j,kq}=\frac{\sigma_k\sigma_q\mathbf{X}_k\mathbf{X}_q}{\sigma_1^2\tau\left(\mathbf{X}_1^2\right)+\sum_{t=2}^{R}\sigma_t^2\mathbf{X}_t^2}=g\left(\mathbf{X}\right).$$

Similarly, let $\tilde{\mathcal{M}}_{kq}=g\left(\eta\right)$, where $\eta=(\eta_1,...,\eta_R)'$. Note that the derivatives $\partial_i^r g(x)$ for $r=1,2,3$ are bounded, with a bound that depends only δ, but not on R.

We have

$$\begin{aligned}\bar{M}_{j,kq} &= \bar{M}_{j,kq}\mathbf{1}_{\mathcal{E}}+\bar{M}_{j,kq}\mathbf{1}_{\mathcal{E}^c}=\bar{M}_{j,kq}\mathbf{1}_{\mathcal{E}}+\tilde{M}_{j,kq}\mathbf{1}_{\mathcal{E}^c}\\ &= \left(\bar{M}_{j,kq}-\tilde{M}_{j,kq}\right)\mathbf{1}_{\mathcal{E}}+\tilde{M}_{j,kq}.\end{aligned}$$

Therefore

$$\left|\mathbb{E}\bar{M}_{j,kq}-\mathbb{E}\tilde{M}_{j,kq}\right|=\left|\mathbb{E}\left[\left(\bar{M}_{j,kq}-\tilde{M}_{j,kq}\right)\mathbf{1}_{\mathcal{E}}\right]\right|\leq p_\delta,$$

so that $\left|\mathbb{E}\bar{M}_{j,kq}-\mathbb{E}\tilde{M}_{j,kq}\right|$ can be made arbitrarily small, uniformly over j, by choosing sufficiently small δ. By similar arguments, we can show that $\left|\mathbb{E}\tilde{\mathcal{M}}_{kq}-\mathbb{E}\bar{\mathcal{M}}_{kq}\right|$ can be made arbitrarily small by choosing sufficiently small δ. Finally, $\left|\mathbb{E}\tilde{M}_{j,kq}-\mathbb{E}\tilde{\mathcal{M}}_{kq}\right|$ can be made arbitrarily small uniformly over j for all sufficiently large T by Lemma 10. Summing up the above arguments, we conclude that $o_j(1)$ in (92) is uniform in j.

By (92) and Chebyshev's inequality,

$$\frac{1}{N}\sum\nolimits_{j=1}^{N} M_{j,kq} = \mathbb{E}\mathcal{M}_{kq} + o(1) + O_{\mathrm{P}}\left(N^{-1/2}\right).$$

Furthermore, by a conditioning argument, it is easy to show that $\mathbb{E}\mathcal{M}_{kq} = 0$ for $k \neq q$. Now recall

$$\|A_1\|^2 = T\sum_{k,q=1}^{K} \alpha_k\alpha_q \frac{1}{N}\sum\nolimits_{j=1}^{N} M_{j,kq}.$$

Therefore, we have

$$\frac{1}{T}\|A_1\|^2 = \sum_{k=1}^{K} \alpha_k^2 \mathbb{E}\mathcal{M}_{kk} + o(1) + O_{\mathrm{P}}\left(N^{-1/2}\right).$$

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

$$\begin{aligned}
\|A_2\|^2/T &= \frac{1}{N}\sum_{j=1}^{N} \frac{\left(\sum\nolimits_{k=K+1}^{T-1} \alpha_k\sigma_k e_{j\cdot}v_k\right)^2}{\sum\nolimits_{t=1}^{T-1} \sigma_t^2 v_t' e_{j\cdot}' e_{j\cdot} v_t} \\
&\le \left(1 - \sum\nolimits_{i=1}^{K} \alpha_i^2\right) \frac{1}{N}\sum_{j=1}^{N} \frac{\sum\nolimits_{k=K+1}^{T-1} \sigma_k^2 \left(e_{j\cdot}v_k\right)^2}{\sum\nolimits_{t=1}^{T-1} \sigma_t^2 \left(e_{j\cdot}v_t\right)^2}.
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\frac{\sum\nolimits_{k=K+1}^{T-1} \sigma_k^2 \left(e_{j\cdot}v_k\right)^2}{\sum\nolimits_{t=1}^{T-1} \sigma_t^2 \left(e_{j\cdot}v_t\right)^2} &= 1 - \mathbb{E}\frac{\sum\nolimits_{k=1}^{K} \sigma_k^2 \left(e_{j\cdot}v_k\right)^2}{\sum\nolimits_{t=1}^{T-1} \sigma_t^2 \left(e_{j\cdot}v_t\right)^2} = 1 - \sum_{k=1}^{K} \mathbb{E}\mathcal{M}_{kk} + o(1) \\
&= \delta_K + o(1),
\end{aligned}$$

where δ_K can be made arbitrarily small by choosing sufficiently large K, and $o(1)$ is uniform in j, but may depend on K. Therefore,

$$\|A_2\|^2/T \le \left(1 - \sum\nolimits_{i=1}^{K} \alpha_i^2\right)\left(\delta_K + o(1) + O_{\mathrm{P}}\left(N^{-1/2}\right)\right).$$

For $\hat{\lambda}_1$, we have

$$\hat{\lambda}_1 = \|A\|^2 \le \|A_1\|^2 + 2\|A_1\|\|A_2\| + \|A_2\|^2.$$

This inequality and the above bounds on $\|A_1\|^2/T$ and $\|A_2\|^2/T$ yield

$$\hat{\lambda}_1/T \le \sum_{k=1}^{K} \alpha_k^2 \mathbb{E}\mathcal{M}_{kk} + \Delta_K + o_{\mathrm{P}}(1),$$

where Δ_K can be made arbitrarily small by choosing sufficiently large K, and the convergence of $o_{\mathrm{P}}(1)$ to 0 as $T \to \infty$ may depend on the choice of K. This implies that

$$\hat{\lambda}_1/T \leq \sum_{k=1}^{T-1} \alpha_k^2 \mathbb{E}\mathcal{M}_{kk} + o_{\mathrm{P}}(1).$$

It is easy to see that $\mathbb{E}\mathcal{M}_{aa} > \mathbb{E}\mathcal{M}_{bb}$ for $a < b$. Indeed, consider functions $\mathcal{H}_a, \mathcal{H}_b : [0,\infty)^2 \to \mathbb{R}$

$$\begin{aligned} \mathcal{H}_a(x,y) &= \frac{x\eta_a^2}{\sum_{t\geq 1, t\neq a,b} (t)^{-2}\eta_t^2 + x\eta_a^2 + y\eta_b^2}, \\ \mathcal{H}_b(x,y) &= \frac{x\eta_b^2}{\sum_{t\geq 1, t\neq a,b} (t)^{-2}\eta_t^2 + x\eta_b^2 + y\eta_a^2} \end{aligned}$$

Functions $\mathcal{H}_a$ and $\mathcal{H}_b$ are increasing in x and decreasing in y and this monotonicity is strict (unless $\eta_a^2 = 0$ or $\eta_b^2 = 0$, which is a zero probability event). Therefore, with probability one, for $a < b$, we have

$$\begin{aligned} \mathbb{E}\mathcal{M}_{aa} &= \mathbb{E}\mathcal{H}_a(a^{-2}, b^{-2}) > \mathbb{E}\mathcal{H}_a\left(\frac{a^{-2}+b^{-2}}{2}, \frac{a^{-2}+b^{-2}}{2}\right) \text{ and} \\ \mathbb{E}\mathcal{M}_{bb} &= \mathbb{E}\mathcal{H}_b(b^{-2}, a^{-2}) < \mathbb{E}\mathcal{H}_b\left(\frac{a^{-2}+b^{-2}}{2}, \frac{a^{-2}+b^{-2}}{2}\right). \end{aligned}$$

On the other hand,

$$\mathbb{E}\mathcal{H}_a\left(\frac{a^{-2}+b^{-2}}{2}, \frac{a^{-2}+b^{-2}}{2}\right) = \mathbb{E}\mathcal{H}_b\left(\frac{a^{-2}+b^{-2}}{2}, \frac{a^{-2}+b^{-2}}{2}\right)$$

Therefore, $\mathbb{E}\mathcal{M}_{aa} > \mathbb{E}\mathcal{M}_{bb}$. This implies that

$$\sum_{k=1}^{T-1} \alpha_k^2 \mathbb{E}\mathcal{M}_{kk} \leq \alpha_1^2 \mathbb{E}\mathcal{M}_{11} + \left(1-\alpha_1^2\right)\mathbb{E}\mathcal{M}_{22}$$

and thus,

$$\hat{\lambda}_1/T \leq \alpha_1^2 \mathbb{E}\mathcal{M}_{11} + \left(1-\alpha_1^2\right)\mathbb{E}\mathcal{M}_{22} + o_{\mathrm{P}}(1). \tag{97}$$

On the other hand, $\hat{\lambda}_1$ must be no smaller than $v_1'\hat{\Sigma}v_1$, which yields

$$\hat{\lambda}_1/T \geq \mathbb{E}\mathcal{M}_{11} + o_{\mathrm{P}}(1). \tag{98}$$

Thus,

$$\left(1-\alpha_1^2\right)\mathbb{E}\mathcal{M}_{11} \leq \left(1-\alpha_1^2\right)\mathbb{E}\mathcal{M}_{22} + o_{\mathrm{P}}(1),$$

which only holds if

$$\alpha_1^2 \xrightarrow{\mathrm{P}} 1. \tag{99}$$

This yields statement (i) of the theorem.

To establish statement (ii), note that (97) and (98) imply

$$\left|\hat{\lambda}_1/T-\mathbb{E}\mathcal{M}_{11}\right|\leq\left|1-\alpha_1^2\right|(\mathbb{E}\mathcal{M}_{11}+\mathbb{E}\mathcal{M}_{22})+o_{\mathrm{P}}(1).$$

Combining this with (99), we conclude that

$$\hat{\lambda}_1/T=\mathbb{E}\mathcal{M}_{11}+o_{\mathrm{P}}(1).$$

This yields (ii) because $\mathbb{E}\mathcal{M}_{11}=\nu_1$ (the latter being defined in the statement of Theorem OW4).

Further,

$$\operatorname{tr}\hat{\Sigma}=\operatorname{tr}\left(MX'D^{-1}XM/N\right)=\operatorname{tr}\left(D^{-1}XMX'/N\right).$$

But, by definition, $D=\operatorname{diag}\left\{XMX'/T\right\}.$ Therefore, $\operatorname{tr}\hat{\Sigma}=T$ and

$$\hat{\lambda}_1/T=\hat{\lambda}_1/\operatorname{tr}\hat{\Sigma},$$

which yields statement (iii) of the theorem.

For $k=m>1$, the theorem follows by mathematical induction. Indeed, suppose it holds for $k<m$. Consider a representation $\hat{F}_m=\sum_{q=1}^{T-1}\alpha_q w_q$. Since $\hat{F}_m'\hat{F}_j=0$ for all $j<m$, and since $\left|\hat{F}_j'w_j\right|=1+o_{\mathrm{P}}(1)$ by the induction hypothesis, we must have $\alpha_j=o_{\mathrm{P}}(1)$ for all $j<m$. In particular,

$$\hat{F}_m'\hat{\Sigma}\hat{F}_m=\sum\nolimits_{q,r=m}^{T-1}\alpha_q\alpha_r\sigma_q\sigma_r v_q'e'D^{-1}ev_r/N+o_{\mathrm{P}}(T).$$

In addition to this equality, we must have

$$\sum_{i=1}^{m-1}\hat{\lambda}_i+\hat{F}_m'\hat{\Sigma}\hat{F}_m\geq\sum_{i=1}^{m}w_i'MU'e'D^{-1}eUMw_i/N=T\sum_{i=1}^{m}\mathbb{E}\mathcal{M}_{ii}+o_{\mathrm{P}}(T).$$

Combining the above two displays, and using the induction hypothesis, this time regarding the validity of the identities $\hat{\lambda}_i=T\mathbb{E}\mathcal{M}_{ii}+o_{\mathrm{P}}(T)$ for all $i<m$, we obtain

$$\sum_{q,r=m}^{T-1}\alpha_q\alpha_r\sigma_q\sigma_r v_q'e'D^{-1}ev_r/N\geq T\mathbb{E}\mathcal{M}_{mm}+o_{\mathrm{P}}(T). \tag{100}$$

Statements (i), (ii), and (iii) for $k=m$ now follow by arguments that are very similar to those used above for the case $k=1$.

That is, we represent the sum on the left hand side of (100) in the form $A'A$, where $A=\frac{1}{\sqrt{N}}\sum_{k=1}^{T-1}\alpha_k\sigma_k D^{-1/2}ev_k$. Then proceed along the lines of the above proof to obtain an upper bound on $A'A$, similar to the right hand side of (97). Then, combining this upper bound with the lower bound (100), we prove the convergence $\alpha_m^2\xrightarrow{\mathrm{P}}1$. Finally, we proceed to establishing parts (ii) and (iii) using part (i). We omit further details to save space.

3.3.2 Proof of Lemma 9

Identity (91) yields

$$\mathbb{E}\left(v_q' e_{j\cdot}' e_{j\cdot} v_r\right) = -\mathbb{E}\left[\left(d_{-q} - d_q\right)\left(d_{-r} - d_r\right)\right] / (2T),$$

and

$$Cov\left(v_q' e_{j\cdot}' e_{j\cdot} v_r, v_p' e_{j\cdot}' e_{j\cdot} v_l\right) = \frac{1}{4T^2} Cov\left(\left(d_{-q} - d_q\right)\left(d_{-r} - d_r\right), \left(d_{-p} - d_p\right)\left(d_{-l} - d_l\right)\right).$$

To evaluate the latter expectation and covariance, we use Theorem 4.3.2 of Brillinger (2001) (B01), which describes joint cumulants of finite Fourier transforms. First, we need to represent the expectation and covariance in terms of the joint cumulants. By their definition, and by Theorem 2.3.1 (B01, p.19),

$$\mathbb{E}\left(v_q' e_{j\cdot}' e_{j\cdot} v_r\right) = -\frac{1}{2T} \sum\nolimits_{s_1, s_2 \in \{-1,+1\}} s_1 s_2 \operatorname{cum}\left(d_{s_1 q}, d_{s_2 r}\right). \tag{101}$$

Similarly, $Cov\left(v_q' e_{j\cdot}' e_{j\cdot} v_r, v_p' e_{j\cdot}' e_{j\cdot} v_l\right)$ equals

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

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

$$\begin{aligned} & Cov\left(v_q' e_{j\cdot}' e_{j\cdot} v_r, v_p' e_{j\cdot}' e_{j\cdot} v_l\right) = \frac{1}{4T^2} \sum\nolimits_{s_1, s_2, s_3, s_4 \in \{-1,+1\}} s_1 s_2 s_3 s_4 \\ & \times \left\{\operatorname{cum}\left(d_{s_1 q}, d_{s_3 p}\right) \operatorname{cum}\left(d_{s_2 r}, d_{s_4 l}\right) + \operatorname{cum}\left(d_{s_1 q}, d_{s_4 l}\right) \operatorname{cum}\left(d_{s_2 r}, d_{s_3 p}\right)\right. \\ & \left. + \operatorname{cum}\left(d_{s_1 q}, d_{s_2 r}, d_{s_3 p}, d_{s_4 l}\right)\right\}. \end{aligned} \tag{102}$$

Lemma 11 *Under assumptions of Theorem OW4, there exists an absolute constant C such that, for any $q, r, p, l = 1, ..., T-1$, and any $s_1, s_2, s_3, s_4 \in \{-1, +1\}$,*

$$\left|\operatorname{cum}\left(d_{s_1 q}, d_{s_2 r}\right) - 2\pi H_{s_1 q, s_2 r} f_j\left(\varpi_q / 2\right)\right| \leq C B^2, \tag{103}$$

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

$$\left|\operatorname{cum}\left(d_{s_1 q}, d_{s_2 r}, d_{s_3 p}, d_{s_4 l}\right) - (2\pi)^3 H_{s_1 q, s_2 r, s_3 p, s_4 l} f_{j4}\right| \leq C \varkappa_4 B^4, \tag{104}$$

where $H_{s_1 q, s_2 r, s_3 p, s_4 l} = \sum\nolimits_{t=0}^{T-1} e^{-\mathrm{i}t\left(s_1 \varpi_q + s_2 \varpi_r + s_3 \varpi_p + s_4 \varpi_l\right)/2}$ and f_{j4} is the 4-th order cumulant spectrum of the series e_{jt}, $t \in \mathbb{Z}$ at frequencies $s_1 \varpi_q / 2$, $s_2 \varpi_r / 2$, $s_3 \varpi_p / 2$.

Proof: The proof of Theorem 4.3.2 in B01 implies that the left hand side of (103) can be bounded by $C\sum_{k=0}^{\infty}(1+k)\left|\Gamma_j(k)\right|$, where C is an absolute constant and

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

Here $\theta_{jt} = (\Psi_t)_{jj}$ for $t \geq 0$ and $\theta_{jt} = 0$ for $t < 0$. On the other hand,

$$\begin{aligned}
&\sum_{k=0}^{\infty}(1+k)\left|\Gamma_j(k)\right| \leq \sum_{k=0}^{\infty}(1+k)\sum_{t=-\infty}^{\infty}\left|\theta_{jt}\right|\left|\theta_{j,t-k}\right| \\
\leq\ &\sum_{k=0}^{\infty}\sum_{t=-\infty}^{\infty}(1+|t-k|)\left|\theta_{jt}\right|\left|\theta_{j,t-k}\right| + \sum_{k=0}^{\infty}\sum_{t=-\infty}^{\infty}(1+|t|)\left|\theta_{jt}\right|\left|\theta_{j,t-k}\right| \leq 2B^2,
\end{aligned} \tag{105}$$

where the last inequality follows from assumption A2b. This yields (103).

Similarly, from the proof of Theorem 4.3.2 in B01, we know that the left hand side of (104) can be bounded by

$$C\sum_{k_1,k_2,k_3=-\infty}^{\infty}(1+|k_1|+|k_2|+|k_3|)\left|c_{j4}(k_1,k_2,k_3)\right|, \tag{106}$$

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

$$\begin{aligned}
&\sum_{t_1,t_2,t_3,t_4=-\infty}^{\infty}\theta_{j,t_1-k_1}\theta_{j,t_2-k_2}\theta_{j,t_3-k_3}\theta_{j,t_4}\operatorname{cum}\left(\varepsilon_{j,-t_1},\varepsilon_{j,-t_2},\varepsilon_{j,-t_3},\varepsilon_{j,-t_4}\right) \\
=\ &\sum_{t=-\infty}^{\infty}\theta_{j,t-k_1}\theta_{j,t-k_2}\theta_{j,t-k_3}\theta_{j,t}\left(\mathbb{E}\varepsilon_{j,-t}^4-3\right) \\
\leq\ &\sum_{t=-\infty}^{\infty}\left|\theta_{j,t-k_1}\theta_{j,t-k_2}\theta_{j,t-k_3}\theta_{j,t}\right|\varkappa_4,
\end{aligned}$$

where the last line follows from A1. By an argument similar to (105), expression (106) can be bounded by $C\varkappa_4B^4$, where C is an absolute constant. This yields (104). □

Returning to the proof of Lemma 9, consider (101). Inequality (103) implies that

$$\left|\mathbb{E}\left(v_q'e_{j\cdot}'e_{j\cdot}v_r\right) + \frac{1}{2T}\sum_{s_1,s_2\in\{-1,+1\}}s_1s_2H_{s_1q,s_2r}2\pi f_j\left(\varpi_q/2\right)\right| \leq \frac{2KB^2}{T}. \tag{107}$$

Further, for $q = r$,

$$\sum_{s_1,s_2\in\{-1,+1\}}s_1s_2H_{s_1q,s_2r} = \sum_{s_1,s_2\in\{-1,+1\}}s_1s_2\sum_{t=0}^{T-1}e^{-\mathrm{i}t(s_1+s_2)\pi q/T} = -2T. \tag{108}$$

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

$$\sum_{s_1,s_2\in\{-1,+1\}}s_1s_2H_{s_1q,s_2r} = \sum_{s_1,s_2\in\{-1,+1\}}s_1s_2\sum_{t=0}^{T-1}e^{-\mathrm{i}t(s_1q+s_2r)\pi/T} = 0. \tag{109}$$

Here, the latter equality holds because s_1q+s_2r is an even nonzero integer, such that $|s_1q+s_2r| <$

$2T$ (recall that $1 \le q, r \le T-1$). For $q \neq r$ and such that $s_1 q + s_2 r$ is odd for all $s_1, s_2 \in \{-1, +1\}$, we have

$$\sum\nolimits_{t=0}^{T-1} e^{-\mathrm{i}t(s_1 q + s_2 r)\pi/T} = \frac{-2}{e^{-\mathrm{i}(s_1 q + s_2 r)\pi/T} - 1}.$$

Nevertheless, $\sum\nolimits_{s_1, s_2 \in \{-1,+1\}} s_1 s_2 H_{s_1 q, s_2 r}$ still equals zero because

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

Therefore, (109) still holds. Using identities (108) and (109) in (107), we obtain statement (i) of Lemma 9.

Next, consider (102). By (103) and (104), the difference between

$$\begin{aligned}&\left(\frac{\pi}{T}\right)^2 \sum\nolimits_{s_1, s_2, s_3, s_4 \in \{-1,+1\}} s_1 s_2 s_3 s_4 \{2\pi H_{s_1 q, s_2 r, s_3 p, s_4 l} f_{j4} \\ &+ (H_{s_1 q, s_3 p} H_{s_2 r, s_4 l} + H_{s_1 q, s_4 l} H_{s_2 r, s_3 p}) f_j(\varpi_q/2) f_j(\varpi_r/2)\}\end{aligned}$$

and $Cov\left(v_q' e_{j\cdot}' e_{j\cdot} v_r, v_p' e_{j\cdot}' e_{j\cdot} v_l\right)$ is no larger by absolute value than

$$\begin{aligned}\frac{1}{4T^2} \quad &\sum\nolimits_{s_1, s_2, s_3, s_4 \in \{-1,+1\}} \{C\varkappa_4 B^4 + 2C^2 B^4 + 2\pi C B^2 (|H_{s_1 q, s_3 p} f_j(\varpi_q/2)| \\ &+ |H_{s_2 r, s_4 l} f_j(\varpi_r/2)| + |H_{s_1 q, s_4 l} f_j(\varpi_q/2)| + |H_{s_2 r, s_3 p} f_j(\varpi_r/2)|)\},\end{aligned}$$

which, in its turn, is bounded from above by $C(1 + \varkappa_4) B^4/T$, where C is an absolute constant (we remind the reader that throughout the paper, the value of the absolute constant C may change from one appearance to another). Indeed, such a bound follows from (94) and the fact that $|H_{a,b}| \le T$. Further, from the above analysis of $\mathbb{E}\left(v_q' e_{j\cdot}' e_{j\cdot} v_r\right)$,

$$\sum\nolimits_{s_1, s_2, s_3, s_4 \in \{-1,+1\}} s_1 s_2 s_3 s_4 (H_{s_1 q, s_3 p} H_{s_2 r, s_4 l} + H_{s_1 q, s_4 l} H_{s_2 r, s_3 p}) = 4T^2 (\delta_{qp}\delta_{rl} + \delta_{ql}\delta_{rp}).$$

Therefore, from (94),

$$\left|\sum\nolimits_{s_1, s_2, s_3, s_4 \in \{-1,+1\}} s_1 s_2 s_3 s_4 (H_{s_1 q, s_3 p} H_{s_2 r, s_4 l} + H_{s_1 q, s_4 l} H_{s_2 r, s_3 p}) f_j(\varpi_q/2) f_j(\varpi_r/2)\right|$$

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

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

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

$$\left|\sum\nolimits_{s_1, s_2, s_3, s_4 \in \{-1,+1\}} s_1 s_2 s_3 s_4 2\pi H_{s_1 q, s_2 r, s_3 p, s_4 l} f_{j4}\right| \le \frac{T B^4 \varkappa_4}{(2\pi)^2}.$$

Overall, we conclude that $\left|Cov\left(v_q' e_{j\cdot}' e_{j\cdot} v_r, v_p' e_{j\cdot}' e_{j\cdot} v_l\right)\right|$ is no larger than

$$\left(\frac{\pi}{T}\right)^2\left(\frac{TB^4\varkappa_4}{(2\pi)^2}+\frac{4T^2B^4\left(\delta_{qp}\delta_{rl}+\delta_{ql}\delta_{rp}\right)}{(2\pi)^2}\right)+\frac{C\left(1+\kappa_4\right)B^4}{T},$$

which yields statement (ii) of Lemma 9.

3.3.3 Proof of Lemma 10

Our proof is based on the following theorem, established in Chatterjee (2006).

Theorem 12 *(Chatterjee, 2006) Suppose x and y are random vectors in $\mathbb{R}^m$ with y having independent components. For $1 \leq i \leq m$, let*

$$\begin{aligned} R_i &:= \mathbb{E}\left|\mathbb{E}\left(x_i\left|x_1,...,x_{i-1}\right.\right)-\mathbb{E}\left(y_i\right)\right|, \\ B_i &:= \mathbb{E}\left|\mathbb{E}\left(x_i^2\left|x_1,...,x_{i-1}\right.\right)-\mathbb{E}\left(y_i^2\right)\right|. \end{aligned}$$

Let M_3 be a bound on $\max_i\left(\mathbb{E}\left|x_i\right|^3+\mathbb{E}\left|y_i\right|^3\right)$. Suppose $h:\mathbb{R}^m\to\mathbb{R}$ is a thrice continuously differentiable function, and for $r=1,2,3$, let $L_r(h)$ be a finite constant such that $\left|\partial_i^r h(z)\right|\leq L_r(h)$ for each i and z, where ∂_i^r denotes the r-fold derivative in the i-th coordinate. Then

$$\left|\mathbb{E}h\left(x\right)-\mathbb{E}h\left(y\right)\right|\leq\sum\nolimits_{i=1}^m\left(R_iL_1(h)+\frac{1}{2}B_iL_2(h)\right)+\frac{1}{6}mL_3(h)M_3.$$

Let us denote $\left(\Psi_k\right)_{jj}$ as θ_{jk}, as in the previous section. With this notation, we have

$$e_{jt}=\sum\nolimits_{k=0}^{\infty}\theta_{jk}\varepsilon_{j,t-k}.$$

Let

$$x_i=\left\{\begin{array}{cc}\varepsilon_{j,T+1-i} & \text{for } i=1,...,2T \\ \sum\nolimits_{k=2T+1-i}^{\infty}\theta_{j,k+2T}\varepsilon_{j,T+1-i-k} & \text{for } i=2T+1,...,3T\end{array}\right.$$

and $m=3T$. Then, for $t=1,...,T$, we have

$$e_{jt}=\sum\nolimits_{k=0}^{T+t-1}\theta_{jk}x_{k+T-t+1}+x_{3T-t+1}$$

so that the R-dimensional vector $\mathbf{X}$ with the k-th coordinate $e_{j\cdot}v_k$ can be thought of as a function $\mathbf{X}(x):\mathbb{R}^m\to\mathbb{R}^R$. Further, let

$$y_i=\left\{\begin{array}{cc}\text{i.i.d. } N(0,1) & \text{for } i=1,...,2T \\ 0 & \text{for } i=2T+1,...,3T\end{array}\right..$$

Since $x_i, i = 1, ..., 2T$, are independent, we have

$$R_i = B_i = 0 \text{ for } i = 1, ..., 2T. \tag{110}$$

For $i > 2T$, we have

$$\begin{aligned} R_i &= \mathbb{E}\left|\mathbb{E}\left(x_i \left| x_1, ..., x_{i-1}\right.\right)\right| \\ &\leq \left(\mathbb{E}\left[\mathbb{E}\left(x_i^2 \left| x_1, ..., x_{i-1}\right.\right)\right]\right)^{1/2} = \left[\mathbb{E}\left(x_i^2\right)\right]^{1/2} \end{aligned}$$

and

$$B_i = \mathbb{E}\left[\mathbb{E}\left(x_i^2 \left| x_1, ..., x_{i-1}\right.\right)\right] = \mathbb{E}\left(x_i^2\right).$$

On the other hand, for $i > 2T$,

$$\mathbb{E}\left(x_i^2\right) = \sum\nolimits_{k=2T+1-i}^{\infty} \theta_{j,k+2T}^2.$$

By assumption A2b, $\sum\nolimits_{k=0}^{\infty} (1+k) |\theta_{jk}| \leq B$. Therefore, $|\theta_{jk}| \leq B/(1+k)$ and, for $2T < i \leq 3T$,

$$\mathbb{E}\left(x_i^2\right) \leq \frac{B}{4T+1-i} \sum\nolimits_{k=2T+1-i}^{\infty} |\theta_{j,k+2T}| \leq \frac{B^2}{(4T+1-i)^2} \leq \frac{B^2}{T^2}.$$

Hence,

$$|R_i| \leq B/T \text{ and } |B_i| \leq B^2/T^2 \text{ for } i = 2T+1, ..., 3T. \tag{111}$$

Further, for $i = 1, ..., 2T$,

$$\begin{aligned} \mathbb{E}|x_i|^3 + \mathbb{E}|y_i|^3 &= \mathbb{E}|\varepsilon_{j,T+1-i}|^3 + 2\sqrt{2/\pi} < \left(\mathbb{E}|\varepsilon_{j,T+1-i}|^4\right)^{3/4} + 2 \\ &\leq (\varkappa_4 + 3)^{3/4} + 2 \leq \varkappa_4 + 5. \end{aligned} \tag{112}$$

Here the second to the last inequality follows from A1. For $2T < i \leq 3T$, we have

$$\begin{aligned} \mathbb{E}|x_i|^3 + \mathbb{E}|y_i|^3 &= \mathbb{E}|x_i|^3 = \mathbb{E}\left|\sum\nolimits_{k=2T+1-i}^{\infty} \theta_{j,k+2T}\varepsilon_{j,T+1-i-k}\right|^3 \\ &\leq \left(\mathbb{E}\left(\sum\nolimits_{k=2T+1-i}^{\infty} \theta_{j,k+2T}\varepsilon_{j,T+1-i-k}\right)^4\right)^{3/4} \\ &\leq \left(\left(\sum\nolimits_{k=2T+1-i}^{\infty} \theta_{j,k+2T}^2\right)^2 + \varkappa_4 \sum\nolimits_{k=2T+1-i}^{\infty} \theta_{j,k+2T}^4\right)^{3/4} \\ &\leq (1+\varkappa_4)^{3/4} B^3/T^3 \leq (1+\varkappa_4) B^3/T^3 \leq \varkappa_4 + 5 \end{aligned}$$

for all sufficiently large T.

Next, let $h(x) = g(\mathbf{X}(x))$. We have

$$\left|\partial_i^1 h(x)\right| \leq \sum\nolimits_{t=1}^{R} \left|\partial_t^1 g(\mathbf{X})\right| \left|\partial_i^1 \mathbf{X}_t(x)\right| \leq M_g \sum\nolimits_{t=1}^{R} \left|\partial_i^1 \mathbf{X}_t(x)\right|.$$

On the other hand,

$$\begin{aligned}\partial_i^1 \mathbf{X}_t(x) &= \partial_i^1 (e_j.v_t) = \sum\nolimits_{s=1}^{T} v_{ts} \partial_i^1 \left(\sum\nolimits_{k=0}^{T+s-1} \theta_{jk} x_{k+T-s+1} + x_{3T-s+1} \right) \\ &= \begin{cases} \sum\nolimits_{s=T+1-i}^{T} v_{ts} \theta_{j,i+s-T-1} & \text{for } i = 1, ..., 2T \\ v_{t,3T+1-i} & \text{for } i = 2T+1, ..., 3T \end{cases}.\end{aligned}$$

Since $|v_{ts}| \leq \sqrt{2/T}$ and $\sum_{k=0}^{\infty} (1+k) |\theta_{jk}| \leq B$, we have

$$\left|\partial_i^1 \mathbf{X}_t(x)\right| \leq (B+1) \sqrt{2/T}.$$

Here, we use $B+1$ instead of B to take into account a possibility that $B < 1$. Combining the latter display with the above inequality for $\left|\partial_i^1 h(x)\right|$, we obtain

$$\left|\partial_i^1 h(x)\right| \leq M_g C_1 / T^{1/2}, \tag{113}$$

where $C_1 = \sqrt{2} R (B+1)$.

Further,

$$\begin{aligned}\left|\partial_i^2 h(x)\right| &\leq \sum\nolimits_{t_1,t_2=1}^{R} \left|\partial_{t_1 t_2}^2 g(\mathbf{X})\right| \left|\partial_i^1 \mathbf{X}_{t_1}(x)\right| \left|\partial_i^1 \mathbf{X}_{t_2}(x)\right| \\ &\quad + \sum\nolimits_{t=1}^{R} \left|\partial_t^1 g(\mathbf{X})\right| \left|\partial_i^2 \mathbf{X}_t(x)\right| \\ &= \sum\nolimits_{t_1,t_2=1}^{R} \left|\partial_{t_1 t_2}^2 g(\mathbf{X})\right| \left|\partial_i^1 \mathbf{X}_{t_1}(x)\right| \left|\partial_i^1 \mathbf{X}_{t_2}(x)\right| \\ &\leq M_g \sum\nolimits_{t_1,t_2=1}^{R} \left|\partial_i^1 \mathbf{X}_{t_1}(x)\right| \left|\partial_i^1 \mathbf{X}_{t_2}(x)\right|\end{aligned}$$

so that

$$\left|\partial_i^2 h(x)\right| \leq M_g C_2 / T, \tag{114}$$

where $C_2 = C_1^2$. Similarly,

$$\left|\partial_i^3 h(x)\right| \leq \sum\nolimits_{t_1,t_2,t_3=1}^{R} \left|\partial_{t_1 t_2 t_3}^3 g(\mathbf{X})\right| \left|\partial_i^1 \mathbf{X}_{t_1}(x)\right| \left|\partial_i^1 \mathbf{X}_{t_2}(x)\right| \left|\partial_i^1 \mathbf{X}_{t_3}(x)\right|$$

so that

$$\left|\partial_i^3 h(x)\right| \leq M_g C_3 / T^{3/2}, \tag{115}$$

where $C_3 = C_1^3$.

Using inequalities established above in Theorem 12, we obtain

$$|\mathbb{E}h\left(x\right)-\mathbb{E}h\left(y\right)|\leq\frac{M_{g}C_{1}B}{T^{1/2}}+\frac{M_{g}C_{2}B^{2}}{2T^{2}}+\frac{M_{g}C_{3}\left(\varkappa_{4}+5\right)}{2T^{1/2}}.$$

On the other hand,

$$\begin{aligned}|\mathbb{E}g\left(\mathbf{X}\right)-\mathbb{E}g\left(\mathbf{Y}\right)| &= |\mathbb{E}g\left(\mathbf{X}(x)\right)-\mathbb{E}g\left(\mathbf{X}(y)\right)+\mathbb{E}g\left(\mathbf{X}(y)\right)-\mathbb{E}g\left(\mathbf{Y}\right)| \\ &= |\mathbb{E}h\left(x\right)-\mathbb{E}h\left(y\right)+\mathbb{E}g\left(\mathbf{X}(y)\right)-\mathbb{E}g\left(\mathbf{Y}\right)| \\ &\leq |\mathbb{E}h\left(x\right)-\mathbb{E}h\left(y\right)|+|\mathbb{E}g\left(\mathbf{X}(y)\right)-\mathbb{E}g\left(\mathbf{Y}\right)|.\end{aligned}$$

Note that $\mathbf{X}(y)$ and $\mathbf{Y}$ are normally distributed vectors with zero means but different covariance matrices, which we denote Σ_y and $\Sigma_{\mathbf{Y}}$, respectively. By definition,

$$\Sigma_{\mathbf{Y}}=2\pi f_{j}\left(0\right)I_{R}.$$

Define η_{T+1-i}, $i=1,2,...$ as y_i for $i=1,...,2T$, and as i.i.d. $N(0,1)$ random variables independent from $y_1,...,y_{2T}$ for $i>2T$. Then,

$$\begin{aligned}\mathbf{X}_t(y) &= \sum\nolimits_{s=1}^{T}\sum\nolimits_{k=0}^{T+s-1}\theta_{jk}y_{k+T-s+1}v_{ts} \\ &= \sum\nolimits_{s=1}^{T}\sum\nolimits_{k=0}^{T+s-1}\theta_{jk}\eta_{s-k}v_{ts} \\ &= \sum\nolimits_{s=1}^{T}\sum\nolimits_{k=0}^{\infty}\theta_{jk}\eta_{s-k}v_{ts}-\sum\nolimits_{s=1}^{T}\sum\nolimits_{k=T+s}^{\infty}\theta_{jk}\eta_{s-k}v_{ts} \\ &= \mathbf{X}_t(\eta)-\sum\nolimits_{s=1}^{T}\sum\nolimits_{k=T+s}^{\infty}\theta_{jk}\eta_{s-k}v_{ts} \\ &= \mathbf{X}_t(\eta)-\sum\nolimits_{\tau=T}^{\infty}\eta_{-\tau}\sum\nolimits_{s=1}^{T}\theta_{j,\tau+s}v_{ts}\end{aligned}$$

By Lemma 9,

$$|\mathbb{E}\left(\mathbf{X}_{t_1}(\eta)\mathbf{X}_{t_2}(\eta)\right)-2\pi f_j\left(\varpi_{t_1}/2\right)\delta_{t_1t_2}|\leq CB^2/T.$$

Further,

$$\begin{aligned}&\mathbb{E}\left(\prod\nolimits_{i=1}^{2}\sum\nolimits_{\tau=T}^{\infty}\eta_{-\tau}\sum\nolimits_{s=1}^{T}\theta_{j,\tau+s}v_{t_is}\right)=\sum\nolimits_{\tau=T}^{\infty}\left(\prod\nolimits_{i=1}^{2}\sum\nolimits_{s=1}^{T}\theta_{j,\tau+s}v_{t_is}\right) \\ \leq\ &\sum\nolimits_{\tau=T}^{\infty}\sum\nolimits_{s=1}^{T}\theta_{j,\tau+s}^{2}\leq\sum\nolimits_{s=1}^{T}\frac{B^2}{\left(T+s\right)^2}<\frac{B^2}{T},\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E}\left(\mathbf{X}_{t_1}(\eta)\sum\nolimits_{\tau=T}^{\infty}\eta_{-\tau}\sum\nolimits_{s=1}^{T}\theta_{j,\tau+s}v_{t_2 s}\right) \\
= \; & \mathbb{E}\left(\sum\nolimits_{s_1=1}^{T}\sum\nolimits_{\tau_1=-s_1}^{\infty}\eta_{-\tau_1}\theta_{j,\tau_1+s_1}v_{t_1 s_1}\sum\nolimits_{\tau_2=T}^{\infty}\eta_{-\tau_2}\sum\nolimits_{s_2=1}^{T}\theta_{j,\tau_2+s_2}v_{t_2 s_2}\right) \\
= \; & \sum\nolimits_{\tau=T}^{\infty}\sum\nolimits_{s_1=1}^{T}\theta_{j,\tau+s_1}v_{t_1 s_1}\sum\nolimits_{s_2=1}^{T}\theta_{j,\tau+s_2}v_{t_2 s_2} < \frac{B^2}{T}.
\end{aligned}$$

This yields

$$\left|\mathbb{E}\left(\mathbf{X}_{t_1}(y)\mathbf{X}_{t_2}(y)\right) - 2\pi f_j\left(\varpi_{t_1}/2\right)\delta_{t_1 t_2}\right| \leq \tilde{C}/T$$

for some constant $\tilde{C}$ that depends on B. This and inequality (95), yield

$$\left|\mathbb{E}\left(\mathbf{X}_{t_1}(y)\mathbf{X}_{t_2}(y)\right) - 2\pi f_j\left(0\right)\delta_{t_1 t_2}\right| \leq \hat{C}/T \tag{116}$$

for any positive integers $t_1, t_2 \leq R$, where $\hat{C}$ depends on B and R.

Further,

$$\left|g\left(\mathbf{X}(y)\right) - g\left(\mathbf{Y}\right)\right| \leq M_g\left\|\mathbf{X}(y) - \mathbf{Y}\right\|.$$

Therefore,

$$\begin{aligned}
\left|\mathbb{E}g\left(\mathbf{X}(y)\right) - \mathbb{E}g\left(\mathbf{Y}\right)\right| & \leq M_g\mathbb{E}\left\|\mathbf{X}(y) - \mathbf{Y}\right\| \\
& \leq M_g\left(\mathbb{E}\left\|\mathbf{X}(y) - \mathbf{Y}\right\|^2\right)^{1/2}
\end{aligned}$$

On the other hand, we may assume that y and $\mathbf{Y}$ are independent, and thus

$$\mathbb{E}\left\|\mathbf{X}(y) - \mathbf{Y}\right\|^2 = \operatorname{tr}\left(\Sigma_y - \Sigma_{\mathbf{Y}}\right) \leq R\hat{C}/T,$$

where the last inequality follows from (116). Hence, finally,

$$\begin{aligned}
\left|\mathbb{E}g\left(\mathbf{X}\right) - \mathbb{E}g\left(\mathbf{Y}\right)\right| \leq \; & \frac{M_g C_1 B}{T^{1/2}} + \frac{M_g C_2 B^2}{2T^2} \\
& + \frac{M_g C_3\left(\varkappa_4 + 5\right)}{2T^{1/2}} + \frac{M_g R^{1/2}\hat{C}^{1/2}}{T^{1/2}}.
\end{aligned}$$

This yields the statement of Lemma 10.

4 The "number of factors"

4.1 Proof of Proposition OW5

Let us denote $V(k) + k\hat{\sigma}^2 p_j(N,T)$ as $IPC_j(k)$. For positive integers k, $IPC_j(k) > IPC_j(k-1)$ if and only if $\hat{\lambda}_k/T < \hat{\sigma}^2 p_j(N,T)$. The latter inequality is equivalent to

$$\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma} < \left(1 - \sum\nolimits_{j=1}^{k_{\max}} \hat{\lambda}_j/\operatorname{tr}\hat{\Sigma}\right) p_j(N,T). \tag{117}$$

On the other hand, by Theorem OW1 (iii), for any fixed positive integer k, $\hat{\lambda}_k/\operatorname{tr}\hat{\Sigma} \xrightarrow{\mathrm{P}} 6/(k\pi)^2$. Since $p_j(N,T) \to \infty$ as $N,T \to \infty$ for $j = 1,2,3$, inequality (117) is satisfied with probability arbitrarily close to one for all sufficiently large N,T. This yields statement (i) of Proposition OW5.

To establish part (ii), we need the following lemma.

Lemma 13 *Under assumptions of Proposition OW5, for $k_{\max} = [\gamma\delta_{NT}]$, we have*

$$\hat{\sigma}^2 = O_{\mathrm{P}}\left(\frac{T\operatorname{tr}\Omega}{N\delta_{NT}} + \frac{(T+N_\varepsilon)\,N^{2\alpha}\delta_{NT}}{NT}\right).$$

Proof: We rely on notations and definitions from the proof of Theorem OW1 in Section 2.1. Let $\tilde{V}(k) = \operatorname{tr}\tilde{\Sigma}/T - \sum\nolimits_{j=1}^{k}\tilde{\lambda}_j/T$. Then,

$$T\tilde{V}(k_{\max}) = \sum\nolimits_{r=k_{\max}+1}^{T-1}\tilde{\lambda}_r \le \sum\nolimits_{r=k_{\max}+1}^{T-1} w_r'\tilde{\Sigma}w_r = \sum\nolimits_{r=k_{\max}+1}^{T-1}\frac{1}{N}\sigma_r^2 v_r'\varepsilon' W\varepsilon v_r.$$

Denote $\sum\nolimits_{r=k_{\max}+1}^{T-1}\sigma_r^2$ as $s_{k_{\max}}$. Then Corollary 3 and the fact that $\operatorname{tr}W = \operatorname{tr}\Omega$ yield

$$\sum_{r=k_{\max}+1}^{T-1}\sigma_r^2 v_r'\varepsilon' W\varepsilon v_r = s_{k_{\max}}\operatorname{tr}\Omega + o_{\mathrm{P}}(s_{k_{\max}}\operatorname{tr}\Omega).$$

Since $\sigma_r = (2\sin(\pi r/(2T)))^{-1} \le T/(2r)$ for $r = 1,...,T-1$, we have

$$s_{k_{\max}} \le \sum\nolimits_{r=k_{\max}+1}^{T-1} T^2/\left(4r^2\right) \le T^2/(4k_{\max}),$$

and therefore,

$$\tilde{V}(k_{\max}) \le \frac{T\operatorname{tr}\Omega}{4Nk_{\max}} + o_{\mathrm{P}}\left(\frac{T\operatorname{tr}\Omega}{Nk_{\max}}\right). \tag{118}$$

Next, similarly to (26), we have the following inequality

$$\left|(TV(k_{\max}))^{1/2} - \left(T\tilde{V}(k_{\max})\right)^{1/2}\right| \le \|\Psi^*(L)\,\varepsilon M\|\min\left\{1, \sqrt{T/N}\right\}.$$

Hence,

$$\hat{\sigma}^2 = V(k_{\max}) \leq 2\tilde{V}(k_{\max}) + \frac{2}{T}\left\|\Psi^*(L)\,\varepsilon M\right\|^2 \min\{1, T/N\}.$$

This inequality together with (25) and (118) yield the statement of the lemma. □

By Theorem OW1 (ii), for any fixed k, $\left(\hat{\lambda}_k/T\right)^{-1} = O_{\mathrm{P}}\left(N/(T \operatorname{tr}\Omega)\right)$. Therefore, by Lemma 13, $\left(\hat{\lambda}_k/T\right)^{-1}\hat{\sigma}^2 = O_{\mathrm{P}}\left(m_{NT}\right)$ and

$$\left(\hat{\lambda}_k/T\right)^{-1}\hat{\sigma}^2 p_j\left(N,T\right) = O_{\mathrm{P}}\left(m_{NT} p_j\left(N,T\right)\right) \xrightarrow{\mathrm{P}} 0.$$

Hence, for any fixed k, $\hat{\lambda}_k/T > \hat{\sigma}^2 p_j\left(N,T\right)$ with probability arbitrarily close to one for all sufficiently large N, T. This implies that $IPC_j(k) < IPC_j(k-1)$ with probability arbitrarily close to one for all sufficiently large N, T, and thus, $\hat{k}_j \xrightarrow{\mathrm{P}} \infty$.

5 Problem detection

5.1 Proof of Lemma OW6

Recall that $R_r = \operatorname{tr} P_{\Delta\hat{F}} P_{\hat{f}}$. Below, we obtain the required expansion of R_r, by first, expanding $P_{\Delta\hat{F}}$ and $P_{\hat{f}}$, and then combining these results. In the next two subsections we use the perturbation theory (e.g. Kato (1980)) to obtain expansions for $P_{\Delta\hat{F}}$ and $P_{\hat{f}}$.

5.1.1 Perturbation analysis in levels

Consider the following identity

$$X'X = \tilde{F}\Lambda'\Lambda\tilde{F}' + e'M_\Lambda e,$$

where $\tilde{F} = F + e'\Lambda\left(\Lambda'\Lambda\right)^{-1}$. Let $\lambda = \operatorname{diag}\{\lambda_1, ..., \lambda_r\}$ be the diagonal matrix of the r largest eigenvalues of matrix $\tilde{F}\Lambda'\Lambda\tilde{F}'/\left(NT^2\right)$ and let $\bar{F}$ be the $(T+1)\times r$ matrix of corresponding normalized eigenvectors. We have

$$X'X/\left(NT^2\right) = \bar{F}\lambda\bar{F}' + e'M_\Lambda e/\left(NT^2\right), \tag{119}$$

or equivalently,

$$\Phi = \Phi_0 + \Phi_1/\left(\delta_{NT}T\right), \tag{120}$$

where $\Phi = X'X/\left(NT^2\right)$, $\Phi_0 = \tilde{F}\Lambda'\Lambda\tilde{F}'/\left(NT^2\right) = \bar{F}\lambda\bar{F}'$, and $\Phi_1 = \delta_{NT}e'M_\Lambda e/\left(NT\right)$.

Note that $\sqrt{NT^2\left\|\Phi_0\right\|}$ equals $\left\|\Lambda F' + P_\Lambda e\right\|$. By assumptions B1, B2, and B3, the latter norm is $O_{\mathrm{P}}\left(\sqrt{NT^2}\right)$. Hence, $\left\|\Phi_0\right\| = O_{\mathrm{P}}(1)$. Further by B3, $\left\|\Phi_1\right\| = O_{\mathrm{P}}(1)$ and therefore, Φ can be interpreted as a small perturbation of Φ_0, and the perturbation theory can be used to link the eigenstructure of Φ to that of Φ_0. Specifically, (see Kato's (1980, p. 75-77) formulae (2.3) and (2.14))

$$P_{\hat{F}} = P_{\bar{F}} + \frac{1}{\delta_{NT}T}\sum\nolimits_{i=1}^{r} P_i^{(1)} + \epsilon, \tag{121}$$

where $\|\epsilon\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2} T^{-2}\right)$ and

$$P_i^{(1)} = -P_i \Phi_1 S_i - S_i \Phi_1 P_i$$

with P_i being the projection on the i-th column of $\bar{F}$, and S_i being the so-called reduced resolvent of Φ_0 evaluated at λ_i. Precisely,

$$S_i = \sum\nolimits_{j=1, j\neq i}^{r} \frac{1}{\lambda_j - \lambda_i} P_j - \frac{1}{\lambda_i} M_{\bar{F}}.$$

Remark 14 *We assume that the eigenvalues $\lambda_1, ..., \lambda_r$ are distinct. If they are not and only $r_1 < r$ of the eigenvalues are distinct, the sum in the representation (121) should run only up to $i = r_1$ with P_i denoting the so-called total eigenprojection on the space spanned by all the eigenvectors corresponding to the i-th distinct eigenvalue among $\lambda_1, ..., \lambda_r$. The analysis below should then be changed in a relatively straightforward manner, without affecting the final result.*

Multiplying both sides of (121) by $\bar{F}$ yields

$$\begin{aligned} \hat{F}\left(\hat{F}'\bar{F}\right) &= \bar{F} - \frac{1}{\delta_{NT} T} \sum\nolimits_{i=1}^{r} \sum\nolimits_{j=1, j\neq i}^{r} \frac{1}{\lambda_j - \lambda_i} \left[P_i \Phi_1 \bar{F}_j e_j' + P_j \Phi_1 \bar{F}_i e_i'\right] \\ &+ \frac{1}{\delta_{NT} T} \sum\nolimits_{i=1}^{r} \frac{1}{\lambda_i} M_{\bar{F}} \Phi_1 \bar{F}_i e_i' + \epsilon \bar{F}, \end{aligned} \tag{122}$$

where e_i is the i-th column of I_r matrix and $\bar{F}_j$ is the j-th column of $\bar{F}$. Further, multiplying both sides of (122) by Δ from the left yields

$$\Delta \hat{F}\left(\hat{F}'\bar{F}\right) = \Delta \bar{F} + \alpha,$$

where

$$\begin{aligned} \alpha &= -\frac{1}{\delta_{NT} T} \sum\nolimits_{i=1}^{r} \sum\nolimits_{j=1, j\neq i}^{r} \frac{1}{\lambda_j - \lambda_i} \Delta \left[P_i \Phi_1 \bar{F}_j e_j' + P_j \Phi_1 \bar{F}_i e_i'\right] \\ &+ \frac{1}{\delta_{NT} T} \sum\nolimits_{i=1}^{r} \frac{1}{\lambda_i} \Delta M_{\bar{F}} \Phi_1 \bar{F}_i e_i' + \Delta \epsilon \bar{F}. \end{aligned} \tag{123}$$

Therefore,

$$P_{\Delta \hat{F}} = \left(\Delta \bar{F} + \alpha\right)\left(\bar{F}'\Delta'\Delta\bar{F} + \bar{F}'\Delta'\alpha + \alpha'\Delta\bar{F} + \alpha'\alpha\right)^{-1}\left(\Delta\bar{F} + \alpha\right)'.$$

It will be convenient to write this equation in the following form

$$P_{\Delta \hat{F}} = P_{\Delta \bar{F}} + \delta_1 + \delta_2, \tag{124}$$

where

$$\begin{aligned} \delta_1 &= \alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' + \Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' \\ &- P_{\Delta\bar{F}}\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' - \Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' P_{\Delta\bar{F}}, \end{aligned} \tag{125}$$

and

$$\begin{aligned} \delta_2 &= \alpha \left(\bar{F}'\Delta'\Delta\bar{F} + \bar{F}'\Delta'\alpha + \alpha'\Delta\bar{F} + \alpha'\alpha\right)^{-1}\alpha' \\ &\quad + \alpha A\bar{F}'\Delta' + \Delta\bar{F}A\alpha' + \Delta\bar{F}B\bar{F}'\Delta' \end{aligned}$$

with

$$\begin{aligned} A &= \left(\bar{F}'\Delta'\Delta\bar{F} + \bar{F}'\Delta'\alpha + \alpha'\Delta\bar{F} + \alpha'\alpha\right)^{-1} - \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1} \text{ and} \\ B &= A + \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\left(\bar{F}'\Delta'\alpha + \alpha'\Delta\bar{F}\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}. \end{aligned}$$

5.1.2 Perturbation analysis in differences

Multiplying both sides of (120) by $T\Delta$ from the left and by Δ' from the right yields

$$\Omega = \Omega_0 + \Omega_1/\delta_{NT},$$

where $\Omega = \Delta X'X\Delta'/(TN)$, $\Omega_0 = \Delta\tilde{F}\Lambda'\Lambda\tilde{F}'\Delta'/(TN)$ and $\Omega_1 = \delta_{NT}\Delta e' M_\Lambda e\Delta'/(TN)$. By B1, B2 and B3, $\|\Omega_0\| = O_{\mathrm{P}}(1)$ and $\|\Omega_1\| = O_{\mathrm{P}}(1)$, so that Ω can be viewed as a small perturbation of Ω_0.

Let $\hat{f}$ and $\bar{f}$ be the $T \times r$ matrices of the normalized r principal eigenvectors of Ω and Ω_0, respectively. We will denote the corresponding eigenvalues as $\hat{\mu}_1, ..., \hat{\mu}_r$ and $\mu_1, ..., \mu_r$, respectively. Then, a higher order perturbation analysis than that used in (121) yields (see Kato's (1980, p. 75-77))

$$P_{\hat{f}} = P_{\bar{f}} + \delta_{NT}^{-1}\sum\nolimits_{i=1}^{r} R_i^{(1)} + \delta_{NT}^{-2}\sum\nolimits_{i=1}^{r} R_i^{(2)} + \epsilon^{(d)}, \tag{126}$$

where $\|\epsilon^{(d)}\| = O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right)$,

$$\begin{aligned} R_i^{(1)} &= -R_i\Omega_1 Q_i - Q_i\Omega_1 R_i, \text{ and} \\ R_i^{(2)} &= R_i\Omega_1 Q_i\Omega_1 Q_i + Q_i\Omega_1 R_i\Omega_1 Q_i + Q_i\Omega_1 Q_i\Omega_1 R_i \\ &\quad - R_i\Omega_1 R_i\Omega_1 Q_i^2 - R_i\Omega_1 Q_i^2\Omega_1 R_i - Q_i^2\Omega_1 R_i\Omega_1 R_i \end{aligned}$$

with R_i being the projection on the i-th column of $\bar{f}$ and

$$Q_i = \sum\nolimits_{j\neq i}^{r} \frac{1}{\mu_j - \mu_i} R_j - \frac{1}{\mu_i} M_{\bar{f}}.$$

It follows from equations (126) and (124) that

$$\mathrm{tr}\left[P_{\Delta\hat{F}} P_{\hat{f}}\right] = \sigma_1 + \sigma_2 + \sigma_3,$$

where

$$\begin{aligned}
\sigma_1 &= \operatorname{tr}\left[P_{\Delta\bar{F}}\left(P_{\bar{f}} + \delta_{NT}^{-1}\sum\nolimits_{i=1}^{r} R_i^{(1)} + \delta_{NT}^{-2}\sum\nolimits_{i=1}^{r} R_i^{(2)}\right)\right],\\
\sigma_2 &= \operatorname{tr}\left[\delta_1\left(P_{\bar{f}} + \delta_{NT}^{-1}\sum\nolimits_{i=1}^{r} R_i^{(1)}\right)\right],\\
\sigma_3 &= \operatorname{tr}\left[P_{\Delta\hat{F}}\epsilon^{(d)}\right] + \operatorname{tr}\left[\delta_1\delta_{NT}^{-2}\sum\nolimits_{i=1}^{r} R_i^{(2)}\right]\\
&\quad + \operatorname{tr}\left[\delta_2\left(P_{\bar{f}} + \delta_{NT}^{-1}\sum\nolimits_{i=1}^{r} R_i^{(1)} + \delta_{NT}^{-2}\sum\nolimits_{i=1}^{r} R_i^{(2)}\right)\right].
\end{aligned}$$

The remaining part of the proof consists of an asymptotic analysis of elements σ_1, σ_2, and σ_3.

5.1.3 Analysis of σ_1

Note that $\Delta\bar{F}$ and $\bar{f}$ span the same subspaces. Therefore, $P_{\Delta\bar{F}} = P_{\bar{f}}$ and $\operatorname{tr} P_{\Delta\bar{F}}P_{\bar{f}} = \operatorname{tr} P_{\Delta\bar{F}} = r$. Further, by definition of $R_i^{(1)}$

$$\begin{aligned}
\operatorname{tr}\left[P_{\Delta\bar{F}}R_i^{(1)}\right] &= -\operatorname{tr}\left[P_{\Delta\bar{F}}R_i\Omega_1 Q_i\right] - \operatorname{tr}\left[P_{\Delta\bar{F}}Q_i\Omega_1 R_i\right]\\
&= -\sum\nolimits_{j\neq i}^{r}\frac{1}{\mu_j-\mu_i}\left(\operatorname{tr}\left[R_i\Omega_1 R_j\right] + \operatorname{tr}\left[R_j\Omega_1 R_i\right]\right).
\end{aligned}$$

Therefore, $\sum_{i=1}^{r}\operatorname{tr}\left[P_{\Delta\bar{F}}R_i^{(1)}\right]$ equals

$$-\sum\nolimits_{i,j:i\neq j}^{r}\frac{1}{\mu_j-\mu_i}\left(\operatorname{tr}\left[R_i\Omega_1 R_j\right] + \operatorname{tr}\left[R_j\Omega_1 R_i\right]\right),$$

which is a sum of all elements of an anti-symmetric matrix. Hence,

$$\sum\nolimits_{i=1}^{r}\operatorname{tr}\left[P_{\Delta\bar{F}}R_i^{(1)}\right] = 0.$$

Next,

$$\begin{aligned}
P_{\Delta\bar{F}}R_i^{(2)} &= P_{\Delta\bar{F}}R_i\Omega_1 Q_i\Omega_1 Q_i + P_{\Delta\bar{F}}Q_i\Omega_1 R_i\Omega_1 Q_i\\
&\quad + P_{\Delta\bar{F}}Q_i\Omega_1 Q_i\Omega_1 R_i - P_{\Delta\bar{F}}R_i\Omega_1 R_i\Omega_1 Q_i^2\\
&\quad - P_{\Delta\bar{F}}R_i\Omega_1 Q_i^2\Omega_1 R_i - P_{\Delta\bar{F}}Q_i^2\Omega_1 R_i\Omega_1 R_i.
\end{aligned}$$

Since $P_{\Delta\bar{F}} = P_{\bar{f}}$, we have $P_{\Delta\bar{F}}M_{\bar{f}} = 0$. Therefore, as trace is invariant with respect to the interchange of the order in a product of two matrices, we have for the first term on the right hand side

$$
\begin{aligned}
& \operatorname{tr}\left[P_{\Delta \bar{F}} R_i \Omega_1 Q_i \Omega_1 Q_i\right] \\
= \;& \sum\nolimits_{j,p: j \neq i \text{ and } p \neq i}^{r} \frac{1}{\mu_j-\mu_i} \frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_i \Omega_1 R_p \Omega_1 R_j\right] \\
& -\sum\nolimits_{j: j \neq i}^{r} \frac{1}{\mu_j-\mu_i} \frac{1}{\mu_i} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_j\right];
\end{aligned} \tag{127}
$$

for the second term

$$
\begin{aligned}
& \operatorname{tr}\left[P_{\Delta \bar{F}} Q_i \Omega_1 R_i \Omega_1 Q_i\right] \\
= \;& \sum\nolimits_{j,p: j \neq i \text{ and } p \neq i}^{r} \frac{1}{\mu_j-\mu_i} \frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j \Omega_1 R_i \Omega_1 R_p\right];
\end{aligned} \tag{128}
$$

for the third term

$$
\begin{aligned}
& \operatorname{tr}\left[P_{\Delta \bar{F}} Q_i \Omega_1 Q_i \Omega_1 R_i\right] \\
= \;& \sum\nolimits_{j,p: j \neq i \text{ and } p \neq i}^{r} \frac{1}{\mu_j-\mu_i} \frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j \Omega_1 R_p \Omega_1 R_i\right] \\
& -\sum\nolimits_{j: j \neq i}^{r} \frac{1}{\mu_j-\mu_i} \frac{1}{\mu_i} \operatorname{tr}\left[R_j \Omega_1 M_{\bar{f}} \Omega_1 R_i\right].
\end{aligned} \tag{129}
$$

To similarly expand the rest of the terms, note that

$$
Q_i^2=\sum\nolimits_{j \neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} R_j+\frac{1}{\mu_i^2} M_{\bar{f}}.
$$

Therefore, for the fourth term

$$
-\operatorname{tr}\left[P_{\Delta \bar{F}} R_i \Omega_1 R_i \Omega_1 Q_i^2\right]=-\sum\nolimits_{j: j \neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} \operatorname{tr}\left[R_i \Omega_1 R_i \Omega_1 R_j\right]; \tag{130}
$$

for the fifth term

$$
\begin{aligned}
& -\operatorname{tr}\left[P_{\Delta \bar{F}} R_i \Omega_1 Q_i^2 \Omega_1 R_i\right] \\
= \;& -\sum\nolimits_{j: j \neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} \operatorname{tr}\left[R_i \Omega_1 R_j \Omega_1 R_i\right]-\frac{1}{\mu_i^2} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_i\right];
\end{aligned} \tag{131}
$$

and for the final, sixth, term

$$
-\operatorname{tr}\left[P_{\Delta \bar{F}} Q_i^2 \Omega_1 R_i \Omega_1 R_i\right]=-\sum\nolimits_{j: j \neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} \operatorname{tr}\left[R_j \Omega_1 R_i \Omega_1 R_i\right]. \tag{132}
$$

Combining (127) and (132) yields

$$
\begin{aligned}
& \sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} R_i \Omega_1 Q_i \Omega_1 Q_i - P_{\Delta\bar{F}} Q_i^2 \Omega_1 R_i \Omega_1 R_i\right] \\
= \; & \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_i \Omega_1 R_p \Omega_1 R_j\right] \\
& + \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} \operatorname{tr}\left[R_i \Omega_1 R_j \Omega_1 R_j - R_j \Omega_1 R_i \Omega_1 R_i\right] \\
& - \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_i} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_j\right] \\
= \; & \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_i \Omega_1 R_p \Omega_1 R_j\right] \\
& - \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_i} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_j\right].
\end{aligned}
$$

Combining (128) and (131) yields

$$
\begin{aligned}
& \sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} Q_i \Omega_1 R_i \Omega_1 Q_i - P_{\Delta\bar{F}} R_i \Omega_1 Q_i^2 \Omega_1 R_i\right] \\
= \; & \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j \Omega_1 R_i \Omega_1 R_p\right] \\
& + \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} \operatorname{tr}\left[R_j \Omega_1 R_i \Omega_1 R_j - R_i \Omega_1 R_j \Omega_1 R_i\right] \\
& - \sum\nolimits_{i=1}^{r} \frac{1}{\mu_i^2} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_i\right] \\
= \; & \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j \Omega_1 R_i \Omega_1 R_p\right] \\
& - \sum\nolimits_{i=1}^{r} \frac{1}{\mu_i^2} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_i\right].
\end{aligned}
$$

Combining (129) and (130) yields

$$
\begin{aligned}
& \sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} Q_i \Omega_1 Q_i \Omega_1 R_i - P_{\Delta\bar{F}} R_i \Omega_1 R_i \Omega_1 Q_i^2\right] \\
= \; & \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j \Omega_1 R_p \Omega_1 R_i\right] \\
& + \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\left(\mu_j-\mu_i\right)^2} \operatorname{tr}\left[R_j \Omega_1 R_j \Omega_1 R_i - R_i \Omega_1 R_i \Omega_1 R_j\right] \\
& - \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_i} \operatorname{tr}\left[R_j \Omega_1 M_{\bar{f}} \Omega_1 R_i\right] \\
= \; & \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j \Omega_1 R_p \Omega_1 R_i\right] \\
& - \sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_i} \operatorname{tr}\left[R_j \Omega_1 M_{\bar{f}} \Omega_1 R_i\right].
\end{aligned}
$$

Summing up the results of the last three displays, we obtain

$$
\begin{aligned}
&\sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} R_i^{(2)}\right] \\
= \;& \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_i\Omega_1 R_p \Omega_1 R_j\right] \\
& -\sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_i} \operatorname{tr}\left[R_i\Omega_1 M_{\bar{f}} \Omega_1 R_j\right] \\
& +\sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j\Omega_1 R_i \Omega_1 R_p\right] \\
& -\sum\nolimits_{i=1}^{r} \frac{1}{\mu_i^2} \operatorname{tr}\left[R_i\Omega_1 M_{\bar{f}} \Omega_1 R_i\right] \\
& +\sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j\Omega_1 R_p \Omega_1 R_i\right] \\
& -\sum\nolimits_{i,j:j\neq i}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_i} \operatorname{tr}\left[R_j\Omega_1 M_{\bar{f}} \Omega_1 R_i\right].
\end{aligned}
$$

The sum of the third and last lines of the latter display can be interpreted as a sum of all elements of an anti-symmetric matrix. Hence, it equals zero, and the expression for $\sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} R_i^{(2)}\right]$ simplifies as follows

$$
\begin{aligned}
&\sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} R_i^{(2)}\right] \\
= \;& \sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_i\Omega_1 R_p \Omega_1 R_j\right] \\
& +\sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j\Omega_1 R_i \Omega_1 R_p\right] \\
& +\sum\nolimits_{i,j,p:j\neq i\neq p\neq j}^{r} \frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i} \operatorname{tr}\left[R_j\Omega_1 R_p \Omega_1 R_i\right] \\
& -\sum\nolimits_{i=1}^{r} \frac{1}{\mu_i^2} \operatorname{tr}\left[R_i\Omega_1 M_{\bar{f}} \Omega_1 R_i\right].
\end{aligned}
$$

Renaming indices in the second and the third sums on the right hand side using rules $(i,j,p) \longmapsto (p,i,j)$ and $(i,j,p) \longmapsto (j,i,p)$, respectively, we represent the sum of the first three sums as a single sum over i,j,p s.t. $j\neq i\neq p\neq j$ of $\operatorname{tr}\left[R_i\Omega_1 R_p\Omega_1 R_j\right]$ multiplied by

$$
\begin{aligned}
&\frac{1}{\mu_j-\mu_i}\frac{1}{\mu_p-\mu_i}+\frac{1}{\mu_i-\mu_p}\frac{1}{\mu_j-\mu_p}+\frac{1}{\mu_i-\mu_j}\frac{1}{\mu_p-\mu_j} \\
= \;& \frac{\left(\mu_j-\mu_p\right)-\left(\mu_j-\mu_i\right)+\left(\mu_p-\mu_i\right)}{\left(\mu_j-\mu_i\right)\left(\mu_p-\mu_i\right)\left(\mu_j-\mu_p\right)}=0.
\end{aligned}
$$

Hence, finally,

$$
\sum\nolimits_{i=1}^{r} \operatorname{tr}\left[P_{\Delta\bar{F}} R_i^{(2)}\right] = -\sum\nolimits_{i=1}^{r} \frac{1}{\mu_i^2} \operatorname{tr}\left[R_i\Omega_1 M_{\bar{f}} \Omega_1 R_i\right],
$$

and overall,

$$\sigma_1 = r - \delta_{NT}^{-2} \sum\nolimits_{i=1}^{r} \frac{1}{\mu_i^2} \operatorname{tr}\left[R_i \Omega_1 M_{\bar{f}} \Omega_1 R_i\right].$$

Finally, note that $\sum\nolimits_{i=1}^{r} \mu_i^{-2} R_i = \left(\Omega_0^{+}\right)^2$, where Ω_0^{+} is the Moore-Penrose pseudoinverse of $\Omega_0 = \Delta\tilde{F}\Lambda'\Lambda\tilde{F}'\Delta'/\left(NT\right)$. Further, the space spanned by the columns of $\bar{f}$ is the same as that spanned by the columns of $\Delta\tilde{F}$. Therefore, we have

$$\sigma_1 = r - \delta_{NT}^{-2} \operatorname{tr}\left[\Omega_0^{+}\Omega_1 M_{\Delta\tilde{F}} \Omega_1 \Omega_0^{+}\right] = r - \delta_{NT}^{-2} \left\|\Omega_0^{+}\Omega_1 M_{\Delta\tilde{F}}\right\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

5.1.4 Analysis of σ_2

By definition (125) of δ_1,

$$\begin{aligned}
&\operatorname{tr}\left[\delta_1 P_{\bar{f}}\right] = \operatorname{tr}\left[\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' P_{\bar{f}}\right] + \operatorname{tr}\left[\Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' P_{\bar{f}}\right] \\
&- \operatorname{tr}\left[P_{\Delta\bar{F}}\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' P_{\bar{f}}\right] - \operatorname{tr}\left[\Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' P_{\Delta\bar{F}} P_{\bar{f}}\right] = 0,
\end{aligned}$$

where for the last equality we used the fact that $P_{\bar{f}} = P_{\Delta\bar{F}}$ and that trace is invariant with respect to the interchange of the order in a product of two matrices.

Further, using the same fact and the definition of Q_i, we obtain after some algebra

$$\begin{aligned}
& \operatorname{tr}\left[\delta_1 R_i^{(1)}\right] \\
= \; & -\sum\nolimits_{j:j\neq i}^{r} \frac{1}{\mu_j - \mu_i} \operatorname{tr}\left[\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta'\left(R_i\Omega_1 R_j + R_j\Omega_1 R_i\right)\right] \\
& -\sum\nolimits_{j:j\neq i}^{r} \frac{1}{\mu_j - \mu_i} \operatorname{tr}\left[\Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha'\left(R_j\Omega_1 R_i + R_i\Omega_1 R_j\right)\right] \\
& +\sum\nolimits_{j:j\neq i}^{r} \frac{1}{\mu_j - \mu_i} \operatorname{tr}\left[P_{\Delta\bar{F}}\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta'\left(R_i\Omega_1 R_j + R_j\Omega_1 R_i\right)\right] \\
& +\sum\nolimits_{j:j\neq i}^{r} \frac{1}{\mu_j - \mu_i} \operatorname{tr}\left[\Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' P_{\Delta\bar{F}}\left(R_i\Omega_1 R_j + R_j\Omega_1 R_i\right)\right] \\
& +\frac{1}{\mu_i} \operatorname{tr}\left[\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' R_i\Omega_1 M_{\bar{f}}\right] + \frac{1}{\mu_i} \operatorname{tr}\left[\Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' M_{\bar{f}}\Omega_1 R_i\right].
\end{aligned}$$

Summing the first four sums on the right hand side over i going from 1 to r, we obtain zero because the result can be interpreted as a sum of all elements of an anti-symmetric matrix. Therefore,

$$\sum\nolimits_{i=1}^{r} \operatorname{tr}\left[\delta_1 R_i^{(1)}\right] = \sum\nolimits_{i=1}^{r} \frac{2}{\mu_i} \operatorname{tr}\left[\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' R_i\Omega_1 M_{\bar{f}}\right].$$

By definition (123) of α, and since $\Delta \bar{F} M_{\bar{f}} = 0$, we have

$$\begin{aligned} M_{\bar{f}}\alpha &= \frac{1}{\delta_{NT}T}\sum\nolimits_{j=1}^{r}\frac{1}{\lambda_j}M_{\bar{f}}\Delta M_{\bar{F}}\Phi_1\bar{F}_j e_j' + M_{\bar{f}}\Delta\epsilon\bar{F} \\ &= \frac{1}{\delta_{NT}T}\sum\nolimits_{j=1}^{r}\frac{1}{\lambda_j}M_{\bar{f}}\Delta\Phi_1\bar{F}_j e_j' + M_{\bar{f}}\Delta\epsilon\bar{F}. \end{aligned}$$

This yields

$$\begin{aligned} &\frac{1}{\delta_{NT}}\sum\nolimits_{i=1}^{r}\operatorname{tr}\left[\delta_1 R_i^{(1)}\right] \\ = {}& \frac{2}{\delta_{NT}^2 T}\sum\nolimits_{i=1}^{r}\sum\nolimits_{j=1}^{r}\frac{1}{\lambda_j}\frac{1}{\mu_i}\operatorname{tr}\left[\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' R_i\Omega_1 M_{\bar{f}}\Delta\Phi_1\bar{F}_j e_j'\right] \\ &+\frac{1}{\delta_{NT}}\sum\nolimits_{i=1}^{r}\frac{2}{\mu_i}\operatorname{tr}\left[\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' R_i\Omega_1 M_{\bar{f}}\Delta\epsilon\bar{F}\right]. \end{aligned}$$

Now recall that $\lambda_1, ..., \lambda_r$ are the principal eigenvalues of $\tilde{F}\Lambda'\Lambda\tilde{F}'/\left(NT^2\right)$. Therefore, their square roots are the principal singular values of $\left(\Lambda F' + P_\Lambda e\right)/\left(\sqrt{N}T\right)$. In particular, $\lambda_r^{1/2}$ must be no smaller than the r-th largest singular value of $\Lambda F'/\left(\sqrt{N}T\right)$ minus $\left\|P_\Lambda e/\left(\sqrt{N}T\right)\right\| = O_{\mathrm{P}}\left(1/\sqrt{\delta_{NT}T}\right)$. On the other hand, the r-th largest singular value of $\Lambda F'/\left(\sqrt{N}T\right)$ equals the square root from the smallest eigenvalue of $\left(F'F/T^2\right)\left(\Lambda'\Lambda/N\right)$. By B1-B2, such smallest eigenvalue must be bounded away from zero in probability. Therefore, $\lambda_r^{-1} = O_{\mathrm{P}}(1)$ and hence,

$$\lambda_j^{-1} = O_{\mathrm{P}}(1) \text{ for } j \leq r. \tag{133}$$

Further, $\mu_1, ..., \mu_r$ are the principal eigenvalues of $\Omega_0 = \Delta\left(\Lambda F' + P_\Lambda e\right)'\left(\Lambda F' + P_\Lambda e\right)\Delta'/\left(TN\right)$. By arguments similar to those just used, μ_r^{-1} is of the same order as the inverse of the smallest eigenvalue of $F'\Delta'\Delta F\Lambda'\Lambda/\left(TN\right)$. Therefore, by B1 and B2, $\mu_r^{-1} = O_{\mathrm{P}}(1)$ and hence,

$$\mu_j^{-1} = O_{\mathrm{P}}(1) \text{ for } j \leq r. \tag{134}$$

Next, the identity

$$\frac{\left(\Lambda F' + P_\Lambda e\right)'\left(\Lambda F' + P_\Lambda e\right)}{NT^2} = \bar{F}\lambda\bar{F}'$$

yields the following representation

$$\left(F + e'\Lambda\left(\Lambda'\Lambda\right)^{-1}\right)\left(\Lambda'\Lambda\right)^{1/2}/\left(T\sqrt{N}\right) = \bar{F}\lambda^{1/2}V,$$

where V is a r-dimensional orthonormal matrix. Therefore,

$$\bar{F} = \frac{1}{T}\left(F + e'\Lambda\left(\Lambda'\Lambda\right)^{-1}\right)\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}V'\lambda^{-1/2}, \tag{135}$$

and

$$\Delta\bar{F}=\frac{1}{T}\left(\Delta F+\Delta e'\Lambda\left(\Lambda'\Lambda\right)^{-1}\right)\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}V'\lambda^{-1/2}.$$

It follows that

$$\left\|\Delta\bar{F}\right\|=O_{\mathrm{P}}\left(\left\|\Delta F+\Delta e'\Lambda\left(\Lambda'\Lambda\right)^{-1}\right\|/T\right),$$

and hence, by B1-B3,

$$\left\|\Delta\bar{F}\right\|=O_{\mathrm{P}}\left(T^{-1/2}\right)\text{ and }\left\|\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\right\|=O_{\mathrm{P}}\left(T\right).\tag{136}$$

Using (134) and (136), and recalling that $\|\epsilon\|=O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-2}\right)$, we obtain

$$\frac{1}{\delta_{NT}}\sum\nolimits_{i=1}^{r}\frac{2}{\mu_i}\operatorname{tr}\left[\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta'R_i\Omega_1M_{\bar{f}}\Delta\epsilon\bar{F}\right]=O_{\mathrm{P}}\left(\delta_{NT}^{-3}T^{-1/2}\right)$$

Therefore,

$$\begin{aligned}&\frac{1}{\delta_{NT}}\sum\nolimits_{i=1}^{r}\operatorname{tr}\left[\delta_1R_i^{(1)}\right]\\=\;&\frac{2}{\delta_{NT}^2T}\sum\nolimits_{i=1}^{r}\sum\nolimits_{j=1}^{r}\frac{1}{\lambda_j}\frac{1}{\mu_i}\operatorname{tr}\left[\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta'R_i\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_je_j'\right]\\&+O_{\mathrm{P}}\left(\delta_{NT}^{-3}T^{-1/2}\right).\end{aligned}$$

Let us now analyze the trace on the right hand side of the above equation. The absolute value of that trace equals

$$\begin{aligned}&\left|e_j'\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta'R_i\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_j\right|\\=\;&\left|e_j'\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta'\bar{f}_i\right|\left|\bar{f}_i'\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_j\right|\\=\;&O_{\mathrm{P}}\left(T^{1/2}\right)\left|\bar{f}_i'\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_j\right|.\end{aligned}\tag{137}$$

Since $M_{\bar{f}}=I_r-\bar{f}\bar{f}'$, we have

$$\begin{aligned}\left|\bar{f}_i'\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_j\right|&\le\left|\bar{f}_i'\Omega_1\Delta\Phi_1\bar{F}_j\right|+\left\|\bar{f}_i'\Omega_1\bar{f}\right\|\left\|\bar{f}'\Delta\Phi_1\bar{F}_j\right\|\\&\le\left\|\bar{f}'\Omega_1\Delta\Phi_1\bar{F}\right\|+\left\|\Omega_1\right\|\left\|\bar{f}'\Delta\Phi_1\bar{F}\right\|.\end{aligned}$$

We point out that $\bar{f}$ can be thought of as the matrix of left singular vectors of

$$T^{-1/2}\Delta\left(F+e'\Lambda\left(\Lambda'\Lambda\right)^{-1}\right)\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2},$$

and $\mu^{1/2}$ can be thought of as the diagional matrix of the corresponding singular values. Hence, there exists an orthogonal matrix W such that

$$\bar{f}\mu^{1/2}W = T^{-1/2}\Delta\tilde{F}\left(\Lambda'\Lambda/N\right)^{1/2},$$

where $\tilde{F} = F + e'\Lambda\left(\Lambda'\Lambda\right)^{-1}$, and therefore,

$$\bar{f} = T^{-1/2}\Delta\tilde{F}\left(\Lambda'\Lambda/N\right)^{1/2}W'\mu^{-1/2}.$$

Using this together with (135), we obtain

$$\begin{aligned}\left\|\bar{f}'\Omega_1\Delta\Phi_1\bar{F}\right\| &\leq T^{-3/2}\left\|\tilde{F}'\Delta'\Omega_1\Delta\Phi_1\tilde{F}\right\|\left\|\frac{\Lambda'\Lambda}{N}\right\|\left\|\mu^{-1/2}\right\|\left\|\lambda^{-1/2}\right\| \\ &= \left\|\tilde{F}'\Delta'\Omega_1\Delta\Phi_1\tilde{F}\right\|O_{\mathrm{P}}\left(T^{-3/2}\right),\end{aligned}$$

where we used B2, (133), and (134).

Further,

$$\begin{aligned}\left\|\tilde{F}'\Delta'\Omega_1\Delta\Phi_1\tilde{F}\right\| &\leq \left\|F'\Delta'\Omega_1\Delta\Phi_1F\right\| + \left\|\tilde{F}'\Delta'\Omega_1\Delta\Phi_1e'\Lambda\left(\Lambda'\Lambda\right)^{-1}\right\| \\ &\quad+\left\|\left(\Lambda'\Lambda\right)^{-1}\Lambda'e\Delta'\Omega_1\Delta\Phi_1F\right\|.\end{aligned}$$

Since $\left\|\left(\Lambda'\Lambda\right)^{-1}\Lambda'e\right\|^2 = O_{\mathrm{P}}\left(T/\delta_{NT}\right)$, $\|F\|^2 = O_{\mathrm{P}}\left(T^2\right) = \left\|\tilde{F}\right\|^2$ and $\left\|\Delta'\Omega_1\Delta\Phi_1\right\| = O_{\mathrm{P}}(1)$, we have

$$\left\|\tilde{F}'\Delta'\Omega_1\Delta\Phi_1\tilde{F}\right\| \leq \left\|F'\Delta'\Omega_1\Delta\Phi_1F\right\| + O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}T^{3/2}\right)$$

and

$$\left\|\bar{f}'\Omega_1\Delta\Phi_1\bar{F}\right\| \leq \left\|F'\Delta'\Omega_1\Delta\Phi_1F\right\|O_{\mathrm{P}}\left(T^{-3/2}\right) + O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right).$$

Similarly,

$$\begin{aligned}\left\|\bar{f}'\Delta\Phi_1\bar{F}\right\| &\leq \left\|\tilde{F}'\Delta'\Delta\Phi_1\tilde{F}\right\|O_{\mathrm{P}}\left(T^{-3/2}\right) \\ &\leq \left\|F'\Delta'\Delta\Phi_1F\right\|O_{\mathrm{P}}\left(T^{-3/2}\right) + O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\left|\bar{f}_i'\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_j\right| &\leq \left\|F'\Delta'\Omega_1\Delta\Phi_1F\right\|O_{\mathrm{P}}\left(T^{-3/2}\right) \\ &\quad+\left\|F'\Delta'\Delta\Phi_1F\right\|O_{\mathrm{P}}\left(T^{-3/2}\right) + O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right).\end{aligned}$$

The lemma below implies that $\left\|F'\Delta'\Omega_1\Delta\Phi_1F\right\| = O_{\mathrm{P}}(T)$ and $\left\|F'\Delta'\Delta\Phi_1F\right\| = O_{\mathrm{P}}(T)$. Hence,

$$\left|\bar{f}_i'\Omega_1M_{\bar{f}}\Delta\Phi_1\bar{F}_j\right| = O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right), \tag{138}$$

and, by (137),

$$\left| e_j' \left(\bar{F}'\Delta'\Delta\bar{F} \right)^{-1} \bar{F}'\Delta' R_i \Omega_1 M_{\bar{f}} \Delta \Phi_1 \bar{F}_j \right| = O_{\mathrm{P}} \left((T/\delta_{NT})^{1/2} \right).$$

Therefore, overall,

$$\frac{1}{\delta_{NT}} \sum\nolimits_{i=1}^{r} \operatorname{tr} \left[\delta_1 R_i^{(1)} \right] = O_{\mathrm{P}} \left(\delta_{NT}^{-5/2} T^{-1/2} \right)$$

and

$$\sigma_2 = O_{\mathrm{P}} \left(\delta_{NT}^{-5/2} T^{-1/2} \right).$$

as well.

Lemma 15 *Let $F_{t+1} = F_t + f_{t+1}$ be an r-dimensional process with $\mathbb{E} \|F_0\|^4 < \infty$ and $f_t = \sum\nolimits_{r=0}^{\infty} \Pi_r \eta_{t-r}$, where $\eta_s = (\eta_{1s}, ..., \eta_{rs})'$ are such that η_{is} are i.i.d., $\mathbb{E}\eta_{is} = 0$, $\mathbb{E}\eta_{is}^2 = 1$, $\mathbb{E}\eta_{is}^4 < \infty$, and $\sum\nolimits_{r=0}^{\infty} (1+r) \|\Pi_r\| < \infty$. Denote $(F_0, F_1, ..., F_T)'$ as F and $(f_1, ..., f_T)$ as f. Let Δ be a $T \times (T+1)$ matrix with elements $\Delta_{ii} = -1, \Delta_{i,i+1} = 1$, and all other elements zero. Finally, let A be a random $(T+1) \times (T+1)$ matrix, independent from F and such that $\|A\| = O_{\mathrm{P}}(1)$ as $T \to \infty$. Then*

$$\left\| f'\Delta A F \right\| = O_{\mathrm{P}}(T).$$

Proof: Let U be a $(T+1) \times T$ matrix with elements $U_{ij} = 1$ if $i > j$ and all the other elements zero. Denote a $(T+1)$-vector of ones as l. We have $F = lF_0' + Uf$, and

$$f'\Delta AF = f'\Delta AlF_0' + f'\Delta AUf.$$

For the first term, we have

$$\left\| f'\Delta AlF_0' \right\| \le \|f\| \|\Delta\| \|l\| \|A\| \|F_0\|.$$

Since $\|l\| = \sqrt{T+1}$, $\|f\| = O_{\mathrm{P}}\left(\sqrt{T}\right)$, $\|A\| = O_{\mathrm{P}}(1)$, $\|\Delta\| = O(1)$, and $\|F_0\| = O_{\mathrm{P}}(1)$, the above inequality yields

$$\left\| f'\Delta AlF_0 \right\| = O_{\mathrm{P}}(T).$$

Hence, it remains to show that $\|f'\Delta AUf\| = O_{\mathrm{P}}(T)$.

Let $g_t = \sum\nolimits_{r=0}^{T} \Pi_r \eta_{t-r}$, $R_t = f_t - g_t$, and $g = (g_1, ..., g_T)'$, $R = (R_1, ..., R_T)'$. We have

$$\begin{aligned} \left\| f'\Delta AUf \right\| &\le \left\| g'\Delta AUg \right\| + \left\| g'\Delta AUR \right\| + \left\| R'\Delta AUg \right\| + \left\| R'\Delta AUR \right\| \\ &\le \left\| g'\Delta AUg \right\| + \|\Delta A\| \left(\|g\| \|UR\| + \|R\| \|Ug\| + \|R\| \|UR\| \right). \end{aligned} \tag{139}$$

Note that

$$\mathbb{E} \|R\|^2 \le \mathbb{E} \|R\|_F^2 = \operatorname{tr} \mathbb{E} \left(RR' \right) = T \sum\nolimits_{k=T+1}^{\infty} \|\Pi_k\|_F^2 \le Tr \sum\nolimits_{k=T+1}^{\infty} \|\Pi_k\|^2.$$

On the other hand, $\|\Pi_k\| < c/(1+k)$, where $c = \sum_{k=0}^{\infty}(1+k)\|\Pi_k\|$, and $\sum_{k=T+1}^{\infty} c^2/(1+k)^2 < c^2/(T+1)$. Therefore, $\mathbb{E}\|R\|^2 < c^2 r$ and, as a consequence,

$$\|R\| = O_{\mathrm{P}}(1). \tag{140}$$

Further,

$$\begin{aligned}
\mathbb{E}\|UR\|^2 &\leq \mathbb{E}\|UR\|_F^2 \\
&= \sum_{j=1}^{T}\mathbb{E}(R_1+..+R_j)'(R_1+..+R_j) \\
&= \sum_{j=1}^{T}\left[j\mathbb{E}R_t'R_t + 2(j-1)\mathbb{E}R_t'R_{t-1}+...+2\mathbb{E}R_t'R_{t+1-j}\right] \\
&\leq T^2\sum_{i=0}^{\infty}\left|\mathbb{E}R_t'R_{t-i}'\right|.
\end{aligned}$$

But

$$\begin{aligned}
\left|\mathbb{E}R_t'R_{t-i}\right| &= \left|\sum_{k=T+1}^{\infty}\mathrm{tr}\left(\Pi_k'\Pi_{k+i}\right)\right| \leq \left|\sum_{k=T+1}^{\infty}\|\Pi_k\|_F\|\Pi_{k+i}\|_F\right| \\
&\leq \frac{cr}{2+T}\sum_{k=T+1}^{\infty}\|\Pi_{k+i}\|_F.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\|UR\|^2 &\leq crT\sum_{i=0}^{\infty}\sum_{k=T+1}^{\infty}\|\Pi_{k+i}\|_F \leq cr^2T\left(\|\Pi_{T+1}\|+2\|\Pi_{T+2}\|+...\right) \\
&\leq cr^2T\sum_{k=0}^{\infty}(1+k)\|\Pi_k\| \leq c^2r^2T.
\end{aligned}$$

This yields

$$\|UR\| = O_{\mathrm{P}}\left(\sqrt{T}\right). \tag{141}$$

Using (140) and (141) in (139), and noting that $\|\Delta A\| = O_{\mathrm{P}}(1), \|g\| = O_{\mathrm{P}}\left(\sqrt{T}\right)$, and $\|Ug\| = O_{\mathrm{P}}(T)$, we obtain

$$\left\|f'\Delta AUf\right\| \leq \left\|g'\Delta AUg\right\| + O_{\mathrm{P}}(T).$$

Hence, it remains to show that $\|g'\Delta AUg\| = O_{\mathrm{P}}(T)$.

Let $g_{\cdot j}$ be the j-th column of g. Represent $g_{\cdot j}$ in the form $\bar{\Pi}_j\eta$, where $\bar{\Pi}_j$ is a $T\times 2Tr$ matrix

$$\bar{\Pi}_j = \begin{pmatrix}
\Pi_{T,j\cdot} & ... & ... & \Pi_{1,j\cdot} & \Pi_{0,j\cdot} & & & \\
& \Pi_{T,j\cdot} & ... & ... & \Pi_{1,j\cdot} & \Pi_{0,j\cdot} & & \\
& & \ddots & & & \ddots & \ddots & \\
& & & \Pi_{T,j\cdot} & ... & ... & \Pi_{1,j\cdot} & \Pi_{0,j\cdot}
\end{pmatrix},$$

where $\Pi_{s,j\cdot}$ is the j-th row of Π_s, and $\eta' = \left(\eta'_{1-T}, ..., \eta'_T\right)$. With this notation, we have

$$g'_{\cdot a}\Delta AU g_{\cdot b} = \eta'\bar{\Pi}'_a \Delta AU\bar{\Pi}_b\eta.$$

Let $\mathbb{E}_A$ denote the expectation conditional on A. We have

$$\begin{aligned}\mathbb{E}_A\left(\eta'\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\eta\right)^2 &= \sum\nolimits_{i,j}\left[\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)^2_{ij} + \left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)_{ij}\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)_{ji}\right] \\ &\quad + \sum\nolimits_i\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)^2_{ii}\left(\mathbb{E}\eta^4_t - 3\right) \\ &\quad + \sum\nolimits_{i,j}\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)_{ii}\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)_{jj}.\end{aligned}$$

Since

$$\begin{aligned}\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)_{ij}\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)_{ji} &\leq \left(\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)^2_{ij} + \left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)^2_{ji}\right)/2, \text{and} \\ \sum\nolimits_{i,j}\left(\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)^2_{ij} &= \left\|\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right\|^2_F,\end{aligned}$$

we have

$$\mathbb{E}_A\left(\eta'\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\eta\right)^2 \leq \left(\mathbb{E}\eta^4_t + 2\right)\left\|\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right\|^2_F + \left(\operatorname{tr}\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right)^2. \tag{142}$$

For the first term on the right hand side, we have

$$\left\|\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right\|^2_F \leq \left\|\bar{\Pi}_a\right\|^2\left\|\bar{\Pi}_b\right\|^2\|\Delta\|^2\|A\|^2\|U\|^2_F.$$

On the other hand,

$$\left\|\bar{\Pi}_a\right\| \leq \sum\nolimits_{s=0}^T\sum\nolimits_{t=1}^r\left|\Pi_{s,at}\right| \leq r\sum\nolimits_{s=0}^T\|\Pi_s\| \leq rc.$$

Similarly, $\left\|\bar{\Pi}_b\right\| \leq rc$. Furthermore, $\|U\|^2_F = (T+1)T/2$. This yields

$$\left(\mathbb{E}\eta^4_t + 2\right)\left\|\bar{\Pi}'_a\Delta AU\bar{\Pi}_b\right\|^2_F < T^2\left(\mathbb{E}\eta^4_t + 2\right)r^4c^4\|\Delta\|^2\|A\|^2. \tag{143}$$

Let us consider $\operatorname{tr}\bar{\Pi}'_a\Delta AU\bar{\Pi}_b$. We have

$$\operatorname{tr}\bar{\Pi}'_a\Delta AU\bar{\Pi}_b = \operatorname{tr}AU\bar{\Pi}_b\bar{\Pi}'_a\Delta = \operatorname{tr}\left(AU\Gamma\Delta\right),$$

where $\Gamma_{ij} = \Gamma_{i-j} = \sum\nolimits_{t=0}^{T-|i-j|}\Pi_{t,b\cdot}\Pi'_{t+|i-j|,a\cdot}$.

Let us decompose matrix $U\Gamma\Delta$ as

$$U\Gamma\Delta = \sum\nolimits_{k=1-T}^{T-1}\Gamma_k UD_k\Delta, \tag{144}$$

where D_k is a matrix with elements $\left(D_k\right)_{ij} = 1$ if $i-j=k$ and zero elements otherwise. A direct calculation reveals that

$$UD_k\Delta = D_k - l_k e'_{1\vee(1-k)} + l_{T+1+k}e'_{T+1},$$

where l_k is a $(T+1)$-vector with $(l_k)_i = 0$ for $i \leq k$ and $(l_k)_i = 1$ for $i > k$, e_j is the j-th column of I_{T+1}, and $1 \vee (1-k)$ denotes $\max\{1, 1-k\}$. Of course, for $k < 0$, l_k is just a $(T+1)$-vector with all elements equal one, and for $k \geq 0$, l_{T+1+k} is just a zero vector.

Decomposition (144) yields

$$\operatorname{tr}(AU\Gamma\Delta) = \sum_{k=1-T}^{T-1} \Gamma_k \left[\operatorname{tr}(AD_k) - e'_{1\vee(1-k)} A l_k + e'_{T+1} A l_{T+1+k}\right]$$

so that

$$|\operatorname{tr}(AU\Gamma\Delta)| \leq \sum_{k=1-T}^{T-1} |\Gamma_k| \left[|\operatorname{tr}(AD_k)| + \left|e'_{1\vee(1-k)} A l_k\right| + \left|e'_{T+1} A l_{T+1+k}\right|\right].$$

On the other hand,

$$\begin{aligned} |\operatorname{tr}(AD_k)| &\leq (T+1)\|AD_k\| \leq (T+1)\|A\|, \\ \left|e'_{1\vee(1-k)} A l_k\right| &\leq \sqrt{T+1}\,\|A\|, \text{ and} \\ \left|e'_T A l_{T+k}\right| &\leq \sqrt{T+1}\,\|A\|. \end{aligned}$$

Therefore,

$$|\operatorname{tr}(AU\Gamma\Delta)| \leq 3(T+1)\|A\| \sum_{k=1-T}^{T-1} |\Gamma_k|.$$

Recall that $c = \sum_{k=0}^{\infty} (1+k)\|\Pi_k\|$, and note that

$$\begin{aligned} \sum_{k=1-T}^{T-1} |\Gamma_k| &\leq 2\sum_{k=0}^{T-1} |\Gamma_k| \leq 2\sum_{k=0}^{T-1} \sum_{t=0}^{T-k} \left(\|\Pi_t\|_F \|\Pi_{t+k}\|_F\right) \\ &\leq 2\sum_{k=0}^{T-1} \sum_{t=0}^{T} \left(\|\Pi_t\|_F \|\Pi_{t+k}\|_F\right) \\ &\leq 2r^2 \sum_{k=0}^{T-1} \sum_{t=0}^{T} \left(\|\Pi_t\| \|\Pi_{t+k}\|\right) \leq 2r^2 c^2. \end{aligned}$$

Therefore,

$$\sum_{k=1-T}^{T-1} |\Gamma_k| \leq 2r^2 c^2,$$

and thus,

$$\left|\operatorname{tr} \bar{\Pi}'_a \Delta A U \bar{\Pi}_b\right| = |\operatorname{tr}(AU\Gamma\Delta)| \leq 6r^2c^2(T+1)\|A\|.$$

Using this and (143) in (142), yields

$$\mathbb{E}_A \left(\eta' \bar{\Pi}'_a \Delta A U \bar{\Pi}_b \eta\right)^2 \leq (T+1)^2 B_T \|A\|^2,$$

where

$$B_T = \left(\mathbb{E}\eta_t^4 + 2\right) r^4 c^4 \|\Delta\|^2 + 36 c^4 r^4 < \bar{B}$$

for some finite $\bar{B}$ for all T. Therefore, we have

$$\left(\eta'\bar{\Pi}_a'\Delta AU\bar{\Pi}_b\eta\right)^2 = O_{\mathrm{P}}\left(T^2\right). \tag{145}$$

But

$$\left\|g'\Delta AUg\right\| \leq \left\|g'\Delta AUg\right\|_F = \left(\sum\nolimits_{a,b=1}^{r}\left(\eta'\bar{\Pi}_a'\Delta AU\bar{\Pi}_b\eta\right)^2\right)^{1/2}.$$

Therefore,

$$\left\|g'\Delta AUg\right\| = O_{\mathrm{P}}\left(T\right).$$

This concludes the proof of the lemma. □

Remark 16 *It is easy to show that under the assumptions of the lemma,* $\|f'AF\| = O_{\mathrm{P}}\left(T\ln T\right)$. *For this, the above proof goes through with* Δ *omitted up to the point where we need to find a bound on* $|\mathrm{tr}\left(AU\Gamma\right)|$. *Note that* $\mathrm{tr}\left(AU\Gamma\right) = \mathrm{tr}\left(\Gamma AU\right) = l'\left(\Gamma A\right)^{(U)} l$ *with* l *equal the vector of ones and* $\left(\Gamma A\right)^{(U)}$ *being the upper triangular part of* ΓA. *Hence,*

$$|\mathrm{tr}\left(AU\Gamma\right)| \leq (T+1)\left\|\left(\Gamma A\right)^{(U)}\right\|.$$

Mathias (1993) shows that $\left\|M^{(U)}\right\| \leq O\left(\ln\left(T+1\right)\right)\|M\|$ *for any* $(T+1)\times(T+1)$ *matrix* M. *Hence,*

$$|\mathrm{tr}\left(AU\Gamma\right)| \leq O(T\ln T)\left\|\Gamma A\right\|,$$

and the rest of the proof proceeds as above.

5.1.5 Analysis of σ_3

By definition,

$$\sigma_3 = \mathrm{tr}\left[P_{\Delta\hat{F}}\epsilon^{(d)}\right] + \mathrm{tr}\left[\delta_1\delta_{NT}^{-2}\sum\nolimits_{i=1}^{r} R_i^{(2)}\right] + \mathrm{tr}\left[\delta_2\left(P_{\hat{f}} - \epsilon^{(d)}\right)\right].$$

Since $\left\|\epsilon^{(d)}\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right)$, we have

$$\mathrm{tr}\left[P_{\Delta\hat{F}}\epsilon^{(d)}\right] = O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right).$$

Next, recall that

$$\begin{aligned}\delta_1 \quad = \quad & \alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' + \Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' \\ & -P_{\Delta\bar{F}}\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\bar{F}'\Delta' - \Delta\bar{F}\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha' P_{\Delta\bar{F}}.\end{aligned}$$

But $\left\|\Delta\bar{F}\right\| = O_{\mathrm{P}}\left(T^{-1/2}\right)$, $\left\|\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\right\| = O_{\mathrm{P}}\left(T\right)$, and $\|\alpha\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1}\right)$. Therefore,

$$\|\delta_1\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right).$$

Since $\left\|\delta_{NT}^{-2}\sum_{i=1}^{r}R_i^{(2)}\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2}\right)$, we have

$$\operatorname{tr}\left[\delta_1\delta_{NT}^{-2}\sum_{i=1}^{r}R_i^{(2)}\right] = O_{\mathrm{P}}\left(\delta_{NT}^{-3}T^{-1/2}\right).$$

Further, recall that

$$\begin{aligned}\delta_2 &= \alpha\left(\left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)^{-1}\alpha' \\ &\quad +\alpha A\bar{F}'\Delta' + \Delta\bar{F}A\alpha' + \Delta\bar{F}B\bar{F}'\Delta'\end{aligned}$$

with

$$\begin{aligned}A &= \left(\left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)^{-1} - \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1} \text{ and} \\ B &= A + \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\left(\bar{F}'\Delta'\alpha + \alpha'\Delta\bar{F}\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}.\end{aligned}$$

We have

$$\left\|\left(\left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)^{-1}\right\| = O_{\mathrm{P}}\left(T\right)$$

and hence,

$$\left\|\alpha\left(\left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)^{-1}\alpha'\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-1}\right).$$

Further,

$$\begin{aligned}A &= \left(\left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)^{-1}\left(\bar{F}'\Delta'\Delta\bar{F} - \left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1} \\ &= -\left(\left(\Delta\bar{F}+\alpha\right)'\left(\Delta\bar{F}+\alpha\right)\right)^{-1}\left(\alpha'\Delta\bar{F} + \bar{F}'\Delta'\alpha + \alpha'\alpha\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\end{aligned}$$

so that $\|A\| = O_{\mathrm{P}}\left(T^{1/2}\delta_{NT}^{-1}\right)$ and

$$\left\|\alpha A\bar{F}'\Delta' + \Delta\bar{F}A\alpha'\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-1}\right).$$

Finally, we can represent A as

$$A = \left(A + \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\right)\left(-\alpha'\Delta\bar{F} - \bar{F}'\Delta'\alpha - \alpha'\alpha\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}$$

so that, by definition,

$$\begin{aligned} B &= \left(A + \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\right)\left(-\alpha'\Delta\bar{F} - \bar{F}'\Delta'\alpha - \alpha'\alpha\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1} \\ &\quad + \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\left(\bar{F}'\Delta'\alpha + \alpha'\Delta\bar{F}\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1} \\ &= A\left(-\alpha'\Delta\bar{F} - \bar{F}'\Delta'\alpha - \alpha'\alpha\right)\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1} \\ &\quad - \left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\alpha'\alpha\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}. \end{aligned}$$

Since $\|A\| = O_{\mathrm{P}}\left(T^{1/2}\delta_{NT}^{-1}\right)$, $\left\|\left(\bar{F}'\Delta'\Delta\bar{F}\right)^{-1}\right\| = O_{\mathrm{P}}(T)$, $\left\|\Delta\bar{F}\right\| = O_{\mathrm{P}}\left(T^{-1/2}\right)$, and $\|\alpha\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1}\right)$, we have

$$\|B\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2}\right).$$

Therefore,

$$\left\|\Delta\bar{F}B\bar{F}'\Delta'\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-1}\right),$$

and overall,

$$\delta_2 = O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-1}\right).$$

Since $\left\|P_{\hat{f}} - \epsilon^{(d)}\right\| = O_{\mathrm{P}}(1)$, this implies that

$$\mathrm{tr}\left[\delta_2\left(P_{\hat{f}} - \epsilon^{(d)}\right)\right] = O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-1}\right).$$

Thus,

$$\sigma_3 = O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right).$$

Combining the above results for σ_1, σ_2, and σ_3, we conclude that

$$R_r = \mathrm{tr}\left[P_{\Delta\hat{F}}P_{\hat{f}}\right] = r - \delta_{NT}^{-2}\left\|\Omega_0^{+}\Omega_1 M_{\Delta\tilde{F}}\right\|_F^2 + O_{\mathrm{P}}\left(\delta_{NT}^{-3}\right).$$

5.2 Proof of Lemma OW7

We will use perturbation analysis to study the asymptotic behavior of $\widehat{\mathcal{S}}_r$. Let

$$\Phi(\varkappa) = \Phi_0 + \varkappa\Phi_1,$$

where $\Phi_0 = \bar{F}\lambda\bar{F}'$ and $\Phi_1 = \delta_{NT}e'M_\Lambda e/(NT)$, as in (120), and $\varkappa$ is a complex-valued variable. Let $R(z,\varkappa) = (\Phi(\varkappa) - zI_T)^{-1}$ be the resolvent of $\Phi(\varkappa)$. Finally, let $R(z) = R(z,0)$ be the resolvent of Φ_0.

Kato (1980, p.67) shows that

$$R(z,\varkappa) = R(z)\sum\nolimits_{p=0}^{\infty}\left(-\left(\Phi(\varkappa) - \Phi_0\right)R(z)\right)^p, \tag{146}$$

where the series on the right hand side converges for

$$\|(\Phi(\varkappa)-\Phi_0)R(z)\|<1,$$

which is satisfied for any z, where $R(z)$ is defined, for sufficiently small $\varkappa$. The series also can be written as

$$R(z,\varkappa)=R(z)+\sum\nolimits_{n=1}^{\infty}\varkappa^{n}R^{(n)}(z), \tag{147}$$

where

$$R^{(n)}(z)=\sum\nolimits_{n_1+\ldots+n_p=n,n_j\geq 1}(-1)^{p}R(z)\Phi_{n_1}R(z)\Phi_{n_2}R(z)\ldots\Phi_{n_p}R(z). \tag{148}$$

Note that $\hat{F}$ is the matrix of the r principal eigenvectors of $\Phi\left(\delta_{NT}^{-1}T^{-1}\right)$ and $\bar{F}$ is the matrix of the r principal eigenvectors of Φ_0. As explained in Kato (1980, p.68), the projections $P_{\hat{F}}$ and $P_{\bar{F}}$ on the spaces spanned by the columns of $\hat{F}$ and $\bar{F}$, respectively, equal

$$P_{\hat{F}} = \hat{F}\hat{F}'=-\frac{1}{2\pi i}\oint_{\Gamma}R\left(z,\delta_{NT}^{-1}T^{-1}\right)\mathrm{d}z \text{ and} \tag{149}$$

$$P_{\bar{F}} = \bar{F}\bar{F}'=-\frac{1}{2\pi i}\oint_{\Gamma}R(z)\,\mathrm{d}z, \tag{150}$$

where Γ is a contour in the complex plane that encircles counterclockwise the r of the largest eigenvalues of $\Phi\left(\delta_{NT}^{-1}T^{-1}\right)$ and Φ_0 (but not the rest of the eigenvalues of these matrices). Similarly,

$$\hat{F}\hat{\lambda}\hat{F}' = -\frac{1}{2\pi i}\oint_{\Gamma}zR\left(z,\delta_{NT}^{-1}T^{-1}\right)\mathrm{d}z \text{ and} \tag{151}$$

$$\bar{F}\lambda\bar{F}' = -\frac{1}{2\pi i}\oint_{\Gamma}zR(z)\,\mathrm{d}z. \tag{152}$$

Equations (151-152) and (147) lead to the representation (compare to Kato's (1980) formulae (2.3) and (2.14) on pp. 75-77)

$$\hat{F}\hat{\lambda}\hat{F}'=\bar{F}\lambda\bar{F}'+\frac{1}{\delta_{NT}T}\sum\nolimits_{i=1}^{r}\tilde{P}_i^{(1)}+\tilde{\epsilon}, \tag{153}$$

where

$$\tilde{P}_i^{(1)}=-\lambda_i P_i\Phi_1 S_i-\lambda_i S_i\Phi_1 P_i+P_i\Phi_1 P_i$$

with P_i being the projection on the i-th column of $\bar{F}$ and $S_i=\sum_{j=1,j\neq i}^{r}\frac{1}{\lambda_j-\lambda_i}P_j-\frac{1}{\lambda_i}M_{\bar{F}}$ being the reduced resolvent of Φ_0 evaluated at λ_i. Furthermore,

$$\tilde{\epsilon}=-\frac{1}{2\pi i}\sum\nolimits_{p=2}^{\infty}\left(-\frac{1}{\delta_{NT}T}\right)^{p}\oint_{\Gamma}zR(z)\left(\Phi_1 R(z)\right)^{p}\mathrm{d}z,$$

which yields $\|\tilde{\epsilon}\|=O_{\mathrm{P}}\left(\delta_{NT}^{-2}T^{-2}\right).$

We have

$$
\begin{aligned}
& \sum\nolimits_{i=1}^{r} \lambda_i \left(P_i \Phi_1 S_i + S_i \Phi_1 P_i\right) \\
= & \sum\nolimits_{i=1}^{r} \lambda_i P_i \Phi_1 \left(\sum\nolimits_{j\neq i} \frac{P_j}{\lambda_j - \lambda_i} - \frac{1}{\lambda_i} M_{\bar{F}}\right) \\
& + \sum\nolimits_{i=1}^{r} \lambda_i \left(\sum\nolimits_{j\neq i} \frac{P_j}{\lambda_j - \lambda_i} - \frac{1}{\lambda_i} M_{\bar{F}}\right) \Phi_1 P_i \\
= & -\sum\nolimits_{i=1}^{r} \sum\nolimits_{j\neq i} P_i \Phi_1 P_j - P_{\bar{F}} \Phi_1 M_{\bar{F}} - M_{\bar{F}} \Phi_1 P_{\bar{F}} \\
= & -P_{\bar{F}} \Phi_1 P_{\bar{F}} - P_{\bar{F}} \Phi_1 M_{\bar{F}} - M_{\bar{F}} \Phi_1 P_{\bar{F}} + \sum\nolimits_{i=1}^{r} P_i \Phi_1 P_i \\
= & -P_{\bar{F}} \Phi_1 - M_{\bar{F}} \Phi_1 P_{\bar{F}} + \sum\nolimits_{i=1}^{r} P_i \Phi_1 P_i
\end{aligned}
$$

Therefore, $\sum\nolimits_{i=1}^{r} \tilde{P}_i^{(1)} = P_{\bar{F}} \Phi_1 + M_{\bar{F}} \Phi_1 P_{\bar{F}}$, and from (153),

$$
\begin{aligned}
\hat{F}\hat{\lambda}\hat{F}' &= \bar{F}\lambda\bar{F}' + \frac{1}{\delta_{NT} T}\left(P_{\bar{F}} \Phi_1 + M_{\bar{F}} \Phi_1 P_{\bar{F}}\right) + \tilde{\epsilon} \\
&= \bar{F}\lambda\bar{F}' + \frac{1}{\delta_{NT} T}\left(\Phi_1 - M_{\bar{F}} \Phi_1 M_{\bar{F}}\right) + \tilde{\epsilon}.
\end{aligned}
$$

This yields

$$
\Delta\hat{F}\hat{\lambda}\hat{F}'\Delta' - \Delta\bar{F}\lambda\bar{F}'\Delta' = \frac{1}{\delta_{NT} T}\Delta\left(\Phi_1 - M_{\bar{F}} \Phi_1 M_{\bar{F}}\right)\Delta' + \hat{\epsilon}, \tag{154}
$$

where $\|\hat{\epsilon}\| = O_{\mathrm{P}}\left(\delta_{NT}^{-2} T^{-2}\right)$.

Note that

$$
\left\|\Delta\left(\Phi_1 - M_{\bar{F}} \Phi_1 M_{\bar{F}}\right)\Delta'\right\| \leq \left\|\Delta P_{\bar{F}} \Phi_1 \Delta'\right\| + \left\|\Delta \Phi_1 P_{\bar{F}} \Delta'\right\| + \left\|\Delta P_{\bar{F}} \Phi_1 P_{\bar{F}} \Delta'\right\|.
$$

But (see (136))

$$
\left\|\Delta P_{\bar{F}}\right\| = \left\|\Delta \bar{F}\bar{F}'\right\| \leq \left\|\Delta \bar{F}\right\| \left\|\bar{F}\right\| = O_{\mathrm{P}}\left(T^{-1/2}\right).
$$

Therefore,

$$
\left\|\Delta\left(\Phi_1 - M_{\bar{F}} \Phi_1 M_{\bar{F}}\right)\Delta'\right\| = O_{\mathrm{P}}\left(T^{-1/2}\right),
$$

and

$$
\left\|\Delta\hat{F}\hat{\lambda}\hat{F}'\Delta' - \Delta\bar{F}\lambda\bar{F}'\Delta'\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1} T^{-3/2}\right). \tag{155}
$$

Let μ_i and $\bar{f}_i$ be the i-th largest eigenvalue and the corresponding normalized eigenvector of $T\Delta\bar{F}\lambda\bar{F}'\Delta'$, and let $\tilde{\mu}_i$ and $\tilde{f}_i$ be the i-th largest eigenvalue and the corresponding eigenvector of $T\Delta\hat{F}\hat{\lambda}\hat{F}'\Delta'$. [2] Equation (155) implies that

$$
\max_{i\leq r}\left\|P_{\bar{f}_i} - P_{\tilde{f}_i}\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1} T^{-1/2}\right),
$$

[2] We assume that $\mu_1, ..., \mu_r$ are distinct. If they are not, the analysis below needs some relatively straightforward modification, but the results do not change (see Remark 14).

so that

$$\left\|M_{\bar{f}}-M_{\Delta\hat{F}}\right\|=\left\|\sum\nolimits_{i=1}^{r}\left(P_{\tilde{f}_i}-P_{\bar{f}_i}\right)\right\|=O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right). \tag{156}$$

Furthermore,

$$\max_{i\leq r}\left|\mu_i-\tilde{\mu}_i\right|=O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right),$$

and thus,

$$\left\|\sum\nolimits_{i=1}^{r}\left(\frac{1}{\mu_i}P_{\bar{f}_i}-\frac{1}{\tilde{\mu}_i}P_{\tilde{f}_i}\right)\right\|=O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right).$$

But

$$\sum\nolimits_{i=1}^{r}\left(\frac{1}{\mu_i}P_{\bar{f}_i}-\frac{1}{\tilde{\mu}_i}P_{\tilde{f}_i}\right)=\Omega_0^{+}-\hat{\Omega}_0^{+}.$$

Therefore,

$$\left\|\Omega_0^{+}-\hat{\Omega}_0^{+}\right\|=O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right). \tag{157}$$

Equations (156) and (157) yield

$$\left\|\Omega_0^{+}\Omega_1M_{\bar{f}}-\hat{\Omega}_0^{+}\hat{\Omega}_1M_{\Delta\hat{F}}\right\|=\left\|\Omega_0^{+}\Omega_1M_{\bar{f}}-\Omega_0^{+}\hat{\Omega}_1M_{\bar{f}}\right\|+O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right)$$

so that

$$\left\|\Omega_0^{+}\Omega_1M_{\bar{f}}\right\|_F^2-\left\|\hat{\Omega}_0^{+}\hat{\Omega}_1M_{\Delta\hat{F}}\right\|_F^2=\left\|\Omega_0^{+}\Omega_1M_{\bar{f}}\right\|_F^2-\left\|\Omega_0^{+}\hat{\Omega}_1M_{\bar{f}}\right\|_F^2+O_{\mathrm{P}}\left(\delta_{NT}^{-1}T^{-1/2}\right). \tag{158}$$

Next, the definitions of Ω_1 and $\hat{\Omega}_1$ yield

$$\Delta\hat{F}\hat{\lambda}\Omega_1-\hat{\Omega}_1=\delta_{NT}T\left(\hat{F}'\Delta'-\Delta\bar{F}\lambda\bar{F}'\Delta'\right). \tag{159}$$

This identity and equation (155) imply that

$$\left\|\Omega_1-\hat{\Omega}_1\right\|=O_{\mathrm{P}}\left(T^{-1/2}\right).$$

Therefore,

$$\begin{aligned}\left\|\Omega_0^{+}\Omega_1M_{\bar{f}}\right\|_F^2-\left\|\Omega_0^{+}\hat{\Omega}_1M_{\bar{f}}\right\|_F^2 &= \operatorname{tr}\Omega_0^{+}\Omega_1M_{\bar{f}}\Omega_1\Omega_0^{+}-\operatorname{tr}\Omega_0^{+}\hat{\Omega}_1M_{\bar{f}}\hat{\Omega}_1\Omega_0^{+}\\ &= 2\operatorname{tr}\Omega_0^{+}\left(\Omega_1-\hat{\Omega}_1\right)M_{\bar{f}}\Omega_1\Omega_0^{+}\\ &\quad -\operatorname{tr}\Omega_0^{+}\left(\Omega_1-\hat{\Omega}_1\right)M_{\bar{f}}\left(\Omega_1-\hat{\Omega}_1\right)\Omega_0^{+}\\ &= 2\operatorname{tr}\Omega_0^{+}\left(\Omega_1-\hat{\Omega}_1\right)M_{\bar{f}}\Omega_1\Omega_0^{+}+O_{\mathrm{P}}\left(T^{-1}\right).\end{aligned}$$

Further, by definition of Ω_0^{+},

$$\operatorname{tr}\Omega_0^{+}\left(\Omega_1-\hat{\Omega}_1\right)M_{\bar{f}}\Omega_1\Omega_0^{+}=\sum\nolimits_{i=1}^{r}\frac{1}{\mu_i^2}\bar{f}_i'\left(\Omega_1-\hat{\Omega}_1\right)M_{\bar{f}}\Omega_1\bar{f}_i,$$

and, from (154) and (159),

$$\Omega_1 - \hat{\Omega}_1 = \Delta P_{\bar{F}} \Phi_1 \Delta' + \Delta \Phi_1 P_{\bar{F}} \Delta' - \Delta P_{\bar{F}} \Phi_1 P_{\bar{F}} \Delta' + \delta_{NT} T \hat{\epsilon}.$$

Since $P_{\bar{F}} \Delta' M_{\bar{f}} = 0$ and since $\|\delta_{NT} T \hat{\epsilon}\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1} T^{-1}\right)$, we have

$$\begin{aligned} \operatorname{tr} \Omega_0^+ \left(\Omega_1 - \hat{\Omega}_1\right) M_{\bar{f}} \Omega_1 \Omega_0^+ &= \sum\nolimits_{i=1}^r \frac{1}{\mu_i^2} \bar{f}_i' \Delta P_{\bar{F}} \Phi_1 \Delta' M_{\bar{f}} \Omega_1 \bar{f}_i + O_{\mathrm{P}}\left(\delta_{NT}^{-1} T^{-1}\right) \\ &= \sum\nolimits_{i=1}^r \frac{1}{\mu_i^2} \bar{f}_i' \Delta \bar{F} \bar{F}' \Phi_1 \Delta' M_{\bar{f}} \Omega_1 \bar{f}_i + O_{\mathrm{P}}\left(\delta_{NT}^{-1} T^{-1}\right). \end{aligned} \tag{160}$$

Finally, recall equation (138):

$$\left|\bar{f}_i' \Omega_1 M_{\bar{f}} \Delta \Phi_1 \bar{F}_j\right| = O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right).$$

This implies that

$$\left\|\bar{F}' \Phi_1 \Delta' M_{\bar{f}} \Omega_1 \bar{f}_i\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right).$$

Further, similarly to (138), we have

$$\left\|\bar{f}_i' \Delta \bar{F} / \mu_i^2\right\| = O_{\mathrm{P}}\left(\delta_{NT}^{-1/2}\right).$$

Using the latter two displays in (160), we obtain

$$\operatorname{tr} \Omega_0^+ \left(\Omega_1 - \hat{\Omega}_1\right) M_{\bar{f}} \Omega_1 \Omega_0^+ = O_{\mathrm{P}}\left(\delta_{NT}^{-1}\right),$$

and thus,

$$\left\|\Omega_0^+ \Omega_1 M_{\bar{f}}\right\|_F^2 - \left\|\hat{\Omega}_0^+ \hat{\Omega}_1 M_{\Delta \hat{F}}\right\|_F^2 = O_{\mathrm{P}}\left(\delta_{NT}^{-1}\right). \square$$

References

[1] Anselone, P. M. (1967) "Collectively Compact Operator Approximations," *Technical Report* 76, Computer Science Department, Stanford University.

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

[3] Chatterjee, S. (2006) "A Generalization of the Lindeberg Principle," *The Annals of Probability* 34, 2061–2076.

[4] Hannan, E. J. (1970) *Multiple Time Series.* John Wiley & Sons, Inc.

[5] Karatzas, I. and Shreve, S. E. (1998) *Brownian Motion and Stochastic Calculus.* Second Edition. Springer.

[6] Kato, T. (1980) *Perturbation Theory for Linear Operators*, Springer-Verlag.

[7] Latala, R. (2004) "Some Estimates of Norms of Random Matrices," *Proceedings of the American Mathematical Society* 133, 1273–1282.

[8] Mathias, R. (1993). "The Hadamard operator norm of a circulant and applications," *SIAM Journal of Matrix Analysis and Applications* 14, 1152–1167.

[9] Onatski, A. (2015) "Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models," *Journal of Econometrics* 186, 388–406.

[10] Onatski, A. (2018) "Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise," Cambridge Working Paper Economics 1808, University of Cambridge.

[11] Onatski, A. and Wang, C. (2019) "Spurious Factor Analysis," manuscript, University of Cambridge.

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

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