Technical Appendix to "Asymptotics of the principal components estimator of large factor models with weak factors"

Alexei Onatski*
Economics Department, Columbia University

May 2, 2009

Abstract

This Appendix contains proofs of all the propositions of the paper "Asymptotics of the principal components estimator of large factor models with weak factors".

1 Proof of Theorem 1

In the proofs below, $\lambda_i(M)$ denotes the i-th largest by absolute value eigenvalue of matrix M and norm $\|M\|$ equals the square root of the largest eigenvalue of $M'M$. We will prove parts ii) and iii) of the theorem. A proof of part i) is similar to the proof of part ii) and we omit it to save space. In what follows, we will assume that $\operatorname{Var}\varepsilon_{it} = 1$. Such an assumption is without loss of generality. Indeed, multiplication of the data by a non-zero scalar α leaves the statement of Theorem 1 unchanged. It only leads to scaling of d_i, σ^2 and $\hat{L}$ by α^2, α^2 and α, respectively. Taking $\alpha = \sigma^{-1}$, we normalize $\operatorname{Var}\varepsilon_{it}$ to one. We will also assume, without loss of generality, that constants a_i, $i = 1, ..., n$ and $b_i, i = 1, ..., T$ in Assumption 2 are positive. If some of them are negative, we simply change the sign of the corresponding $\mathcal{L}_i$ or $\mathcal{F}_i$. Furthermore, if some of them are exactly zero, we can make them positive but decreasing to zero as $n \to \infty$ so fast that the first and second-order asymptotics of the principal components estimator does not change.

1.1 Truncation and renormalization

Let $\bar{\varepsilon}_{it} = (\operatorname{Var}\tilde{\varepsilon}_{it})^{-1/2}(\tilde{\varepsilon}_{it} - E\tilde{\varepsilon}_{it})$ with $\tilde{\varepsilon}_{it} = \varepsilon_{it} I_{|\varepsilon_{it}| \leq \ln n}$ be a truncated, centralized and renormalized version of ε_{it} and let $\bar{X} = LF' + A\bar{\varepsilon}B$. For the eigenvalues of XX'/T and $\bar{X}\bar{X}'/T$,

---

*420 West 118th St., International Affairs Building, Economics Department, Room 1011, New York, NY, 10027. Tel: (212) 854-3685. Fax: (212) 854 8059. E-mail: ao2027@columbia.edu

$\mu_i \equiv \lambda_i \left(XX'/T \right)$ and $\bar{\mu}_i \equiv \lambda_i \left(\bar{X}\bar{X}'/T \right)$, respectively, we have: $\max_{i \leq n} \left| \mu_i - \bar{\mu}_i \right| = o_p(1)$. To prove this fact, we will need the following result, which was established in Theorem 3.1 of Yin, Bai and Krishnaiah (1988):

Lemma 1. (Yin, Bai and Krishnaiah, 1988) *Let η be an $n \times T$ matrix with i.i.d. entries η_{it} with $E\eta_{it} = 0$ and $E\eta_{it}^4 < \infty$. Then, $T^{-1/2} \left\| \eta \right\| \overset{a.s.}{\to} (1 + \sqrt{c}) \left(E\eta_{it}^2 \right)^{1/2}$ as n and T go to infinity so that $n/T \to c$.*

By Corollary 7.3.8 of Horn and Johnson (1985), $\max_{i \leq n} \left| \sqrt{\mu_i} - \sqrt{\bar{\mu}_i} \right| \leq \left\| A \left(\varepsilon - \bar{\varepsilon} \right) B / \sqrt{T} \right\| \leq \left\| A \right\| \left\| B \right\| \left\| \varepsilon - \bar{\varepsilon} \right\| / \sqrt{T}$. By Assumption 3ii), $\left\| A \right\| \overset{a.s.}{=} O\left(1\right)$ and $\left\| B \right\| \overset{a.s.}{=} O\left(1\right)$. Further, note that matrix $\varepsilon - \bar{\varepsilon}$ has i.i.d. entries with finite fourth moment, zero mean and variance $2 - 2E\tilde{\varepsilon}_{it}^2 / \sqrt{\operatorname{Var} \tilde{\varepsilon}_{it}}$, which is no larger than $2E \left(\varepsilon_{it}^2 I_{\left| \varepsilon_{it} \right| > \ln n} \right)$, and hence, converges to zero as $n \to \infty$. Therefore, by Lemma 1, $\left\| \varepsilon - \bar{\varepsilon} \right\| / \sqrt{T} \overset{a.s.}{=} o(1)$ and we have: $\max_{i \leq n} \left| \sqrt{\mu_i} - \sqrt{\bar{\mu}_i} \right| \overset{a.s.}{=} o(1)$. On the other hand, $\left| \mu_i - \bar{\mu}_i \right| = \left| \sqrt{\mu_i} - \sqrt{\bar{\mu}_i} \right| \left| \sqrt{\mu_i} + \sqrt{\bar{\mu}_i} \right| \leq \left| \sqrt{\mu_i} - \sqrt{\bar{\mu}_i} \right| \left| \sqrt{\mu_1} + \sqrt{\bar{\mu}_1} \right|$. We have: $\sqrt{\mu_1} = \left\| T^{-1/2} X \right\| \leq \left\| T^{-1/2} LF' \right\| + \left\| A \right\| \left\| T^{-1/2} \varepsilon \right\| \left\| B \right\|$. By Assumptions 1ii) and 1iii), $\left\| T^{-1/2} LF' \right\| = O_p\left(1\right)$, and by Lemma 1, $\left\| T^{-1/2} \varepsilon \right\| \overset{a.s.}{=} O\left(1\right)$. Hence, $\sqrt{\mu_1} = O_p\left(1\right)$. Similarly, $\sqrt{\bar{\mu}_1} = O_p\left(1\right)$, and therefore, $\max_{i \leq n} \left| \mu_i - \bar{\mu}_i \right| = o_p(1)$.

Such a uniform eigenvalue approximation result implies that, for the purpose of proving part iii) of Theorem 1, we can assume without loss of generality that

$$\max_{i \leq n, t \leq T} \left| \varepsilon_{it} \right| \leq \ln n. \tag{1}$$

Assumption (1) is also without loss of generality for the purpose of proving the convergence of the first q columns of P stated in part ii) of Theorem 1. It is because the projections on the principal q eigenspaces of XX'/T and those on the principal q eigenspaces of $\bar{X}\bar{X}'/T$ converge (in probability) to each other in operator norm.

Indeed, let $\mathcal{T}$ and $\mathcal{T}^{(1)}$ be linear operators acting in $\mathbb{R}^n$, which are represented with respect to the standard basis by matrices XX'/T and $\left(\bar{X}\bar{X}' - XX' \right)/T$, respectively, and let $\mathcal{T}\left(\varkappa\right) = \mathcal{T} + \varkappa \mathcal{T}^{(1)}$. Denote the resolvent of $\mathcal{T}\left(\varkappa\right)$, $\left(\mathcal{T}\left(\varkappa\right) - \zeta \right)^{-1}$, as $R\left(\zeta, \varkappa\right)$ and the resolvent of $\mathcal{T}$, $\left(\mathcal{T} - \zeta \right)^{-1}$, as $R\left(\zeta\right)$. Let Γ be a positively oriented circle in the complex plane with the center at μ_i and radius $r = \frac{1}{2} \min\left(h_i, \rho \right)$, where $h_1 = \left| \mu_1 - \mu_2 \right|$ and $h_i = \min\left\{ \left| \mu_{i-1} - \mu_i \right|, \left| \mu_i - \mu_{i+1} \right| \right\}$ for $i > 1$. Define $P_i\left(\varkappa\right) = -\frac{1}{2\pi\sqrt{-1}} \int_\Gamma R\left(\zeta, \varkappa\right) d\zeta$. Then (see Kato, 1980, p.67-68 and p.88), for all $\left| \varkappa \right| < r^{-1} \left\| \mathcal{T}^{(1)} \right\|^{-1}$, $P_i\left(\varkappa\right)$ is the eigenprojection of $\mathcal{T}\left(\varkappa\right)$ corresponding to its unique eigenvalue inside circle Γ, and $P_i\left(\varkappa\right)$ can be represented in the form of the convergent (in operator norm) series:

$$P_i\left(\varkappa\right) = P_i(0) + \frac{1}{2\pi\sqrt{-1}} \sum_{t=1}^{\infty} \left(-1\right)^{t+1} \varkappa^t \int_\Gamma R\left(\zeta\right) \left(\mathcal{T}^{(1)} R\left(\zeta\right) \right)^t d\zeta.$$

Note that $P_i(0)$ and $P_i(1)$ are the projections on the spaces spanned by the i-th principal eigenvector of XX'/T and of $\bar{X}\bar{X}'/T$, respectively.

As will be seen from the proof of part iii) of Theorem 1, there exists a positive number ρ such that $\Pr(\max_{1\leq i\leq q} h_i < \rho) \to 0$ as $n \to \infty$. Therefore, $\Pr(r = \rho/2) \to 1$ as $n \to \infty$. Further, since $\left\|\mathcal{T}^{(1)}\right\| = \left\|\frac{\bar{X}\bar{X}'}{T} - \frac{XX'}{T}\right\| \leq |\mu_1 - \bar{\mu}_1| = o_p(1)$, we have: $\Pr\left(r^{-1}\left\|\mathcal{T}^{(1)}\right\|^{-1} > 1\right) \to 1$ as $n \to \infty$ so that the series for $P_i(\varkappa)$ displayed above converge for $\varkappa = 1$ with probability arbitrarily close to 1 for large enough n. Moreover, with probability arbitrarily close to 1 for large enough n, we have: $\|P_i(1) - P_i(0)\| \leq \sum_{t=1}^{\infty}\left\|\frac{1}{2\pi}\int_\Gamma R(\zeta)\left(\mathcal{T}^{(1)}R(\zeta)\right)^t d\zeta\right\| \leq \sum_{t=1}^{\infty} r^{-t}\left\|\mathcal{T}^{(1)}\right\|^t = \frac{\left\|\mathcal{T}^{(1)}\right\|}{r-\left\|\mathcal{T}^{(1)}\right\|} = o_p(1)$, which proves that the projections on the principal q eigenspaces of XX'/T and those on the principal q eigenspaces of $\bar{X}\bar{X}'/T$ converge in probability in operator norm.

In what follows, we will, therefore, assume that (1) holds. We will explicitly relax this assumption only when proving the convergence of the last $k-q$ columns of P.

1.2 A key lemma

Define $\tilde{X} = \sigma^{-1}[\mathcal{L}_1, ..., \mathcal{L}_n]' X [\mathcal{F}_1, ..., \mathcal{F}_T]$. Note that the j-th column of P equals the first k components of the unit-length j-th principal eigenvector of $\frac{1}{T}\tilde{X}\tilde{X}'$, and $\frac{\hat{L}'\hat{L}}{\sigma^2}$ equals a diagonal matrix with the first k eigenvalues of $\frac{1}{T}\tilde{X}\tilde{X}'$ on the diagonal. Lemma 2 below relates the eigenvalues and eigenvectors of the high-dimensional matrix $\frac{1}{T}\tilde{X}\tilde{X}'$ to the unit eigenvalues and the corresponding eigenvectors of the low-dimensional matrix-valued function $M^{(1)}(x)$, defined as follows.

Let us partition matrix ε as $[\varepsilon_1, \varepsilon_2]$, where ε_1 are the first k columns of ε. Further, let us denote $n \times n$ and $T \times T$ diagonal matrices with the j-th diagonal elements a_j and b_j as $\mathcal{A}_0$ and $\mathcal{B}_0$, respectively. Finally, let $\mathcal{A}$ and $\mathcal{B}$ be the $(n-k)\times(n-k)$ and $(T-k)\times(T-k)$ diagonal matrices with the j-th diagonal elements a_{k+j} and b_{k+j}, respectively. We define

$$\begin{aligned} M^{(1)}(x) &\equiv \tilde{\Psi}'\left(xI_n - \tilde{\Lambda}\right)^{-1}\tilde{\Psi}, \\ M^{(2)}(x) &\equiv \tilde{\Psi}'\left(xI_n - \tilde{\Lambda}\right)^{-2}\tilde{\Psi}, \text{ and} \\ M^{(3)}(x) &\equiv [I_k, 0]\left(xI_n - \tilde{\Lambda}\right)^{-1}\tilde{\Psi}, \end{aligned}$$

where $\tilde{\Psi}' = \frac{1}{\sqrt{T}}\left[\begin{array}{cc}\Delta', & 0\end{array}\right] + \frac{1}{\sqrt{T}}\varepsilon_1'\mathcal{A}_0$, $\Delta = \left(\frac{L'L}{\sigma^2}\right)^{1/2}(F'F)^{1/2}$ and $\tilde{\Lambda} = \frac{1}{T}\mathcal{A}_0\varepsilon_2\mathcal{B}^2\varepsilon_2'\mathcal{A}_0$. If $xI_n - \tilde{\Lambda}$ is not invertible, we set $M^{(j)}(x) = 0_{k\times k}$ for $j = 1, 2, 3$.

Lemma 2. *Let $\mu \neq \lambda_i\left(\tilde{\Lambda}\right)$, $i = 1, ..., n$ so that $\mu I_n - \tilde{\Lambda}$ is invertible. Then:*

i) *μ is an eigenvalue of $\frac{1}{T}\tilde{X}\tilde{X}'$ of multiplicity larger than or equal to s if and only if there* *exists a positive integer* $m \leq k+1-s$ *such that* $x=\mu$ *satisfies equations*

$$\lambda_m\left(M^{(1)}(x)\right)=1,...,\lambda_{m+s-1}\left(M^{(1)}(x)\right)=1, \tag{2}$$

ii) *If* v *is an eigenvector of* $M^{(1)}(\mu)$ *corresponding to eigenvalue* 1, *then*

$$y=\left(v'M^{(2)}(\mu)v\right)^{-1/2}\left(\mu I_n-\tilde{\Lambda}\right)^{-1}\tilde{\Psi}v \tag{3}$$

is a unit-length eigenvector of $\frac{1}{T}\tilde{X}\tilde{X}'$ *corresponding to eigenvalue* μ.

iii) *If* 1 *is a simple eigenvalue of* $M^{(1)}(\mu)$, *then* μ *is a simple eigenvalue of* $\frac{1}{T}\tilde{X}\tilde{X}'$. *Furthermore, if* μ *is the* j*-th largest eigenvalue of* $\frac{1}{T}\tilde{X}\tilde{X}'$ *and* v *is a corresponding eigenvector of* $M^{(1)}(\mu)$, *then the* j*-th column of matrix* P *from part ii) of Theorem 1 equals* $\left(v'M^{(2)}(\mu)v\right)^{-1/2}M^{(3)}(\mu)v$.

iv) *Consider matrix* $\frac{1}{T}\tilde{X}\tilde{X}'+\varkappa e_ie_i'$, *where* e_i *is the vector with all entries but the* i*-th equal to zero and the* i*-th entry equal to 1, and where* $\varkappa$ *is an arbitrary positive number. We have:* μ *is an eigenvalue of* $\frac{1}{T}\tilde{X}\tilde{X}'+\varkappa e_ie_i'$ *of multiplicity larger than or equal to* s *if and only if there exists a positive integer* $m\leq k+1-s$ *such that* $x=\mu$ *satisfies equations*

$$\lambda_m\left(M^{(1)}_{\varkappa i}(x)\right)=1,...,\lambda_{m+s-1}\left(M^{(1)}_{\varkappa i}(x)\right)=1,$$

where $M^{(1)}_{\varkappa i}(x)\equiv\tilde{\Psi}'_{\varkappa i}\left(xI_n-\tilde{\Lambda}\right)^{-1}\tilde{\Psi}_{\varkappa i}$ *and* $\tilde{\Psi}_{\varkappa i}\equiv\left[\tilde{\Psi},\sqrt{\varkappa}e_i\right]$.

Proof of Lemma 2: Let μ be an eigenvalue of $\frac{1}{T}\tilde{X}\tilde{X}'$ of multiplicity larger than or equal to s and let $y_1,...,y_s$ be orthonormal eigenvectors corresponding to μ. Since $\frac{1}{T}\tilde{X}\tilde{X}'=\tilde{\Lambda}+\tilde{\Psi}\tilde{\Psi}'$, we have: $\left(\tilde{\Lambda}+\tilde{\Psi}\tilde{\Psi}'\right)y_j=\mu y_j$ for $j=1,...,s$. Note that vectors $\tilde{\Psi}'y_1,...,\tilde{\Psi}'y_s$ are linearly independent. Otherwise, if $\sum_{j=1}^s\beta_j\tilde{\Psi}'y_j=0$ for some β_j that are not all equal to zero, we would have: $\tilde{\Lambda}\sum_{j=1}^s\beta_jy_j=\left(\tilde{\Lambda}+\tilde{\Psi}\tilde{\Psi}'\right)\sum_{j=1}^s\beta_jy_j=\mu\sum_{j=1}^s\beta_jy_j$, which violates our assumption that $\mu\neq\lambda_i\left(\tilde{\Lambda}\right)$, $i=1,...,n$. Equation $\left(\tilde{\Lambda}+\tilde{\Psi}\tilde{\Psi}'\right)y_j=\mu y_j$ implies that $\tilde{\Psi}'\left(\mu I_n-\tilde{\Lambda}\right)^{-1}\tilde{\Psi}\tilde{\Psi}'y_j=\tilde{\Psi}'y_j$. Hence, the space spanned by $\tilde{\Psi}'y_j$, $j=1,...,s$ is an invariant subspace of $M_n(\mu)$ with the corresponding eigenvalue equal to 1. This proves the "only if" part of i).

Suppose now that (2) holds with $x=\mu$. Let $v_1,...,v_s$ be orthonormal eigenvectors of $M^{(1)}(\mu)$ corresponding to eigenvalue 1. Define $y_1,...,y_s$ by (3) with v replaced by $v_1,...,v_s$, respectively. Vectors $y_1,...,y_s$ are unit-length vectors by definition of $M^{(2)}(\mu)$. Furthermore, they are linearly independent because, otherwise, if $\sum_{j=1}^s\beta_jy_j=0$ for some β_j that are not all equal to zero, we would have, for $\gamma_j=\left(v_j'M^{(2)}(\mu)v_j\right)^{-1/2}\beta_j$: $\sum_{j=1}^s\gamma_jv_j=$

$\sum_{j=1}^{s}\gamma_j M^{(1)}(\mu)v_j = \tilde{\Psi}'\sum_{j=1}^{s}\beta_j y_j = 0$, which violates our assumption that $v_1,...,v_s$ are orthonormal. Equation (3) implies that $\tilde{\Psi}'y_j = \left(v'M^{(2)}(\mu)v\right)^{-1/2}M^{(1)}(\mu)v_j = \left(v'M^{(2)}(\mu)v\right)^{-1/2}v_j$ and therefore, $y_j = \left(\mu I_n - \tilde{\Lambda}\right)^{-1}\tilde{\Psi}\tilde{\Psi}'y_j$ for all $j=1,...,s$. The latter equality implies that $\left(\tilde{\Lambda}+\tilde{\Psi}\tilde{\Psi}'\right)y_j = \mu y_j$, which means that y_j, $j=1,...,s$ are linearly independent eigenvectors of $\frac{1}{T}\tilde{X}\tilde{X}'$, each of which corresponds to eigenvalue μ. This proves ii) and the "if" part of i).

Part iii) of the lemma follows from parts i) and ii). Indeed, part i) implies that if 1 is a simple eigenvalue of $M^{(1)}(\mu)$, then μ is a simple eigenvalue of $\frac{1}{T}\tilde{X}\tilde{X}'$. Further, by definition, the j-th column of P equals the first k components of the unit-length j-th principal eigenvector of $\frac{1}{T}\tilde{X}\tilde{X}'$. This fact, part ii) of the lemma and the definition of $M^{(3)}(\mu)$ imply that the j-th column of P equals $\left(v'M^{(2)}(\mu)v\right)^{-1/2}M^{(3)}(\mu)v$.

Proof of part iv) of the lemma is almost identical to the proof of part i). We only need to replace $\tilde{\Psi}$ by $\tilde{\Psi}_{\varkappa i}$ and $M^{(2)}(x)$ by $M^{(2)}(\mu) \equiv \tilde{\Psi}'_{\varkappa i}\left(xI_n - \tilde{\Lambda}\right)^{-2}\tilde{\Psi}_{\varkappa i}$ in that proof.□

Below, we prove several technical lemmas to find the probability limits of $M^{(1)}(x)$, $M^{(2)}(x)$, $M^{(3)}(x)$ and $M^{(1)}_{\varkappa i}(x)$. We will then use these limits and Lemma 2 to derive the probability limits of the eigenvalues of $\frac{1}{T}\tilde{X}\tilde{X}'$ and of matrix P.

1.3 Technical lemmata

Lemma 3. (Bai and Silverstein, 1988) *Let* $\{\xi_i, i=1,...,2n\}$ *be i.i.d. random variables with mean zero and variance* 1. *Define* $\xi=(\xi_1,...,\xi_n)$, $\zeta=\left(\xi_{n+1},...,\xi_{2n}\right)$ *and let* Z *be an* $n\times n$ *random matrix independent from* ξ *and* ζ. *Then, for any* $p>0$, *we have:*

$$E\left(\left|\xi'Z\xi - \operatorname{tr} Z\right|^p|Z\right) \leq C_{1p}n^{p/2}\|Z\|^p\left(\left[E\left|\xi_1\right|^4\right]^{p/2} + E\left|\xi_1\right|^{2p}\right), \quad (4)$$

$$E\left(\left|\xi'Z\zeta\right|^p|Z\right) \leq C_{2p}n^{p/2}\|Z\|^p\left(\left[E\left|\xi_1\right|^4\right]^{p/2} + E\left|\xi_1\right|^{2p}\right), \quad (5)$$

where C_{1p} *and* C_{2p} *are constants that depend only on* p.

Proof of Lemma 3: Inequality (4) is a slightly simplified version of the statement of Lemma 2.7 in Bai and Silverstein (1998). Inequality (5) follows from (4). Indeed, consider a vector $\varphi = \left(\xi',\zeta'\right)'$ and consider matrix $\tilde{Z} = \begin{pmatrix} 0 & Z \\ Z' & 0 \end{pmatrix}$. We have: $E\left(\left|\xi'Z\zeta\right|^p|Z\right) = E\left(\left|\frac{1}{2}\varphi'\tilde{Z}\varphi\right|^p|Z\right) \leq 2^{-p}C_{1p}(2n)^{p/2}\left\|\tilde{Z}\right\|^p\left(\left[E\left|\xi_1\right|^4\right]^{p/2} + E\left|\xi_1\right|^{2p}\right)$, where the latter inequality follows from (4) because $\operatorname{tr}\tilde{Z} = 0$. It remains to note that $\left\|\tilde{Z}\right\| = \|Z\|$ and set $C_{2p} = 2^{-p/2}C_{1p}$.□

Lemma 4. *Let* Σ *and* Π *be two independent identically distributed random* $n\times k$

matrices with i.i.d. entries, which have finite fourth moment, $\mu_4 < \infty$. Further, let Z be a random $n \times n$ matrix independent from Σ and Π and such that $n \|Z\|^2 \xrightarrow{p} 0$ as $n \to \infty$. Then, as $n \to \infty$:

$$\left\|\Sigma' Z \Sigma - (\operatorname{tr} Z) I_k\right\| \xrightarrow{p} 0 \text{ and } \left\|\Sigma' Z \Pi\right\| \xrightarrow{p} 0.$$

Proof of Lemma 4: To save space, we omit the proof of $\|\Sigma' Z \Pi\| \xrightarrow{p} 0$ because it is similar to the proof of $\|\Sigma' Z \Sigma - (\operatorname{tr} Z) I_k\| \xrightarrow{p} 0$. Let δ_1 and δ_2 be arbitrary positive numbers. For the i-th diagonal element of $\Sigma' Z \Sigma$, we have: $\Pr\left(|(\Sigma' Z \Sigma)_{ii} - \operatorname{tr} Z| > \delta_1 | Z\right) \leq \delta_1^{-2} E\left(|(\Sigma' Z \Sigma)_{ii} - \operatorname{tr} Z|^2 |Z\right) \leq \delta_1^{-2} 2 C_{12} n \|Z\|^2 \mu_4$, where the first inequality is Chebyshev's inequality and the second inequality follows from Lemma 3. Next, since $n \|Z\|^2 \xrightarrow{p} 0$, there exists N such that for all $n > N$, $\Pr(\delta_1^{-2} 2 C_{12} n \|Z\|^2 \mu_4 < \delta_2/2) > 1 - \delta_2/2$. Therefore, for all $n > N$, $\Pr\left(|(\Sigma' Z \Sigma)_{ii} - \operatorname{tr} Z| > \delta_1\right) = E\left[\Pr\left(|(\Sigma' Z \Sigma)_{ii} - \operatorname{tr} Z| > \delta_1 | Z\right)\right] \leq \delta_2/2 \left(1 - \delta_2/2\right) + \delta_2/2 < \delta_2$, which proves that $|(\Sigma' Z \Sigma)_{ii} - \operatorname{tr} Z| \xrightarrow{p} 0$. The convergence $\left|(\Sigma' Z \Sigma)_{ij}\right| \xrightarrow{p} 0$ for $i \neq j$ can be proven similarly. Since k is fixed as $n \to \infty$, the entry-wise convergence of $\Sigma' Z \Sigma - (\operatorname{tr} Z) I_k$ to zero implies that $\|\Sigma' Z \Sigma - (\operatorname{tr} Z) I_k\| \xrightarrow{p} 0$. $\square$

Let us partition ε_1' into $[\varepsilon_{11}', \varepsilon_{21}']$, where ε_{11} is $k \times k$, and ε_2' into $[\varepsilon_{12}', \varepsilon_{22}']$, where ε_{12} is $k \times (T-k)$. In the lemmas below, we will need the following new notation. Denote matrix $x I_{n-k} - \frac{1}{T} \mathcal{A} \varepsilon_{22} \mathcal{B}^2 \varepsilon_{22}' \mathcal{A}$ as Y; the i-th column of ε_{22} as $\varepsilon_{22,i}$; matrix ε_{22} with the i-th column removed as $\varepsilon_{22,-i}$; matrix $\mathcal{B}$ with i-th row and i-th column removed as $\mathcal{B}_{-i}$; and, finally, matrix $x I_{n-k} - \frac{1}{T} \mathcal{A} \varepsilon_{22,-i} \mathcal{B}_{-i}^2 \varepsilon_{22,-i}' \mathcal{A}$ as Y_i. In order to simplify notation, we do not explicitly indicate the dependence of Y and Y_i on x.

Lemma 5. *Suppose that Assumptions 1-3 hold. Let θ_1 be any number such that $\theta_1 > \bar{w}$, where $\bar{w}$ is as in Theorem 1. Then, for any $x > \theta_1$, Y is a positive definite matrix with $\left\|Y^{-1}\right\| \leq (\theta_1 - \bar{w})^{-1}$ for large n with probability 1. Further, whenever Y is a positive definite matrix, Y_i is also a positive definite matrix with $\left\|Y_i^{-1}\right\| \leq \left\|Y^{-1}\right\|$ and the following interlacing inequalities hold:*

$$\lambda_1\left(Y^{-1}\right) \geq \lambda_1\left(Y_i^{-1}\right) \geq \lambda_2\left(Y^{-1}\right) \geq \ldots \geq \lambda_n\left(Y^{-1}\right) \geq \lambda_n\left(Y_i^{-1}\right) \tag{6}$$

and similarly,

$$\lambda_1\left(\mathcal{A} Y^{-1} \mathcal{A}\right) \geq \lambda_1\left(\mathcal{A} Y_i^{-1} \mathcal{A}\right) \geq \lambda_2\left(\mathcal{A} Y^{-1} \mathcal{A}\right) \geq \ldots \geq \lambda_n\left(\mathcal{A} Y^{-1} \mathcal{A}\right) \geq \lambda_n\left(\mathcal{A} Y_i^{-1} \mathcal{A}\right) \tag{7}$$

Proof of Lemma 5: That Y is positive definite for $x > \theta_1$ and large n with probability 1 follows from Lemma 3 in Onatski (2005), which implies that under Assumptions 1-3, the largest eigenvalue of $\frac{1}{T} \mathcal{A} \varepsilon_{22} \mathcal{B}^2 \varepsilon_{22}' \mathcal{A}$ almost surely converges to the upper boundary of support of the cdf $\mathcal{F}^{c,A,B}$, which is the limit of the empirical distribution of eigenvalues of

$\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$ as n goes to infinity. As follows from the proof of Lemma 3 in Onatski (2005), the upper boundary of support of $\mathcal{F}^{c,A,B}$ equals $\bar{w}$, which is smaller than x by assumption, hence the positive definiteness of $Y(x) = xI_{n-k} - \frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$. The convergence of the largest eigenvalue of $\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$ to $\bar{w}$ also implies that, for any $x > \theta_1$, $\left\|Y^{-1}\right\| \leq (\theta_1 - \bar{w})^{-1}$ for large n with probability 1.

Matrix Y_i is positive definite whenever Y is because $Y_i - Y = \frac{1}{T}b_{i+k}^2\mathcal{A}\varepsilon_{22,i}\varepsilon_{22,i}'\mathcal{A}$ is a positive semidefinite matrix. The latter fact also implies that $\left\|Y_i^{-1}\right\| \leq \left\|Y^{-1}\right\|$ and that (see Corollary 4.3.3 in Horn and Johnson, 1985) $\lambda_j(Y_i) \geq \lambda_j(Y)$ for any $j = 1, ..., n$. Further, since $Y_i - Y$ is a rank-one matrix, we have by interlacing theorem (Theorem 4.3.4 in Horn and Johnson, 1985): $\lambda_{j+1}(Y_i) \leq \lambda_j(Y)$ for $j = 1, ..., n-1$. Combining the latter two inequalities, and using the fact that $\lambda_j\left(M^{-1}\right) = \lambda_{n-j+1}^{-1}(M)$ for any positive definite Hermitian matrix, we obtain: $\lambda_j\left(Y_i^{-1}\right) \leq \lambda_j\left(Y^{-1}\right)$ for any $j = 1, ..., n$ and $\lambda_j\left(Y_i^{-1}\right) \geq \lambda_{j+1}\left(Y^{-1}\right)$ for $j = 1, ..., n-1$, which implies the first set of the interlacing inequalities in the statement of Lemma 5. The second set of the interlacing inequalities can be established similarly by noting that $\mathcal{A}^{-1}Y_i\mathcal{A}^{-1} - \mathcal{A}^{-1}Y\mathcal{A}^{-1} = \frac{1}{T}b_{i+k}^2\varepsilon_{22,i}\varepsilon_{22,i}'$ is a rank-one positive semidefinite matrix. $\square$

Lemma 6. *Suppose that Assumptions 1-3 hold. Let θ_1 be any number such that $\theta_1 > \bar{w}$, where $\bar{w}$ is as in Theorem 1. Then, for any $x > \theta_1$ and any pair of integers (r, s) from the set $\{(1,1), (1,2), (2,1)\}$, we have:*

$$\max_{1 \leq i \leq T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\left[\mathcal{A}Y_i^{-r}\mathcal{A}\right]^s\varepsilon_{22,i} - \frac{1}{T}\operatorname{tr}\left[\mathcal{A}Y^{-r}\mathcal{A}\right]^s\right| \xrightarrow{p} 0,$$

where, if either Y_i or Y is not invertible, we set the maximized absolute difference to an arbitrary non-zero number, say 1.

Proof of Lemma 6: Let us define $\bar{Y}_i \equiv Y_i$ when Y_i is invertible and $\bar{Y}_i \equiv -I_{n-k}$ when Y_i is not invertible. Similarly, define $\bar{Y} \equiv Y$ when Y is invertible and $\bar{Y} \equiv -I_{n-k}$ when Y is not invertible. It is enough to prove the lemma for Y_i replaced by $\bar{Y}_i$ and Y replaced by $\bar{Y}$. It is because, first, $\left\{Y_i \neq \bar{Y}_i \text{ for some } i \leq T-k, \text{ or } Y \neq \bar{Y}\right\} \cap \Omega = \varnothing$, where Ω is the event that happens if and only if Y is positive definite and $\left\|Y^{-1}\right\| \leq (\theta_1 - \bar{w})^{-1}$, and second, by Lemma 5, $\Pr(\Omega) \to 1$ as $n \to \infty$. The above intersection of events is empty because, first, as follows from Lemma 5, Ω implies Ω_i, which is defined as the event that happens if and only if Y_i is positive definite and $\left\|Y_i^{-1}\right\| \leq (\theta_1 - \bar{w})^{-1}$, and second, if Ω and Ω_i hold, then $Y = \bar{Y}$ and $Y_i = \bar{Y}_i$.

Let us decompose the difference $\frac{1}{T}\varepsilon_{22,i}'\left[\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right]^s\varepsilon_{22,i} - \frac{1}{T}\operatorname{tr}\left[\mathcal{A}\bar{Y}^{-r}\mathcal{A}\right]^s$ into a sum $U_{rs}(i) + V_{rs}(i)$, where $U_{rs}(i) = \frac{1}{T}\varepsilon_{22,i}'\left[\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right]^s\varepsilon_{22,i} - \frac{1}{T}\operatorname{tr}\left[\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right]^s$ and $V_{rs}(i) = \frac{1}{T}\operatorname{tr}\left[\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right]^s - \frac{1}{T}\operatorname{tr}\left[\mathcal{A}\bar{Y}^{-r}\mathcal{A}\right]^s$. To prove our lemma, it is enough to show that $\max_{1 \leq i \leq T-k}|U_{rs}(i)| \xrightarrow{p} 0$ and $\max_{1 \leq i \leq T-k}|V_{rs}(i)| \xrightarrow{p} 0$. Below, we will establish the latter two convergences.

Let δ_1 and δ_2 be arbitrary positive numbers. Note that $\Pr(\Omega) > 1 - \delta_2/2$ for large enough n. Therefore, and since $\Omega \subseteq \Omega_i$, we have:

$$\Pr\left(\max_{1\leq i\leq T-k} |U_{rs}(i)| > \delta_1\right) \leq \Pr\left(\max_{1\leq i\leq T-k} |U_{rs}(i)| > \delta_1 \text{ and } \Omega\right) + \delta_2/2 \leq \tag{8}$$
$$\sum_{i=1}^{T-k} \Pr\left(|U_{rs}(i)| > \delta_1 \text{ and } \Omega_i\right) + \delta_2/2 \leq \sum_{i=1}^{T-k} E\left[\Pr\left(|U_{rs}(i)| > \delta_1 \text{ and } \Omega_i | \bar{Y}_i\right)\right] + \delta_2/2$$

If either $\bar{Y}_i$ is not positive definite or $\left\|\bar{Y}_i^{-1}\right\| > (\theta_1 - \bar{w})^{-1}$, then $\Pr\left(|U_{rs}(i)| > \delta_1 \text{ and } \Omega_i | \bar{Y}_i\right) = 0$. In contrast, if $\bar{Y}_i$ is positive definite and $\left\|\bar{Y}_i^{-1}\right\| \leq (\theta_1 - \bar{w})^{-1}$, then $\Pr\left(|U_{rs}(i)| > \delta_1 \text{ and } \Omega_i | \bar{Y}_i\right) = \Pr\left(|U_{rs}(i)| > \delta_1 | \bar{Y}_i\right)$. But, by Markov's inequality:

$$\Pr\left(|U_{rs}(i)| > \delta_1 | \bar{Y}_i\right) \leq \delta_1^{-p} E\left(|U_{rs}(i)|^p | \bar{Y}_i\right) \leq$$
$$\leq \delta_1^{-p} C_{1p} \frac{(n-k)^{p/2}}{T^p} \|\mathcal{A}\|^{2sp} \left\|\bar{Y}_i^{-1}\right\|^{rsp} \left(\left[E|\varepsilon_{jt}|^4\right]^{p/2} + (\ln n)^{2p}\right),$$

where the second line follows from Lemma 3 and from assumption (1). If $\left\|\bar{Y}_i^{-1}\right\| \leq (\theta_1 - \bar{w})^{-1}$, we can make the latter expression smaller than $\delta_2/(2T)$ by choosing $p > 2$ and large enough n. Therefore, $E\left[\Pr\left(|U_{rs}(i)| > \delta_1 \text{ and } \Omega_i | \bar{Y}_i\right)\right] \leq \delta_2/(2T)$ for all $i = 1, ..., T-k$ and large enough n. Using (8), we obtain: $\Pr\left(\max_{1\leq i\leq T-k} |U_{rs}(i)| > \delta_1\right) < \delta_2$ for large enough n. Since δ_1 and δ_2 were arbitrary positive numbers, we have: $\max_{1\leq i\leq T-k} |U_{rs}(i)| \xrightarrow{p} 0$.

Next, when Ω takes place so that $\bar{Y}_i$ and $\bar{Y}$ are positive definite, we have:

$$V_{rs}(i) \equiv \frac{1}{T}\sum_{j=1}^{n-k}\left[\lambda_j\left(\left[\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right]^s\right) - \lambda_j\left(\left[\mathcal{A}\bar{Y}^{-r}\mathcal{A}\right]^s\right)\right] =$$
$$\frac{1}{T}\sum_{j=1}^{n-k}\left[\lambda_j^s\left(\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right) - \lambda_j^s\left(\mathcal{A}\bar{Y}^{-r}\mathcal{A}\right)\right] = -\frac{1}{T}\lambda_1^s\left(\mathcal{A}\bar{Y}^{-r}\mathcal{A}\right) +$$
$$\frac{1}{T}\sum_{j=1}^{n-k-1}\left[\lambda_j^s\left(\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right) - \lambda_{j+1}^s\left(\mathcal{A}\bar{Y}^{-r}\mathcal{A}\right)\right] + \frac{1}{T}\lambda_{n-k}^s\left(\mathcal{A}\bar{Y}_i^{-r}\mathcal{A}\right).$$

Therefore, setting $r = 1$ and using the interlacing inequalities (7), we conclude that

$$0 \geq V_{1s}(i) \geq -\frac{1}{T}\lambda_1^s\left(\mathcal{A}\bar{Y}^{-1}\mathcal{A}\right) \geq -\frac{1}{T}\|\mathcal{A}\|^{2s}(\theta_1 - \bar{w})^{-s} \tag{9}$$

whenever Ω holds. Since $\Pr(\Omega) \to 1$ as $n \to \infty$, the latter inequalities imply that $\max_{1\leq i\leq T-k} |V_{1s}(i)| \xrightarrow{p} 0$ for $s = 1$ and $s = 2$.

It remains to prove the convergence to zero of $\max_{1\leq i\leq T-k}\left|V_{21}(i)\right|$. Note that if Ω holds, $\bar{Y}^{-2}-\bar{Y}_i^{-2}$ is a positive semidefinite matrix so that, in particular, all its diagonal elements are not negative. Therefore, if Ω holds, we have: $0\geq V_{21}(i)\geq-\left(\max_{j=1,\ldots,n-k}a_{j+k}^2\right)\frac{1}{T}\operatorname{tr}\left(\bar{Y}^{-2}-\bar{Y}_i^{-2}\right)$. But $\max_{j=1,\ldots,n-k}a_{j+k}^2=\left\|\mathcal{A}\right\|^2$ and $\frac{1}{T}\operatorname{tr}\left(\bar{Y}^{-2}-\bar{Y}_i^{-2}\right)\leq\frac{1}{T}\lambda_1^2\left(\bar{Y}^{-1}\right)=\frac{1}{T}\left\|\bar{Y}^{-1}\right\|^2\leq\frac{1}{T}\left(\theta_1-\bar{w}\right)^{-2}$ when Ω holds, where the first of the latter two inequalities can be obtained from the interlacing inequalities (6) similarly to as (9) was obtained from (7). Thus, when Ω hold, we have: $0\geq V_{21}(i)\geq-\frac{1}{T}\left\|\mathcal{A}\right\|^2\left(\theta_1-\bar{w}\right)^{-2}$, and therefore $\max_{1\leq i\leq T-k}\left|V_{21}(i)\right|\xrightarrow{p}0.\square$

Lemma 7. *Let θ_1 be any number such that $\theta_1>\bar{w}$, where $\bar{w}$ is as in Theorem 1 and let x be any number larger than θ_1. Then, for any complex z such that $\operatorname{Im}z>0$, the equation $m(z)=\int\frac{\lambda d\mathcal{F}^A(\lambda)}{x-\left(z+\int\frac{\tau d\mathcal{F}^B(\tau)}{1-\tau cm(z)}\right)\lambda}$ has a unique solution $m(z)$ such that $\operatorname{Im}m(z)>0$. Function $m(z)$ is analytic for $\operatorname{Im}z>0$ and can be analytically continued to a small open neighborhood of $z=0$. If Assumptions 1-3 hold, then:*

$$\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}\overset{a.s.}{\to}cm(0)\text{ and }\frac{1}{T}\operatorname{tr}\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\overset{a.s.}{\to}cm'(0),$$

where, if Y is not invertible, we set the left hand of the above convergence statements to an arbitrary number, which equals neither $cm(0)$ nor $cm'(0)$. The above convergence statements remain valid if we replace $n-k,\mathcal{A}$ and ε_{22} in the definition $Y\equiv xI_{n-k}-\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$ by $n,\mathcal{A}_0$ and ε_2, respectively.

Proof of Lemma 7: Let $m_n(z)$ and $\tilde{m}_n(z)$ be the Stieltjes transforms of the empirical eigenvalue distributions of $x\mathcal{A}^{-2}-\frac{1}{T}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'$ and $x\frac{T}{n-k}\mathcal{A}^{-2}-\frac{1}{n-k}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'$, respectively. Note that $m_n(z)=\frac{T}{n-k}\tilde{m}_n\left(\frac{T}{n-k}z\right)$. Silverstein and Bai (1995) show that, for any z with $\operatorname{Im}z>0$, as $n\to\infty$, $\tilde{m}_n(z)$ almost surely converges to $\tilde{m}(z)$, which is an analytic function in the $\operatorname{Im}z>0$ domain and which is the unique solution to equation[1] $\tilde{m}(z)=m_A\left(z+c^{-1}\int\frac{\tau d\mathcal{F}^B(\tau)}{1-\tau\tilde{m}(z)}\right)$ that satisfies $\operatorname{Im}\tilde{m}(z)>0$. Here, $m_A(z)$ is the Stieltjes transform of a (possibly defective) non-random distribution function which is the vague limit[2] of the empirical spectral distribution of $x\frac{T}{n-k}\mathcal{A}^{-2}$ as $n\to\infty$.

Note that the cdf of the latter vague limit at λ equals the limit of the proportion of those eigenvalues of $x\frac{T}{n-k}\mathcal{A}^{-2}$, which are no larger than λ. By Assumptions 1i) and 3i), such a limit equals $1-\lim_{\tau\uparrow xc^{-1}\lambda^{-1}}\mathcal{F}^A(\tau)$. Hence, $m_A(z)=\int\frac{\tau d\mathcal{F}^A(\tau)}{xc^{-1}-z\tau}$ and $\tilde{m}(z)=\int\frac{\lambda d\mathcal{F}^A(\lambda)}{xc^{-1}-\left(z+c^{-1}\int\frac{\tau d\mathcal{F}^B(\tau)}{1-\tau\tilde{m}(z)}\right)\lambda}$. Recalling that $m_n(z)=\frac{T}{n-k}\tilde{m}_n\left(\frac{T}{n-k}z\right)$, we conclude that for any z with $\operatorname{Im}z>0$, $m_n(z)$ converges to $m(z)=c^{-1}\tilde{m}\left(c^{-1}z\right)$, which is an analytic

---

[1] Note the difference in notation: their c is our c^{-1}, their n is our T and their T_N is our $-\mathcal{B}^2$ so that their $dH(\tau)$ is our $-d\mathcal{F}^B(-\tau)$.

[2] The vague convergence is a generalization of the weak convergence to sub-probability measures. For a definition of the vague convergence see, for example, Athreya and Lahiri (2006), chapter 9.2.

function in the $\operatorname{Im} z > 0$ domain and which is the unique solution to equation $m(z) = \int \frac{\lambda d\mathcal{F}^A(\lambda)}{x-\left(z+\int \frac{\tau d\mathcal{F}^B(\tau)}{1-\tau c m(z)}\right)\lambda}$ that satisfies $\operatorname{Im} m\left(z\right) > 0$.

Now, let U_0 be an open disk in the complex plane with center at zero and radius $\frac{1}{2}\frac{\theta_1-\bar{w}}{u(\mathcal{F}^A)}$. Note that the smallest eigenvalue of $x\mathcal{A}^{-2} - \frac{1}{T}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'$ is no smaller than $\frac{\theta_1-\bar{w}}{u(\mathcal{F}^A)}$ for large n with probability 1. Therefore, $m_n\left(z\right)$ are analytic in U_0 and bounded there by $\left(\frac{1}{2}\frac{\theta_1-\bar{w}}{u(\mathcal{F}^A)}\right)^{-1}$ for large n with probability 1. Moreover, as has been just shown, $m_n(z)$ almost surely converges to $m(z)$ for any $\operatorname{Im} z > 0$. Therefore, by Vitali-Porter theorem (see p.44 of Schiff, 1993), $m_n(z)$ converge (almost surely) to $m(z)$ uniformly on compact subsets of U_0 and $m(z)$ is analytic in U_0. Note that since $m_n(z) = \frac{1}{n-k}\operatorname{tr}\left(x\mathcal{A}^{-2} - \frac{1}{T}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}' - zI_{n-k}\right)^{-1}$, matrix $x\mathcal{A}^{-2} - \frac{1}{T}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'$ (and therefore also Y) is invertible and $\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A} = \frac{n-k}{T}m_n(0)$ whenever $m_n(z)$ is analytic in U_0. Therefore, we have: $\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A} \overset{a.s.}{\to} cm(0)$.

Next, note that, whenever $m_n(z)$ are analytic in U_0, $\frac{1}{T}\operatorname{tr}\left(\mathcal{A}Y^{-1}\mathcal{A}\right)^2 = \frac{n-k}{T}m_n'(0)$. Since, with probability 1 for large n, $m_n(z)$ are analytic functions on U_0 which converge uniformly on compact subsets of U_0 to $m(z)$, by classical Weirstrass theorem, the derivatives of $m_n(z)$ also converge to the corresponding derivatives of $m(z)$, and this convergence is uniform on the compact subsets of U_0. We therefore have: $\frac{1}{T}\operatorname{tr}\left(\mathcal{A}Y^{-1}\mathcal{A}\right)^2 \overset{a.s}{\to} cm'\left(0\right)$.

To adapt the above proof to the situation when $n-k, \mathcal{A}$ and ε_{22} in the definition $Y \equiv xI_{n-k} - \frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$ are replaced by $n, \mathcal{A}_0$ and ε_2, we only need to replace $n-k, \mathcal{A}$ and ε_{22} in the above arguments by $n, \mathcal{A}_0$ and ε_2, respectively.$\square$

Lemma 8. *Suppose that Assumptions 1-3 hold. Let* θ_1 *be any number such that* $\theta_1 > \bar{w}$, *where* $\bar{w}$ *is as in Theorem 1. Further, for any* $w > \bar{w}$, *let* $(u_{2,w}, v_{2,w})$ *be the bigger of the two solutions to system*

$$\begin{cases} v = w\left(c\int \frac{au}{u-a}d\mathcal{F}^A(a)\right)^{-1} \\ u = w\left(\int \frac{bv}{v-b}d\mathcal{F}^B(b)\right)^{-1} \end{cases} \tag{10}$$

and let

$$\begin{aligned} r_{u,w} &= \int\left(\frac{\lambda}{u_{2,w}-\lambda}\right)^2 d\mathcal{F}^A(\lambda)/\int\frac{\lambda}{u_{2,w}-\lambda}d\mathcal{F}^A(\lambda) \text{ and} \\ r_{v,w} &= \int\left(\frac{\tau}{v_{2,w}-\tau}\right)^2 d\mathcal{F}^B(\tau)/\int\frac{\tau}{v_{2,w}-\tau}d\mathcal{F}^B(\tau). \end{aligned}$$

Then for any $x > \theta_1$ *and function* $m(z)$ *defined in Lemma 7, we have:*

$$cm(0) = v_{2,x}^{-1} \text{ and } cm'(0) = \frac{r_{u,x}u_{2,x}}{\left(1-r_{u,x}r_{v,x}\right)xv_{2,x}}.$$

Proof of Lemma 8: Consider two functions of three complex variables: $f_1(z,u,v) =$

$x + cuv\left(1 + um_A(u)\right)$ and $f_2\left(z, u, v\right) = x - uz + uv\left(1 + vm_B(v)\right),$ where $m_A\left(u\right)$ and $m_B\left(v\right)$ are the Stieltjes transforms of $\mathcal{F}^A$ and $\mathcal{F}^B$, respectively. Further, consider a system $\begin{cases} f_1(z,u,v) = 0 \\ f_2\left(z,u,v\right) = 0 \end{cases}$. Note that f_1 and f_2 are holomorphic functions of z, u and v near the point $\left(z, u, v\right) = \left(0, u_{2,x}, v_{2,x}\right)$. This follows from the fact that $m_A(u)$ and $m_B\left(v\right)$ are holomorphic at $u_{2,x}$ and $v_{2,x}$, respectively, which, in turn, follows from the fact that $u_{2,x} > u\left(\mathcal{F}^A\right)$ and $v_{2,x} > u\left(\mathcal{F}^B\right)$.

According to the holomorphic implicit function theorem (see Krantz (1992), p.54), there exists a unique holomorphic solution $\{u(z), v\left(z\right)\}$ to the above system in a neighborhood of $z = 0$ such that $u(0) = u_{2,x}$ and $v\left(0\right) = v_{2,x}$ as long as $\det f_{1,2}' \neq 0$ at $\left(z, u, v\right) = \left(0, u_{2,x}, v_{2,x}\right)$, where $f_{1,2}' = \begin{pmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{pmatrix}$.

By assumption, the curves in the (u, v)-plane, $u = g_1(v)$ and $u = g_2(v)$, defined by the equations of (10): $v = x\left(c \int \frac{\lambda g_1(v)}{g_1(v) - \lambda} d\mathcal{F}^A(\lambda)\right)^{-1}$ and $g_2(v) = x\left(\int \frac{\tau v}{v - \tau} d\mathcal{F}^B(\tau)\right)^{-1}$, respectively, intersect at $(u, v) = (u_{2,x}, v_{2,x})$ so that $\frac{d}{dv} g_2(v) < \frac{d}{dv} g_1(v)$ at $(u, v) = \left(u_{2,x}, v_{2,x}\right)$. The latter inequality is equivalent to the inequality

$$\frac{d}{du}\left[x\left(c \int \frac{\lambda u}{u - \lambda} d\mathcal{F}^A(\lambda)\right)^{-1}\right] \frac{d}{dv}\left[x\left(\int \frac{\tau v}{v - \tau} d\mathcal{F}^B(\tau)\right)^{-1}\right] < 1. \tag{11}$$

at $\left(u, v\right) = \left(u_{2,x}, v_{2,x}\right)$. Note that by definition of Stieltjes transform, $c \int \frac{\lambda u}{u - \lambda} d\mathcal{F}^A(\lambda) = -cu\left(1 + um_A(u)\right)$ and $\int \frac{\tau v}{v - \tau} d\mathcal{F}^B(\tau) = -v\left(1 + vm_B(v)\right)$ so that, by definition of functions f_1 and f_2, we have:

$$\begin{array}{c} c \int \frac{\lambda u}{u - \lambda} d\mathcal{F}^A(\lambda) = -\frac{\partial f_1}{\partial v}, \ \int \frac{\tau v}{v - \tau} d\mathcal{F}^B(\tau) = -\frac{\partial f_2}{\partial u} \text{ and} \\ \frac{d}{du} c \int \frac{\lambda u}{u - \lambda} d\mathcal{F}^A(\lambda) = -\frac{1}{v} \frac{\partial f_1}{\partial u}, \ \frac{d}{dv} \int \frac{\tau v}{v - \tau} d\mathcal{F}^B(\tau) = -\frac{1}{u} \frac{\partial f_2}{\partial v} \end{array} \tag{12}$$

at $(z, u, v) = (0, u_{2,x}, v_{2,x})$. Using (12) in (11), we obtain:

$$x^2\left(-\frac{\partial f_1}{\partial v}\right)^{-2}\left(-\frac{\partial f_2}{\partial u}\right)^{-2}\left(-\frac{1}{v}\frac{\partial f_1}{\partial u}\right)\left(-\frac{1}{u}\frac{\partial f_2}{\partial v}\right) < 1 \tag{13}$$

at $(z, u, v) = (0, u_{2,x}, v_{2,x})$. But at $(u, v) = \left(u_{2,x}, v_{2,x}\right)$, the curves $u = g_1(v)$ and $u = g_2(v)$ intersect. Therefore, $v_{2,x} u_{2,x} = x^2\left(c \int \frac{\lambda u_{2,x}}{u_{2,x} - \lambda} d\mathcal{F}^A(\lambda)\right)^{-1}\left(\int \frac{\tau v_{2,x}}{v_{2,x} - \tau} d\mathcal{F}^B(\tau)\right)^{-1}$, and hence, $\left(-\frac{\partial f_1}{\partial v}\right)\left(-\frac{\partial f_2}{\partial u}\right) = x^2 / \left(v_{2,x} u_{2,x}\right)$, which, together with (13), implies that $\det f_{1,2}' < 0$ at $\left(z, u, v\right) = \left(0, u_{2,x}, v_{2,x}\right)$.

Now, the vector of derivatives $\left(\frac{du}{dz}, \frac{dv}{dz}\right)'$ evaluated at $z = 0$ equals vector $-\left(f_{1,2}'\right)^{-1}\left(\frac{\partial f_1}{\partial z}, \frac{\partial f_2}{\partial z}\right)'$

evaluated at $(z,u,v)=(0,u_{2,x},v_{2,x})$. But $\left(\frac{\partial f_1}{\partial z},\frac{\partial f_2}{\partial z}\right)=(0,-u_{2,x})$. Thus,

$$\left(\frac{du}{dz},\frac{dv}{dz}\right)=-u_{2,x}\det\left(f_{1,2}'\right)^{-1}\left(\frac{\partial f_1}{\partial v},-\frac{\partial f_1}{\partial u}\right). \tag{14}$$

Equations (12) together with the fact that $u_{2,x}>u\left(\mathcal{F}^A\right)$ and $v_{2,x}>u\left(\mathcal{F}^B\right)$ imply that $\frac{\partial f_1}{\partial v}<0$, $\frac{\partial f_1}{\partial u}>0$ at $(z,u,v)=(0,u_{2,x},v_{2,x})$. Therefore, and since, as has been shown, the determinant in (14) is negative,

$$\frac{du}{dz}<0\text{ and }\frac{dv}{dz}<0 \tag{15}$$

at $z=0$.

The last of the two inequalities in (15) and the fact that $\operatorname{Im}(v(0))=0$ imply that $\operatorname{Im}v^{-1}(z)>0$ for z, which are near 0 and such that $\operatorname{Im}z>0$. Note that, by definition, $v^{-1}(z)=c\int\frac{\lambda d\mathcal{F}^A(\lambda)}{x-\left(z+\int\frac{\tau d\mathcal{F}^B(\tau)}{1-\tau v^{-1}(z)}\right)\lambda}$. On the other hand, according to Lemma 7, for z such that $\operatorname{Im}z>0$, the unique $v^{-1}(z)$ satisfying the latter equation such that $\operatorname{Im}v^{-1}(z)>0$ must equal $cm(z)$. Hence, $cm(z)=v^{-1}(z)$, and $cm(0)=v_{2,x}^{-1}$.

Next, for the derivative of $v^{-1}(z)$ at $z=0$, we have: $\frac{d}{dz}v^{-1}(0)=-v_{2,x}^{-2}\frac{d}{dz}v(0)$. Using (14) and (12), we get:$\left(\frac{du}{dz},\frac{dv}{dz}\right)=-u_{2,x}\det\left(f_{1,2}'\right)^{-1}\left(\frac{\partial f_1}{\partial v},-\frac{\partial f_1}{\partial u}\right)$. Therefore,

$$\frac{d}{dz}v^{-1}(0)=-v_{2,x}^{-2}u_{2,x}\det\left(f_{1,2}'\right)^{-1}\left.\frac{\partial f_1}{\partial u}\right|_{u=u_{2,x}}=-v_{2,x}^{-1}u_{2,x}\det\left(f_{1,2}'\right)^{-1}c\int\frac{\lambda^2}{\left(u_{2,x}-\lambda\right)^2}d\mathcal{F}^A(\lambda),$$

where the latter equality follows from (12). Using the definition of the determinant $\det\left(f_{1,2}'\right)^{-1}$ and, once again, equations (12), we obtain:

$$\begin{aligned}\det\left(f_{1,2}'\right) &= cu_{2,x}v_{2,x}\int\frac{\lambda^2}{\left(u_{2,x}-\lambda\right)^2}d\mathcal{F}^A(\lambda)\int\frac{\tau^2}{\left(v_{2,x}-\tau\right)^2}d\mathcal{F}^B(\tau)-\\ &\quad cu_{2,x}v_{2,x}\int\frac{\lambda}{u_{2,x}-\lambda}d\mathcal{F}^A(\lambda)\int\frac{\tau}{v_{2,x}-\tau}d\mathcal{F}^B(\tau)\end{aligned}$$

so that, finally,

$$cm'(0)=\frac{d}{dz}v^{-1}(0)=-\frac{v_{2,x}^{-2}\int\frac{\lambda^2 d\mathcal{F}^A(\lambda)}{\left(u_{2,x}-\lambda\right)^2}}{\int\frac{\lambda^2 d\mathcal{F}^A(\lambda)}{\left(u_{2,x}-\lambda\right)^2}\int\frac{\tau^2 d\mathcal{F}^B(\tau)}{\left(v_{2,x}-\tau\right)^2}-\int\frac{\lambda d\mathcal{F}^A(\lambda)}{u_{2,x}-\lambda}\int\frac{\tau d\mathcal{F}^B(\tau)}{v_{2,x}-\tau}}=\frac{r_{u,x}u_{2,x}}{\left(1-r_{u,x}r_{v,x}\right)xv_{2,x}},$$

where the latter equality follows from the definition of $r_{u,x}$ and $r_{v,x}$ and from the fact that $(u_{2,x},v_{2,x})$, being a solution to system (10), satisfy $u_{2,x}=x\left(\int\frac{\tau v_{2,x}d\mathcal{F}^B(\tau)}{v_{2,x}-\tau}\right)^{-1}$.□

Lemma 9. *Under assumptions of Lemma 8:*

i) $M^{(1)}(x) \xrightarrow{p} x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1}\sigma^{-2}D+v_{2,x}^{-1}I_k$,

ii) $M^{(2)}(x) \xrightarrow{p} \sigma^{-2}Dx^{-2}\left(1-u_{2,x}^{-1}\right)^{-2}\left(1+\frac{r_{v,x}(1+r_{u,x})}{(1-r_{u,x}r_{v,x})u_{2,x}}\right)+\frac{1+r_{u,x}}{(1-r_{u,x}r_{v,x})xv_{2,x}}I_k$,

iii) $M^{(3)}(x) \xrightarrow{p} \sigma^{-1}x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1}D^{1/2}$.

iv) $M_{\varkappa i}^{(1)}(x) \xrightarrow{p} x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1}\sigma^{-2}\begin{pmatrix} D & \sigma\sqrt{\varkappa d_i}e_i \\ \sigma\sqrt{\varkappa d_i}e_i' & \sigma^2\varkappa \end{pmatrix}+v_{2,x}^{-1}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$, *where* e_i *is a* $k\times 1$ *vector with the* i*-th entry 1 and all the other entries zero.*

Proof of Lemma 9 i): Let us consider the following partitioned matrix:

$$xI_n-\tilde{\Lambda}\equiv\begin{pmatrix} xI_k-\frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{12}' & -\frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{22}'\mathcal{A} \\ -\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{12}' & Y \end{pmatrix}, \tag{16}$$

where $Y\equiv xI_{n-k}-\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$. Lemma 5 proves that, for any $x>\theta_1>\bar{w}$, matrix Y is positive definite (and hence invertible) for large n with probability 1. Replacing $n-k,\mathcal{A}$ and ε_{22} in that proof by $n,\mathcal{A}_0$ and ε_2, respectively, we establish the invertibility of matrix $xI_n-\tilde{\Lambda}\equiv xI_n-\frac{1}{T}\mathcal{A}_0\varepsilon_2\mathcal{B}^2\varepsilon_2'\mathcal{A}_0$ for large n with probability 1. Below, we will work with Y and $xI_n-\tilde{\Lambda}$ as if they were invertible matrices for large n, keeping in mind that this is indeed so almost surely.

The following formula for the inverse of a partitioned matrix A is well known. If A_{22} is not singular, then:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}=\begin{bmatrix} \Upsilon^{-1} & -\Upsilon^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}\Upsilon^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}\Upsilon^{-1}A_{12}A_{22}^{-1} \end{bmatrix}, \tag{17}$$

where $\Upsilon=A_{11}-A_{12}A_{22}^{-1}A_{21}$ is invertible as long as A is invertible. Applying formula (17) to (16), we find that, for any $x>\theta_1>\bar{w}$, $M^{(1)}(x)$ can be decomposed for large n with probability 1 as:

$$M^{(1)}(x)=\frac{1}{T}\left(\Delta+\varepsilon_{11}\right)'K_1^{-1}\left(\Delta+\varepsilon_{11}\right)-\frac{1}{T}\varepsilon_{11}'K_1^{-1}\varepsilon_{11}+K_2+\frac{1}{\sqrt{T}}\left(\Delta'K_3+K_3'\Delta\right),$$

where

$$\begin{aligned}
K_1 &= xI_k - \frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{12}' - \frac{1}{T^2}\varepsilon_{12}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-1}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{12}',\\
K_2 &= \frac{1}{T}\varepsilon_1'\left(x\mathcal{A}_0^{-2} - \frac{1}{T}\varepsilon_2\mathcal{B}^2\varepsilon_2'\right)^{-1}\varepsilon_1,\\
K_3 &= K_1^{-1}\frac{1}{T^{3/2}}\varepsilon_{12}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-1}\mathcal{A}\varepsilon_{21}.
\end{aligned}$$

First, we find the probability limit of K_1. By Lemma 4:

$$\left\|xI_k - \frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{12}' - (x-1)I_k\right\| \xrightarrow{p} 0. \tag{18}$$

Let us denote matrix $\frac{1}{T}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-1}\mathcal{A}\varepsilon_{22}\mathcal{B}^2$ as Z. Note that $n\left\|Z\right\|^2 \leq \frac{n}{T^2}\left\|\mathcal{A}\right\|^4\left\|Y^{-1}\right\|^2\left\|\varepsilon_{22}\right\|^4\left\|\mathcal{B}\right\|^8$ so that by Lemmas 1 and 5, $n\left\|Z\right\|^2 \xrightarrow{a.s.} 0$. Therefore, by Lemma 4:

$$\left\|\frac{1}{T}\varepsilon_{12}Z\varepsilon_{12}' - (\operatorname{tr} Z)\, I_k\right\| \xrightarrow{p} 0. \tag{19}$$

We will now focus on finding the probability limit of $\operatorname{tr} Z$. Note that for a general rank-one perturbation $M - vv'$ of matrix M, we have: $v'\left(M - vv'\right)^{-1}v = \frac{v'M^{-1}v}{1-v'M^{-1}v}$. Using this formula and the definition of Z, we obtain:

$$\begin{aligned}
\operatorname{tr} Z &= \frac{1}{T^2}\sum_{i=1}^{T-k} b_{k+i}^4\varepsilon_{22,i}'\mathcal{A}\left(Y_i - \frac{1}{T}b_{k+i}^2\mathcal{A}\varepsilon_{22,i}\varepsilon_{22,i}'\mathcal{A}\right)^{-1}\mathcal{A}\varepsilon_{22,i}\\
&= \frac{1}{T}\sum_{i=1}^{T-k} b_{k+i}^2\frac{b_{k+i}^2\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}}{1 - b_{k+i}^2\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}}.
\end{aligned}$$

But by Lemma 6, $\max_{1\leq i\leq T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i} - \frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}\right| \xrightarrow{p} 0$, whereas by Lemmas 7 and 8, $\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A} \xrightarrow{p} v_{2,x}^{-1}$. Therefore, $\max_{1\leq i\leq T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i} - v_{2,x}^{-1}\right| \xrightarrow{p} 0$. Further, since $v_{2,x} > u\left(\mathcal{F}^B\right) = \lim_{n\to\infty}\max_{i=1,\ldots,T} b_i^2$, the quantity $1 - b_{k+i}^2\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}$ is separated from zero with probability arbitrarily close to 1 for large enough n. Therefore, we have:

$$\max_{1\leq i\leq T-k}\left|\frac{b_{k+i}^4\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}}{1 - b_{k+i}^2\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}} - \frac{b_{k+i}^4 v_{2,x}^{-1}}{1 - b_{k+i}^2 v_{2,x}^{-1}}\right| \xrightarrow{p} 0$$

so that $\left|\operatorname{tr} Z - \frac{1}{T}\sum_{i=1}^{T-k}\frac{b_{k+i}^4 v_{2,x}^{-1}}{1-b_{k+i}^2 v_{2,x}^{-1}}\right| \xrightarrow{p} 0$. Finally, note that, by Assumption 2iii),

$$\left|\frac{1}{T}\sum_{i=1}^{T-k}\frac{b_{k+i}^4 v_{2,x}^{-1}}{1 - b_{k+i}^2 v_{2,x}^{-1}} - \frac{1}{T}\sum_{i=1}^{T}\frac{b_i^4 v_{2,x}^{-1}}{1 - b_i^2 v_{2,x}^{-1}}\right| = \frac{k}{T}\frac{v_{2,x}^{-1}}{1 - v_{2,x}^{-1}} \xrightarrow{p} 0 \tag{20}$$

and that, by Assumption 3i),

$$\frac{1}{T}\sum_{i=1}^{T}\frac{b_i^4 v_{2,x}^{-1}}{1-b_i^2 v_{2,x}^{-1}} = \int \frac{\tau^2}{v_{2,x}-\tau}dF^{B'B} \rightarrow \int \frac{\tau^2}{v_{2,x}-\tau}d\mathcal{F}^{B}\left(\tau\right). \tag{21}$$

Putting the latter three convergence statements together, we obtain:

$$\left|\operatorname{tr} Z - \int \frac{\tau^2}{v_{2,x}-\tau}d\mathcal{F}^{B}\left(\tau\right)\right| \xrightarrow{p} 0. \tag{22}$$

Combining (18), (19) and (22), we get: $\left\|K_1 - (x-1-\int \frac{\tau^2}{v_{2,x}-\tau}d\mathcal{F}^{B}\left(\tau\right))I_k\right\| \xrightarrow{p} 0$. But $\int \frac{\tau^2}{v_{2,x}-\tau}d\mathcal{F}^{B}\left(\tau\right) = -1 + \int \frac{\tau v_{2,x}}{v_{2,x}-\tau}d\mathcal{F}^{B}\left(\tau\right) = -1 + x u_{2,x}^{-1}$, where the last equality holds because $(u_{2,x}, v_{2,x})$ is a solution to (10). Therefore, finally,

$$\left\|K_1 - x(1-u_{2,x}^{-1})I_k\right\| \xrightarrow{p} 0. \tag{23}$$

Note that since $u_{2,x} > u\left(\mathcal{F}^{A}\right)$, $u_{2,x}$ is larger than 1. Hence, $x\left(1-u_{2,x}^{-1}\right)I_k$ is a positive definite matrix.

Now, let us find the probability limit of K_2. Note that $\left\|\frac{1}{T}\left(x\mathcal{A}_0^{-2} - \frac{1}{T}\varepsilon_2\mathcal{B}^2\varepsilon_2'\right)^{-1}\right\| \leq \frac{1}{T}\left\|\mathcal{A}_0\right\|^2\left(\theta_1-\bar{w}\right)^{-1}$ for large n with probability 1. Therefore, by Lemma 4: $\left\|K_2 - \frac{1}{T}\operatorname{tr}\left(x\mathcal{A}_0^{-2} - \frac{1}{T}\varepsilon_2\mathcal{B}^2\varepsilon_2'\right)^{-1}I_k\right\| \xrightarrow{p} 0$. On the other hand, by Lemmas 7 and 8, $\frac{1}{T}\operatorname{tr}\left(x\mathcal{A}_0^{-2} - \frac{1}{T}\varepsilon_2\mathcal{B}^2\varepsilon_2'\right)^{-1} \xrightarrow{p} v_{2,x}^{-1}$. Therefore,

$$\left\|K_2 - v_{2,x}^{-1}I_k\right\| \xrightarrow{p} 0. \tag{24}$$

Finally, let us find the probability limit of K_3. Denote the $(T-k)\times(n-k)$ matrix $\frac{1}{T^{3/2}}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-1}\mathcal{A}$ as G and let $\bar{G}$ be obtained from G by adding $\max\left\{T-n,0\right\}$ zero columns and $\max\left\{n-T,0\right\}$ zero rows. Similarly, let $\bar{\varepsilon}_{12}$ be obtained from ε_{12} by adding $\max\left\{n-T,0\right\}$ columns with i.i.d. entries distributed as ε_{it}, and let $\bar{\varepsilon}_{21}$ be obtained from ε_{21} by adding $\max\left\{T-n,0\right\}$ rows with i.i.d. entries distributed as ε_{it}. Assume that the elements added are independent from $\varepsilon_{12}, \varepsilon_{21}$ and from G. Then, we have: $\varepsilon_{12}G\varepsilon_{21} = \bar{\varepsilon}_{12}\bar{G}\bar{\varepsilon}_{21}$, where $\bar{G}$ is a square matrix with $\left\|\bar{G}\right\| \leq \frac{1}{T^{3/2}}\left\|\mathcal{B}\right\|^2\left\|\varepsilon_{22}\right\|\left\|\mathcal{A}\right\|^2\left(\theta_1-\bar{w}\right)^{-1}$ for large n with probability 1. Using Lemma 1 and Assumption 1i), we further get: $n\left\|\bar{G}\right\|^2 \xrightarrow{a.s.} 0$ so that, by Lemma 4: $\left\|\varepsilon_{12}G\varepsilon_{21}\right\| = \left\|\bar{\varepsilon}_{12}\bar{G}\bar{\varepsilon}_{21}\right\| \xrightarrow{p} 0$. Combining this finding with the fact that $K_3 = K_1^{-1}\varepsilon_{12}G\varepsilon_{21}$ and with (23), we obtain:

$$\left\|K_3\right\| \xrightarrow{p} 0. \tag{25}$$

The convergence facts (23), (24) and (25) established above together with the fact that, by Assumption 1iii), $\frac{1}{T}\Delta'\Delta \xrightarrow{p} \sigma^{-2}D$, imply that $M^{(1)}(x) \xrightarrow{p} x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1}\sigma^{-2}D + v_{2,x}^{-1}I_k$.□

Proof of Lemma 9 ii): For $M^{(2)}(x)$, using the square of the inverse of a partitioned matrix formula (17), we have:

$$\begin{aligned} M^{(2)}(x) &= \frac{1}{T}(\Delta+\varepsilon_{11})'\left(K_1^{-2}+K_1^{-1}K_4K_1^{-1}\right)(\Delta+\varepsilon_{11}) + \\ & K_5 - \frac{1}{T}\varepsilon_{11}'\left(K_1^{-2}+K_1^{-1}K_4K_1^{-1}\right)\varepsilon_{11} + \frac{1}{\sqrt{T}}\left(\Delta' K_6 + K_6'\Delta\right), \end{aligned}$$

where

$$\begin{aligned} K_4 &= \frac{1}{T^2}\varepsilon_{12}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-2}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{12}', \\ K_5 &= \frac{1}{T}\varepsilon_1'\mathcal{A}_0\left(xI_n - \frac{1}{T}\mathcal{A}_0\varepsilon_2\mathcal{B}^2\varepsilon_2'\mathcal{A}_0\right)^{-2}\mathcal{A}_0\varepsilon_1 \text{ and} \\ K_6 &= K_1^{-1}K_3(I_k+K_4) + K_1^{-1}\frac{1}{T^{3/2}}\varepsilon_{12}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-2}\mathcal{A}\varepsilon_{21}. \end{aligned}$$

Our analysis of K_4 term is similar to that of K_1 term. Let us define $\tilde{Z} \equiv \frac{1}{T^2}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}Y^{-2}\mathcal{A}\varepsilon_{22}\mathcal{B}^2$. Note that $n\left\|\tilde{Z}\right\|^2 \le \frac{n}{T^2}\|\mathcal{A}\|^4\left\|Y^{-1}\right\|^4\|\varepsilon_{22}\|^4\|\mathcal{B}\|^8$ so that by Lemmas 1 and 5, $n\left\|\tilde{Z}\right\|^2 \xrightarrow{a.s.} 0$. Therefore, by Lemma 4:

$$\left\|K_4 - \left(\operatorname{tr}\tilde{Z}\right)I_k\right\| \xrightarrow{p} 0. \tag{26}$$

For a general rank-one perturbation $M - vv'$ of matrix M, we have:

$$v'\left(M-vv'\right)^{-2}v = \frac{v'M^{-2}v}{\left(1-v'M^{-1}v\right)^2}. \tag{27}$$

Using this formula together with the definition of $\operatorname{tr}\tilde{Z}$, we obtain:

$$\begin{aligned} \operatorname{tr}\tilde{Z} &= \frac{1}{T^2}\sum\nolimits_{i=1}^{T-k} b_{k+i}^4\varepsilon_{22,i}'\mathcal{A}\left(Y_i - \frac{1}{T}b_{k+i}^2\mathcal{A}\varepsilon_{22,i}\varepsilon_{22,i}'\mathcal{A}\right)^{-2}\mathcal{A}\varepsilon_{22,i} \\ &= \frac{1}{T}\sum\nolimits_{i=1}^{T-k} b_{k+i}^2\frac{\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\mathcal{A}Y_i^{-2}\mathcal{A}\varepsilon_{22,i}}{\left(1-\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}\right)^2}. \end{aligned} \tag{28}$$

But by Lemma 6, $\max_{1\le i\le T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i} - \frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}\right| \xrightarrow{p} 0$ and $\max_{1\le i\le T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-2}\mathcal{A}\varepsilon_{22,i} - \frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-2}\mathcal{A}\right| \xrightarrow{p} 0$. Further, by Lemmas 7 and 8, $\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A} \xrightarrow{p}$

$v_{2,x}^{-1}$, whereas for $\operatorname{tr}\mathcal{A}Y^{-2}\mathcal{A}$, we have:

$$\begin{aligned}\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-2}\mathcal{A} &= \frac{1}{T}\operatorname{tr}Y^{-1}\mathcal{A}^2Y^{-1}\\ &= \frac{1}{T}x^{-1}\operatorname{tr}Y^{-1}\mathcal{A}^2Y^{-1}\left(xI_{n-k}-\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}+\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}\right)\\ &= \frac{1}{T}x^{-1}\operatorname{tr}Y^{-1}\mathcal{A}^2+\frac{1}{T}x^{-1}\operatorname{tr}Y^{-1}\mathcal{A}^2Y^{-1}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}\\ &= \frac{1}{T}x^{-1}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}+\frac{1}{T^2}x^{-1}\operatorname{tr}\mathcal{B}\varepsilon_{22}'\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\varepsilon_{22}\mathcal{B}. \qquad (29)\end{aligned}$$

For the first term in the latter sum, we have, by Lemmas 7 and 8:

$$\frac{1}{T}x^{-1}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}\xrightarrow{p}x^{-1}v_{2,x}^{-1}. \qquad (30)$$

For the second term, using (27), we obtain:

$$\begin{aligned}\frac{1}{T^2}x^{-1}\operatorname{tr}\mathcal{B}\varepsilon_{22}'\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\varepsilon_{22}\mathcal{B} &= \frac{1}{T^2}x^{-1}\operatorname{tr}\mathcal{B}\varepsilon_{22}'\left[x\mathcal{A}^{-2}-\frac{1}{T}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\right]^{-2}\varepsilon_{22}\mathcal{B}\\ = \frac{1}{T^2}x^{-1}\sum_{i=1}^{T-k}b_{k+i}^2\varepsilon_{22,i}'&\left[x\mathcal{A}^{-2}-\frac{1}{T}\varepsilon_{22,-i}\mathcal{B}_{-i}^2\varepsilon_{22,-i}'-\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}\varepsilon_{22,i}'\right]^{-2}\varepsilon_{22,i}\\ = \frac{1}{T}x^{-1}\sum_{i=1}^{T-k}&\frac{\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\left[x\mathcal{A}^{-2}-\frac{1}{T}\varepsilon_{22,-i}\mathcal{B}_{-i}^2\varepsilon_{22,-i}'\right]^{-2}\varepsilon_{22,i}}{\left(1-\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\left[x\mathcal{A}^{-2}-\frac{1}{T}\varepsilon_{22,-i}\mathcal{B}_{-i}^2\varepsilon_{22,-i}'\right]^{-1}\varepsilon_{22,i}\right)^2}\\ = \frac{1}{T}x^{-1}\sum_{i=1}^{T-k}&\frac{\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\left[\mathcal{A}Y_i^{-1}\mathcal{A}\right]^2\varepsilon_{22,i}}{\left(1-\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}\right)^2}.\end{aligned}$$

But by Lemma 6, $\max_{1\le i\le T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\left[\mathcal{A}Y_i^{-1}\mathcal{A}\right]^2\varepsilon_{22,i}-\frac{1}{T}\operatorname{tr}\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\right|\xrightarrow{p}0$ and $\max_{1\le i\le T-k}\left|\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}-\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}\right|\xrightarrow{p}0$. By Lemmas 7 and 8, $\frac{1}{T}\operatorname{tr}\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\xrightarrow{p}v_{2,x}^{-1}x^{-1}u_{2,x}r_{u,x}\left(1-r_{u,x}r_{v,x}\right)^{-1}$ and $\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A}\xrightarrow{p}v_{2,x}^{-1}$. Therefore, and since $1-b_{k+i}^2v_{2,x}^{-1}$ is separated from zero for large n, we have:

$$\max_{1\le i\le T-k}\left|\frac{\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\left[\mathcal{A}Y_i^{-1}\mathcal{A}\right]^2\varepsilon_{22,i}}{\left(1-\frac{1}{T}b_{k+i}^2\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}\right)}-\frac{b_{k+i}^2v_{2,x}^{-1}x^{-1}u_{2,x}r_{u,x}}{\left(1-b_{k+i}^2v_{2,x}^{-1}\right)^2\left(1-r_{u,x}r_{v,x}\right)}\right|\xrightarrow{p}0,$$

and using equations analogous to (20) and (21), we obtain:

$$\left|\frac{1}{T^2}x^{-1}\operatorname{tr}\mathcal{B}\varepsilon_{22}'\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\varepsilon_{22}\mathcal{B}-\frac{v_{2,x}^{-1}x^{-2}u_{2,x}r_{u,x}}{1-r_{u,x}r_{v,x}}\int\frac{\tau d\mathcal{F}^B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2}\right|\xrightarrow{p}0.$$

Combining the latter result with (29) and (30), we obtain:

$$\left|\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-2}\mathcal{A}-x^{-1}v_{2,x}^{-1}-\frac{v_{2,x}^{-1}x^{-2}u_{2,x}r_{u,x}}{1-r_{u,x}r_{v,x}}\int\frac{\tau d\mathcal{F}^B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2}\right|\xrightarrow{p}0.$$

The latter expression can be simplified. We have:

$$\begin{aligned}&x^{-1}v_{2,x}^{-1}+\frac{v_{2,x}^{-1}x^{-2}u_{2,x}r_{u,x}}{1-r_{u,x}r_{v,x}}\int\frac{\tau d\mathcal{F}^B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2}\\ =\;&x^{-1}v_{2,x}^{-1}\left(1-r_{u,x}r_{v,x}\right)^{-1}\left(1-r_{u,x}r_{v,x}+r_{u,x}u_{2,x}x^{-1}\int\frac{\tau v_{2,x}^2 d\mathcal{F}^B(\tau)}{\left(v_{2,x}-\tau\right)^2}\right)\end{aligned}\tag{31}$$

On the other hand,

$$u_{2,x}x^{-1}=\left(\int\frac{\tau v_{2,x}}{v_{2,x}-\tau}d\mathcal{F}^B(\tau)\right)^{-1}\tag{32}$$

because $(u_{2,x},v_{2,x})$ solve system (10) for $w=x$. Therefore,

$$\begin{aligned}u_{2,x}x^{-1}\int\frac{\tau v_{2,x}^2 d\mathcal{F}^B(\tau)}{\left(v_{2,x}-\tau\right)^2}&=\int\frac{\tau v_{2,x}d\mathcal{F}^B(\tau)}{\left(v_{2,x}-\tau\right)^2}\left(\int\frac{\tau d\mathcal{F}^B(\tau)}{v_{2,x}-\tau}\right)^{-1}\\ =\;\left(\int\frac{\tau^2 d\mathcal{F}^B(\tau)}{\left(v_{2,x}-\tau\right)^2}+\int\frac{\tau d\mathcal{F}^B(\tau)}{v_{2,x}-\tau}\right)&\left(\int\frac{\tau d\mathcal{F}^B(\tau)}{v_{2,x}-\tau}\right)^{-1}=1+r_{v,x}\end{aligned}$$

Substituting this result in (31), we obtain:

$$x^{-1}v_{2,x}^{-1}+\frac{v_{2,x}^{-1}x^{-2}u_{2,x}r_{u,x}}{1-r_{u,x}r_{v,x}}\int\frac{\tau d\mathcal{F}^B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2}=\frac{1+r_{u,x}}{xv_{2,x}\left(1-r_{u,x}r_{v,x}\right)}$$

and therefore,

$$\left|\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-2}\mathcal{A}-\frac{1+r_{u,x}}{xv_{2,x}\left(1-r_{u,x}r_{v,x}\right)}\right|\xrightarrow{p}0.$$

Returning to (28), using equations analogous to (20) and (21), we obtain:

$$\left| \operatorname{tr} \tilde{Z} - \frac{1 + r_{u,x}}{x v_{2,x} \left(1 - r_{u,x} r_{v,x}\right)} \int \frac{\tau^2 d\mathcal{F}^B(\tau)}{\left(1 - \tau v_{2,x}^{-1}\right)^2} \right| \xrightarrow{p} 0.$$

Note that $\int \frac{\tau^2 d\mathcal{F}^B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2} = v_{2,x} r_{v,x} \int \frac{\tau v_{2,x} d\mathcal{F}^B(\tau)}{v_{2,x} - \tau} = \frac{x v_{2,x} r_{v,x}}{u_{2,x}}$, where the last equality follows from (32). Therefore, finally,

$$\left| \operatorname{tr} \tilde{Z} - \frac{r_{v,x}\left(1 + r_{u,x}\right)}{u_{2,x}\left(1 - r_{u,x} r_{v,x}\right)} \right| \xrightarrow{p} 0.$$

Combining this fact with (26), we obtain:

$$\left\| K_4 - \frac{r_{v,x}\left(1 + r_{u,x}\right)}{u_{2,x}\left(1 - r_{u,x} r_{v,x}\right)} I_k \right\| \xrightarrow{p} 0. \tag{33}$$

For K_5, we have, by Lemma 4:

$$\left\| K_5 - I_k \frac{1}{T} \operatorname{tr} \mathcal{A}_0 \left(x I_n - \frac{1}{T} \mathcal{A}_0 \varepsilon_2 \mathcal{B}^2 \varepsilon_2' \mathcal{A}_0 \right)^{-2} \mathcal{A}_0 \right\| \xrightarrow{p} 0.$$

Replacing $n-k$, $\mathcal{A}$ and ε_{22} in the above analysis of $\frac{1}{T} \operatorname{tr} \mathcal{A} Y^{-2} \mathcal{A}$ by n, $\mathcal{A}_0$ and ε_2, respectively, we find that $\frac{1}{T} \operatorname{tr} \mathcal{A}_0 \left(x I_n - \frac{1}{T} \mathcal{A}_0 \varepsilon_2 \mathcal{B}^2 \varepsilon_2' \mathcal{A}_0 \right)^{-2} \mathcal{A}_0$ converges in probability to $\frac{1 + r_{u,x}}{x v_{2,x}\left(1 - r_{u,x} r_{v,x}\right)}$ so that

$$\left\| K_5 - \frac{1 + r_{u,x}}{x v_{2,x}\left(1 - r_{u,x} r_{v,x}\right)} I_k \right\| \xrightarrow{p} 0. \tag{34}$$

Finally, let us find the probability limit of $K_6 \equiv K_1^{-1} K_3 \left(I_k + K_4\right) + K_1^{-1} \frac{1}{T^{3/2}} \varepsilon_{12} \mathcal{B}^2 \varepsilon_{22}' \mathcal{A} Y^{-2} \mathcal{A} \varepsilon_{21}$. Since $\|K_3\| \xrightarrow{p} 0$, the first term in the latter sum converges in probability to zero. As to the second term, repeating the analysis that led us to (25), substituting Y^{-1} by Y^{-2} and $\left(\theta_1 - \bar{w}\right)^{-1}$ by $\left(\theta_1 - \bar{w}\right)^{-2}$, we conclude that it also converges in probability to zero. Hence,

$$\|K_6\| \xrightarrow{p} 0 \tag{35}$$

Finally, combining (23), (33), (34) and (35), we get:

$$M^{(2)}(x) \xrightarrow{p} \frac{D}{\sigma^2 x^2 \left(1 - u_{2,x}^{-1}\right)^2} \left(1 + \frac{r_{v,x}\left(1 + r_{u,x}\right)}{u_{2,x}\left(1 - r_{u,x} r_{v,x}\right)}\right) + \frac{1 + r_{u,x}}{x v_{2,x}\left(1 - r_{u,x} r_{v,x}\right)} I_k.$$

□

Proof of Lemma 9 iii): For $M^{(3)}(x)$, we have:

$$M^{(3)}(x) = \frac{1}{\sqrt{T}} K_1^{-1} (\Delta + \varepsilon_{11}) + K_3.$$

Therefore, using (23) and (25), we get $M^{(3)}(x) \xrightarrow{p} \sigma^{-1} x^{-1} \left(1 - u_{2,x}^{-1}\right)^{-1} D^{1/2}$.□

Proof of Lemma 9 iv): Note that

$$M_{\varkappa i}^{(1)}(x) = \begin{bmatrix} M^{(1)}(x) & \xi \\ \xi' & \varkappa e_i' K_1^{-1} e_i \end{bmatrix},$$

where $\xi' = \sqrt{\frac{\varkappa}{T}} e_i' K_1^{-1} (\Delta + \varepsilon_{11}) + \sqrt{\varkappa} e_i' K_3$. Such a representation for $M_{\varkappa i}^{(1)}(x)$ and the probability limits for K_1 and K_2 obtained in the proof of Lemma 9 i) imply part iv) of Lemma 9.□

1.4 Proof of Theorem 1 iii)

Let q be the integer defined in Theorem 1, that is q is the maximum non-negative integer such that $d_i/\sigma^2 > \bar{w}\left(1 - \bar{u}^{-1}\right)\left(1 - \bar{v}^{-1}\right)$. Since $x\left(1 - u_{2,x}^{-1}\right)\left(1 - v_{2,x}^{-1}\right)$ is a continuous strictly increasing function of $x \geq \bar{w}$, there exists a small enough $\theta_1 > \bar{w}$ such that $d_i/\sigma^2 > \theta_1\left(1 - u_{2,\theta_1}^{-1}\right)\left(1 - v_{2,\theta_1}^{-1}\right)$ for all $i \leq q$ and the inequality changes its sign for $i > q$. This fact is equivalent to the existence of a small enough $\theta_1 > \bar{w}$ such that

$$\theta_1^{-1}\left(1 - u_{2,\theta_1}^{-1}\right)^{-1} d_i \sigma^{-2} + v_{2,\theta_1}^{-1} > 1 \text{ for all } i \leq q \quad (36)$$

$$\theta_1^{-1}\left(1 - u_{2,\theta_1}^{-1}\right)^{-1} d_i \sigma^{-2} + v_{2,\theta_1}^{-1} < 1 \text{ for all } q < i \leq k. \quad (37)$$

Note that $x^{-1}\left(1 - u_{2,x}^{-1}\right)^{-1} d_i \sigma^{-2} + v_{2,x}^{-1}$ is the probability limit of the i-th largest eigenvalue of $M^{(1)}(x)$ as $n \to \infty$. Functions $g_i(x) \equiv x^{-1}\left(1 - u_{2,x}^{-1}\right)^{-1} d_i \sigma^{-2} + v_{2,x}^{-1}$, $i = 1, ..., k$ are strictly decreasing in $x \geq \theta_1$ and they tend to zero as $x \to \infty$. Taking into account these properties of the functions and inequalities (36-37), we conclude that equations $g_i(x) = 1$ have unique solutions $x = w_i$ for $i \leq q$ and $x > \theta_1$, and no solutions for $i > q$ and $x > \theta_1$. Note that $w_1 > w_2 > ... > w_q > \theta_1$.

Let δ be a small positive number such that $\delta < w_q - \theta_1$, let $\delta_1 = \min_{i=1,...,q; j=1,...,k} |g_j(w_i \pm \delta)|$ and $\delta_2 = \min\left\{|g_q(\theta_1)|, |g_{q+1}(\theta_1)|\right\}$. Further, let $\Phi_i^{\pm}$ denote the events that $\max_{j=1,...,k}\left|\lambda_j\left(M^{(1)}(w_i \pm \delta)\right) - g_j(w_i \pm \delta)\right| < \delta_1$ and let Φ be the event that $\max_{j=1,...,k}\left|\lambda_j\left(M^{(1)}(\theta_1)\right) - g_j(\theta_1)\right| < \delta_2$. By Lemma 9, the probability of each of these events can be made arbitrarily close to zero by choosing n large enough. Therefore, for any

$\delta_3 > 0$, for large enough n, $\Pr(\Omega) > 1-\delta_3$, where $\Omega = \cap_{i=1}^{q} \Phi_i^{\pm} \cap \Phi$. Hence, with probability arbitrarily close to 1, for large enough n, we have:

$$\begin{array}{c} \lambda_j \left(M^{(1)} \left(\theta_1 \right) \right) > 1 \text{ if and only if } j \leq q \text{ and} \\ \lambda_j \left(M^{(1)} \left(\theta_1 \right) \right) < 1 \text{ if and only if } j > q, \end{array} \tag{38}$$

and, for all $i = 1, ..., q$:

$$\begin{array}{c} \lambda_j \left(M^{(1)} \left(w_i - \delta \right) \right) > 1 \text{ if and only if } j \leq i \text{ and} \\ \lambda_j \left(M^{(1)} \left(w_i + \delta \right) \right) < 1 \text{ if and only if } j > i \end{array} . \tag{39}$$

Now, if $\lambda_1 \left(\tilde{\Lambda} \right) < \theta_1$ and $\tilde{\Psi}$ is full rank, then the eigenvalues $\lambda_j \left(M^1 (x) \right), j = 1, ..., k$ are strictly decreasing functions of x on $x \geq \theta_1$. It is because for any x_1 and x_2 such that $x_2 > x_1 \geq \theta_1$, matrix $M^1(x_1) - M^1(x_2) = \tilde{\Psi}' \left(\left(x_1 I_n - \tilde{\Lambda} \right)^{-1} - \left(x_2 I_n - \tilde{\Lambda} \right)^{-1} \right) \tilde{\Psi}$ is positive definite (this follows from the fact that the eigenvalues of $\left(x_1 I_n - \tilde{\Lambda} \right)^{-1} - \left(x_2 I_n - \tilde{\Lambda} \right)^{-1}$ equal $\frac{x_2 - x_1}{\left(x_1 - \lambda_j (\tilde{\Lambda}) \right) \left(x_2 - \lambda_j (\tilde{\Lambda}) \right)}$, $j = 1, ..., n$, and therefore, are positive). Note that, as was mentioned above, $\lambda_1 \left(\tilde{\Lambda} \right) \xrightarrow{a.s.} \bar{w}$ and, as is easily verified, $\tilde{\Psi}' \tilde{\Psi} \xrightarrow{p} \sigma^{-2} D + I_k$ so that with probability arbitrarily close to 1, $\lambda_1 \left(\tilde{\Lambda} \right) < \theta_1$ and $\tilde{\Psi}$ is full rank for large enough n. The strict monotonicity of $\lambda_j \left(M^1(x) \right), j = 1, ..., k$ and inequalities (38) and (39) imply that, with probability arbitrarily close to 1, for large enough n, there exists exactly q values of $x \geq \theta_1$ such that $M^{(1)}(x)$ has a (simple) eigenvalue, which equals 1. These q values of $x \geq \theta_1$ are in the δ-neighborhoods of $w_1, ..., w_q$. Since δ was an arbitrary positive number, we conclude, using Lemma 2, that $w_1, ..., w_q$ must be the probability limits of the first q eigenvalues of $\frac{1}{T} \tilde{X} \tilde{X}'$, which proves the first convergence statement of part iii) of Theorem 1.

Furthermore, as follows from above and from Lemma 2, for any $\theta_1 > \bar{w}$, there will be only q eigenvalues of $\frac{1}{T} \tilde{X} \tilde{X}'$ larger than θ_1 for large enough n. On the other hand, $\frac{1}{T} \tilde{X} \tilde{X}' = \tilde{\Psi} \tilde{\Psi}' + \tilde{\Lambda}$ so that the k-th eigenvalue of $\frac{1}{T} \tilde{X} \tilde{X}'$ cannot be smaller than the k-th eigenvalue of $\tilde{\Lambda}$, which converges to $u \left(\mathcal{F}^{c,A,B} \right) = \bar{w}$. Hence, the $q+1$-th,...,q-th eigenvalues of $\frac{1}{T} \tilde{X} \tilde{X}'$ converge to $\bar{w}$, which proves the second convergence statement of part iii) of Theorem 1.

1.5 Proof of Theorem 1 ii)

Now, let us turn to part ii) of Theorem 1. Let μ_j be the j-th largest eigenvalue of $\frac{1}{T} \tilde{X} \tilde{X}'$ with $j \leq q$. Then, since, as has been just shown, $\mu_j \xrightarrow{p} w_j$ and since the probability limit of $M^{(1)}(x)$, $x^{-1} \left(1 - u_{2,x}^{-1} \right)^{-1} \sigma^{-2} D + v_{2,x}^{-1} I_k$, is a continuous function of $x \geq \theta_1$, we have:

$M^{(1)}(\mu_j) \xrightarrow{p} w_j^{-1}\left(1-u_{2,w_j}^{-1}\right)^{-1}\sigma^{-2}D+v_{2,w_j}^{-1}I_k$. Further, since the latter probability limit is a diagonal matrix with strictly decreasing entries on the diagonal, the j-th principal eigenprojection of $M^{(1)}(\mu_j)$ converges in probability to the projection on the subspace spanned by the vector $e_j \equiv (0,...,0,1,0,...,0)'$, where 1 is the j-th coordinate. In other words, we can choose the eigenvectors v_j corresponding to the unit eigenvalue of $M^{(1)}(\mu_j)$ so that they converge in probability to e_j as $n \to \infty$.

Further, by Lemma 9 and since $u_{2,w_j} = u_j$ and $v_{2,w_j} = v_j$, we have: the j,j-th element of $M^{(2)}(\mu_j)$ converges in probability to $\sigma^{-2}d_j w_j^{-2}\left(1-u_j^{-1}\right)^{-2}\left(1+\frac{r_{vj}(1+r_{uj})}{(1-r_{uj}r_{vj})u_j}\right)+\frac{1+r_{uj}}{(1-r_{uj}r_{vj})w_j v_j}$ and the j-th column of $M^{(3)}\left(\mu_j\right)$ converges in probability to $e_j\sigma^{-1}w_j^{-1}\left(1-u_j^{-1}\right)^{-1}d_j^{1/2}$. Therefore, by Lemma 2 iii, the j-th column of R, where $j \leq q$, is proportional to e_j with the coefficient of proportionality P_{jj}, where

$$\begin{aligned}
P_{jj}^{-2} &= \left(\sigma^{-2}d_j w_j^{-2}\left(1-u_j^{-1}\right)^{-2}\left(1+\frac{r_{vj}\left(1+r_{uj}\right)}{\left(1-r_{uj}r_{vj}\right)u_j}\right)+\frac{1+r_{uj}}{\left(1-r_{uj}r_{vj}\right)w_j v_j}\right)\times \\
&\quad \left(\sigma^{-2}w_j^{-2}\left(1-u_j^{-1}\right)^{-2}d_j\right)^{-1} \\
&= 1+\frac{r_{vj}\left(1+r_{uj}\right)}{\left(1-r_{uj}r_{vj}\right)u_j}+\frac{\left(1+r_{uj}\right)\sigma^2 w_j\left(1-u_j^{-1}\right)^2 d_j^{-1}}{\left(1-r_{uj}r_{vj}\right)v_j} \\
&= 1+\frac{r_{vj}\left(1+r_{uj}\right)}{\left(1-r_{uj}r_{vj}\right)u_j}+\frac{\left(1+r_{uj}\right)\left(u_j-1\right)}{\left(1-r_{uj}r_{vj}\right)\left(v_j-1\right)u_j},
\end{aligned}$$

where the last equality follows from the fact that $\sigma^{-2}d_j = w_j\left(1-u_j^{-1}\right)\left(1-v_j^{-1}\right)$. Such a convergence of the j-th column of R is the first convergence statement of part ii) of Theorem 1.

Now, let us focus on the second convergence statement of part ii) of Theorem 1. We no longer assume that (1) is satisfied (this assumption was innocuous for the proof of the first convergence statement as have been explained above). Suppose that $j > q$. Let $y_1, y_2, ..., y_k$ be the unit-length eigenvectors of $\frac{1}{T}\tilde{X}\tilde{X}'$ corresponding to the k of the largest eigenvalues. Let e_i be an $n \times 1$ vector with all entries but the i-th one equal zero and the i-th entry equal to 1. Define $Q_j = \sum_{r=1}^{q} y_r\left(\frac{1}{T}\tilde{X}\tilde{X}'+\varkappa e_i e_i'\right)y_r + y_j'\left(\frac{1}{T}\tilde{X}\tilde{X}'+\varkappa e_i e_i'\right)y_j$, where $\varkappa$ is an arbitrary positive number. Since $y_1, ..., y_k$ are orthonormal,

$$Q_j \leq \sum_{r=1}^{q+1}\lambda_r\left(\frac{1}{T}\tilde{X}\tilde{X}'+\varkappa e_i e_i'\right). \tag{40}$$

Lemmas 2 iv) and 9 iv) imply that the probability limits of the first k eigenvalues of $\frac{1}{T}\tilde{X}\tilde{X}'+\varkappa e_i e_i'$ are as follows. If $i > q$ and $\varkappa_1$ is so small that, for any $0 < \varkappa < \varkappa_1$, the largest eigenvalue of $\bar{w}^{-1}\left(1-\bar{u}^{-1}\right)^{-1}\sigma^{-2}\begin{pmatrix} d_i & \sigma\sqrt{\varkappa d_i}e_i \\ \sigma\sqrt{\varkappa d_i}e_i' & \sigma^2\varkappa \end{pmatrix} - \begin{pmatrix} \bar{v}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$ is less than 1, then eigenvalues $\lambda_r\left(\frac{1}{T}\tilde{X}\tilde{X}' + \varkappa e_i e_i'\right)$ converge to w_r for $r \leq q$ and to $\bar{w}$ for $q < r \leq k$. If $i \leq q$ and $\varkappa_2$ is so small that, for any $0 < \varkappa < \varkappa_2$, the smallest eigenvalue of $\bar{w}^{-1}\left(1-\bar{u}^{-1}\right)^{-1}\sigma^{-2}\begin{pmatrix} d_i & \sigma\sqrt{\varkappa d_i}e_i \\ \sigma\sqrt{\varkappa d_i}e_i' & \sigma^2\varkappa \end{pmatrix} - \begin{pmatrix} \bar{v}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$ is less than 1, then eigenvalues $\lambda_r\left(\frac{1}{T}\tilde{X}\tilde{X}' + \varkappa e_i e_i'\right)$ converge to w_r for $r \leq q$ and $r \neq i$, and to $\bar{w}$ for $q < r \leq k$. For the i-th eigenvalue of $\frac{1}{T}\tilde{X}\tilde{X}' + \varkappa e_i e_i'$, by the formula for the approximation of an eigenvalue of a perturbed matrix (formula 3.6 on p.89 of Kato, 1995), we have: for any $\delta > 0$, there exists $C > 0, N > 0$ and $\varkappa_0 > 0$ such that $\left|\lambda_i\left(\frac{1}{T}\tilde{X}\tilde{X}' + \varkappa e_i e_i'\right) - \lambda_i\left(\frac{1}{T}\tilde{X}\tilde{X}'\right) - \varkappa y_{ii}^2\right| < C\varkappa^2$, for all $n > N$ and $\varkappa < \varkappa_0$ with probability no smaller than $1 - \delta$.

For $i \leq q$, the convergence of $\lambda_r\left(\frac{1}{T}\tilde{X}\tilde{X}' + \varkappa e_i e_i'\right)$ described above, together with (40) imply that:

$$Q_j \leq \sum\nolimits_{r=1}^{q} w_r + \varkappa y_{ii}^2 + \bar{w} + \delta + C\varkappa^2 \tag{41}$$

with probability no smaller than $1 - \delta$ for large enough n. On the other hand, by definition of Q_j :

$$Q_j \geq \sum\nolimits_{r=1}^{q} w_r + \varkappa y_{ii}^2 + \bar{w} + \varkappa y_{ji}^2 - \delta \tag{42}$$

with probability no smaller than $1-\delta$ for large enough n. Combining the latter two displayed inequalities, we have: $\varkappa y_{ji}^2 \leq 2\delta + C\varkappa^2$ with probability no smaller than $1 - 2\delta$ for large enough n. Let us take $\delta = \varkappa^2$. Then, we have: $y_{ji}^2 \leq (2 + C)\,\varkappa$ with probability no smaller than $1 - 2\varkappa^2$ for large enough n. Since $\varkappa$ is an arbitrary positive number, smaller than $\min(\varkappa_0, \varkappa_2)$, we have: $y_{ji}^2 \xrightarrow{p} 0$.

For $i > q$, since the probability limit of $\lambda_i\left(\frac{1}{T}\tilde{X}\tilde{X}' + \varkappa e_i e_i'\right)$ is w_i, we replace inequalities (41) and (42) by $Q \leq \sum_{r=1}^{q} w_r + \bar{w} + \delta$ and $Q \geq \sum_{r=1}^{q} w_r + \bar{w} + \varkappa y_{ji}^2 - \delta$, respectively. So, we have: $y_{ji}^2 \leq 2\varkappa$ with probability no smaller than $1 - 2\varkappa^2$ for large enough n. Since $\varkappa$ is an arbitrary positive number, smaller than $\min(\varkappa_0, \varkappa_1)$, we again have: $y_{ji}^2 \xrightarrow{p} 0$. Note that $y_{ji}^2 = P_{ij}^2$. Hence, for $j > q$ and any $i = 1, ..., k$, $P_{ij}^2 \xrightarrow{p} 0$. This completes the proof.$\square$

2 Proof of Theorem 2

We will prove parts ii) and iii) of the theorem. A proof of part i) is similar to the proof of part ii) and we omit it to save space. We will use notation introduced in the proof of Theorem 1. However, we will not normalize ε so that the variance of its entries equals 1. For any matrix M, we will denote its j-th row as $M_{j\cdot}$ and its j-th column as $M_{\cdot j}$. Further, we will use $M_{r:s}$ to denote the matrix that consists of the columns $r, r+1, ..., s$ of matrix M, and we will use $M_{i:j,r:s}$ to denote the matrix that consists of the intersection of the rows

$i, i+1, ..., j$ and columns $r, r+1, ..., s$ of matrix M.

2.1 A key lemma

Let $\frac{1}{T}\varepsilon_2\varepsilon_2' = O'\Lambda O$ be the spectral decomposition of $\frac{1}{T}\varepsilon_2\varepsilon_2'$. Note that, since by Assumption 2a i), $\varepsilon_2\varepsilon_2'$ is distributed according to Wishart $W\left(\sigma^2 I_n, T-k\right)$, its spectral decomposition can be chosen so that O has the Haar invariant distribution (see Anderson (1984)).[3] Define $\hat{X} = \sigma O \tilde{X}$ and $\Psi = \sigma O \tilde{\Psi} = O_{1:k}\left(L'L\right)^{1/2}\left(\frac{F'F}{T}\right)^{1/2} + \frac{1}{\sqrt{T}}O\varepsilon_1$. Then, matrix $\frac{1}{T}\hat{X}\hat{X}'$ has a convenient representation $\frac{1}{T}\hat{X}\hat{X}' = \Psi\Psi' + \Lambda$ and the same eigenvalues as matrix $\frac{1}{T}XX'$.

Let y_{ij} denote the i-th component of an eigenvector of $\frac{1}{T}\hat{X}\hat{X}'$, corresponding to eigenvalue $\lambda_j\left(\frac{1}{T}XX'\right)$, and λ_i denote the i-th largest diagonal element of Λ. Let us define

$$\begin{aligned} M_n^{(1)}(x) &\equiv \sum_{i=1}^{n} \frac{\Psi_{i\cdot}'\Psi_{i\cdot}}{x-\lambda_i}, \\ M_n^{(2)}(x) &\equiv \sum_{i=1}^{n} \frac{\Psi_{i\cdot}'\Psi_{i\cdot}}{\left(x-\lambda_i\right)^2}, \text{ and} \\ M_n^{(3)}(x) &\equiv \sum_{i=1}^{n} \frac{O_{i,1:k}'\Psi_{i\cdot}}{x-\lambda_i} \end{aligned}$$

The following Lemma is a straightforward consequence of Lemma 2:

Lemma 10: *Let $\mu \neq \lambda_i$, $i = 1, ..., n$ so that $\mu I_n - \Lambda$ is invertible. Then:*

i) *μ is an eigenvalue of $\frac{1}{T}\hat{X}\hat{X}'$ of multiplicity larger than or equal to s if and only if there exists a positive integer $m \leq k+1-s$ such that $x = \mu$ satisfies equations*

$$\lambda_m\left(M_n^{(1)}(x)\right) = 1, ..., \lambda_{m+s-1}\left(M_n^{(1)}(x)\right) = 1, \tag{43}$$

ii) *If v is an eigenvector of $M_n^{(1)}(\mu)$ corresponding to eigenvalue 1, then*

$$y = \left(v'M_n^{(2)}(\mu)v\right)^{-1/2}\left(\mu I_n - \Lambda\right)^{-1}\Psi v \tag{44}$$

is a unit-length eigenvector of $\frac{1}{T}\hat{X}\hat{X}'$ corresponding to eigenvalue μ.

iii) *If 1 is a simple eigenvalue of $M_n^{(1)}(\mu)$, then μ is a simple eigenvalue of $\frac{1}{T}\hat{X}\hat{X}'$. Furthermore, if μ* is *the j-th largest eigenvalue of $\frac{1}{T}\hat{X}\hat{X}'$* and *$v$ is a corresponding*

---

[3] The decomposition is not unique because each of the columns of O can be multiplied by -1 and the last $\max(0, n-T+k)$ columns can be arbitrarily rotated.

eigenvector of $M_n^{(1)}(\mu)$, *then the* j*-th column of matrix* $\tilde{P}$ *from part ii) of Theorem 2 equals* $\left(v' M_n^{(2)}(\mu) v\right)^{-1/2} M_n^{(3)}(\mu) v$.

The key fact for the analysis below was established by Marčenko and Pastur (1967). They showed that the empirical distribution of the elements along the diagonal of Λ defined as $\mathcal{F}^\Lambda \equiv \frac{\#\{\lambda_i \leq \lambda\}}{n}$ almost surely converges to a non-random cumulative distribution function $\mathcal{F}_c$, which has density

$$\begin{aligned} f_c(\lambda) &= \begin{cases} \frac{1}{2\pi\lambda c\sigma^2}\sqrt{(b-\lambda)(\lambda-a)} & \text{if } a \leq \lambda \leq b \\ 0 & \text{otherwise} \end{cases} \\ a &= \left(1-\sqrt{c}\right)^2\sigma^2, \quad b = \left(1+\sqrt{c}\right)^2\sigma^2, \end{aligned} \tag{45}$$

and a point mass $1 - 1/c$ at $\lambda = 0$ if $c > 1$.

To see the significance of the Marčenko-Pastur result for our analysis, assume for a moment that $k = 1$ and note that $M_n^{(1)}(x)$ is a weighted linear combination of terms Ψ_i^2 with weights $(x - \lambda_i)^{-1}$. Now, by definition, $\Psi_i = O_{i,1}(L'L)^{1/2}\left(\frac{F'F}{T}\right)^{1/2} + \frac{1}{\sqrt{T}}O_{i\cdot}\varepsilon_1$. The second element in this sum is independent of the first and, by Assumption 2a, is $N\left(0, \sigma^2/T\right)$. The first term is asymptotically $N(0, d_1/n)$. Indeed, since O has the Haar invariant distribution, the joint distribution of the entries of its first column is the same as that of the entries of $\xi/\|\xi\|$, where $\xi \sim N(0, I_n)$ and $\|\xi\| = \sqrt{\xi'\xi}$. Hence, $M_n^{(1)}(x)$ asymptotically behaves as a weighted sum of $\chi^2(1)$ independent random variables with weights $\frac{1}{n}\left(d_1 + c\sigma^2\right)(x - \lambda_i)^{-1}$. Intuitively, such a sum should converge to $\left(d_1 + c\sigma^2\right)\int (x-\lambda)^{-1} d\mathcal{F}_c(\lambda)$, which we confirm below. The properties of $M_n^{(1)}(x)$ centered by its probability limit and scaled by $\sqrt{n}$ can be analyzed using similar ideas.

2.2 Technical lemmata

Lemma 11: *(McLeish (1974)) Let* $\{X_{n,i}, \mathcal{F}_{n,i}; i = 1, 2, ..., n\}$ *be a martingale difference array on the probability triple* $(\Omega, \mathcal{F}, P)$. *If the following conditions are satisfied: a) Lindeberg's condition: for all* $\varepsilon > 0, \sum_i \int_{|X_{n,i}|>\varepsilon} X_{n,i}^2 dP \to 0, n \to \infty$; *b)* $\sum_{i=1}^n X_{n,i}^2 \xrightarrow{p} 1$, *then* $\sum_{i=1}^n X_{n,i} \xrightarrow{d} N(0,1)$.

Proof of Lemma 11: This is a consequence of Theorem (2.3) of McLeish (1974). Two conditions of the theorem, i) $\max_{i\leq n}|X_{n,i}|$ is uniformly bounded in L_2 norm, and ii) $\max_{i\leq n}|X_{n,i}| \xrightarrow{p} 0$, are replaced here by the Lindeberg condition. As explained in McLeish (1974), since for any ε, $\max_{i\leq n} X_{n,i}^2 \leq \varepsilon^2 + \sum_i X_{n,i}^2 I(|X_{n,i}| > \varepsilon)$ and since $P\{\max_{i\leq n}|X_{n,i}| > \varepsilon\} = P\left\{\sum_i X_{n,i}^2 I(|X_{n,i}| > \varepsilon) > \varepsilon^2\right\}$, both conditions i) and ii) follow from the Lindeberg condition.□

Lemma 12: *(Hall and Heyde) Let* $\{X_{n,i}, \mathcal{F}_{ni}; 1 \leq i \leq n\}$ *be a martingale difference* *array and define* $V_{n,j}^2 = \sum_{i=1}^{j} E\left(X_{n,i}^2|\mathcal{F}_{n,i-1}\right)$ and $U_{n,j}^2 = \sum_{i=1}^{j} X_{n,i}^2$ for $1 \leq j \leq n$. *Suppose that the conditional variances* $V_{n,n}^2$ *are tight, that is* $\sup_n P\left(V_{n,n}^2 > \varepsilon\right) \to 0$ *as* $\varepsilon \to \infty$, *and that the conditional Lindeberg condition holds, that is for all* $\varepsilon > 0$, $\sum_i E\left[X_{n,i}^2 I\left(|X_{n,i}| > \varepsilon\right)|\mathcal{F}_{n,i-1}\right] \xrightarrow{p} 0$. *Then* $\max_j \left|U_{n,j}^2 - V_{n,j}^2\right| \xrightarrow{p} 0$.

Proof of Lemma 12: This is a shortened version of Theorem 2.23 in Hall and Heyde (1980).□

Let $g_j(\lambda)$, $j = 1, ..., J$, be analytic functions of real variable λ on an open interval $\left(\bar{a}, \bar{b}\right)$ containing the support of the Marčenko-Pastur distribution, that is the set $\{0, [a, b]\}$ if $c > 1$, and the segment $[a, b]$ if $c \geq 1$. Further, let $\varsigma^{(n)}$ be an array of $n \times m$ matrices with i.i.d. standard normal entries independent of $\lambda_1, ..., \lambda_n$. In what follows we will omit the superscript n in $\varsigma^{(n)}$ to simplify notations. Finally, denote the set of triples $\{(j, s, t) : 1 \leq j \leq J, 1 \leq s \leq t \leq m\}$ as Θ_1. Then, we have the following

Lemma 13: *Let Assumptions 1a and 2a hold. Then, the joint distribution of random variables* $\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_j(\lambda_i)\left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right); (j, s, t) \in \Theta_1\right\}$ *weakly converges to a multivariate normal distribution as* $n \to \infty$. *The covariance between components* (j, s, t) *and* (j_1, s_1, t_1) *of the limiting distribution is equal to* 0 *when* $(s, t) \neq (s_1, t_1)$, *and to* $(1 + \delta_{st}) \int g_j(\lambda) g_{j_1}(\lambda) d\mathcal{F}_c(\lambda)$ *when* $(s, t) = (s_1, t_1)$.

Proof of Lemma 13: Let real numbers a_1 and b_1 be such that $[a_1, b_1]$ is included in $\left(\bar{a}, \bar{b}\right)$, but itself includes the support of the Marčenko-Pastur law. Define functions $h_j(\lambda), j = 1, ..., J$, so that $h_j(\lambda) = g_j(\lambda)$ for $\lambda \in [a_1, b_1]$, and $h_j(\lambda) = 0$ otherwise. Note that $|h_j(\lambda)| < B$ for any $j = 1, ..., J$ and any λ, where B is a constant larger than $\max_{j=1,...,J} \sup_{\lambda \in [a_1, b_1]} |g_j(\lambda)|$. Note also that since, by Lemma 1, λ_1 almost surely converges to b,
$P\{\exists j \leq J, i \leq n \text{ such that } h_j(\lambda_i) \neq g_j(\lambda_i)\} \to 0$ as $n \to \infty$.

Consider random variables $X_{n,i} = \frac{1}{\sqrt{n}}\sum_{(j,s,t)\in\Theta_1} \gamma_{jst} h_j(\lambda_i)\left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right)$, where γ_{jst} are some constants. Let $\mathcal{F}_{n,i}$ be sigma-algebra generated by $\lambda_1, ..., \lambda_n$ and $\varsigma_{js}; 1 \leq j \leq i, 1 \leq s \leq m$. Clearly, $\{X_{n,i}, \mathcal{F}_{n,i}; i = 1, 2, ..., n\}$ form a martingale difference array. Let K be the number of different triples $(j, s, t) \in \Theta_1$. Consider an arbitrary order in Θ_1. In Hölder's inequality $\sum_{r=1}^{K} a_r b_r \leq \left(\sum_{r=1}^{K} (a_r)^p\right)^{1/p}\left(\sum_{r=1}^{K} (b_r)^q\right)^{1/q}$, which holds for $a_r > 0, b_r > 0, p > 1, q > 1$, and $(1/p) + (1/q) = 1$, take $a_r = \left|\frac{1}{\sqrt{n}}\gamma_{jst} h_j(\lambda_i)\left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right)\right|$, where (j, s, t) is the r-th triple in Θ_1, $b_r = 1$, and $p = 2 + \delta$ for some $\delta > 0$. Then, the inequality implies that $|X_{n,i}|^{2+\delta} \leq K^{1+\delta} B^{2+\delta} \sum_{(j,s,t)\in\Theta_1}\left|\gamma_{jst}\frac{\varsigma_{is}\varsigma_{it} - \delta_{st}}{\sqrt{n}}\right|^{2+\delta}$. Recalling that ς_{is} are i.i.d. standard normal random variables, we have: $\sum_i E|X_{n,i}|^{2+\delta}$ tends to zero as $n \to \infty$, which means that the Lyapunov condition holds for $X_{n,i}$. As is well known, Lyapunov's condition implies Lindeberg's condition. Hence, condition a) of McLeish's proposition is satisfied for $X_{n,i}$.

Now, let us consider $\sum_{i=1}^{n} X_{n,i}^2$. Since convergence in mean implies convergence in probability, the conditional Lindeberg condition is satisfied for $X_{n,i}$ because the unconditional Lindeberg condition is satisfied as checked above. Further, in notations of Hall and Heyde's proposition, we have
$V_{n,n}^2 = \frac{1}{n}\sum_{i=1}^{n} E(\sum_{\substack{(j,s,t)\in\Theta_1,\\ (j_1,s_1,t_1)\in\Theta_1}} \gamma_{jst}\gamma_{j_1 s_1 t_1} h_j(\lambda_i) h_{j_1}(\lambda_i) \left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right)\left(\varsigma_{is_1}\varsigma_{it_1} - \delta_{s_1 t_1}\right)|\mathcal{F}_{n,i-1})$.
It is straightforward to check that the latter expression is equal to $\sum_{\substack{1\leq j\leq J\\ 1\leq j_1\leq J}} \left[\left(\sum_{1\leq s\leq t\leq m} \gamma_{jst}\gamma_{j_1 st}\left(1+\delta_{st}\right)\right)\frac{1}{n}\sum_{i=1}^{n} h_j(\lambda_i) h_{j_1}(\lambda_i)\right]$.

Consider now $\tilde{V}_{n,n}^2 = \sum_{\substack{1\leq j\leq J\\ 1\leq j_1\leq J}} \left[\left(\sum_{1\leq s\leq t\leq m} \gamma_{jst}\gamma_{j_1 st}\left(1+\delta_{st}\right)\right)\frac{1}{n}\sum_{i=1}^{n} g_j(\lambda_i) g_{j_1}(\lambda_i)\right]$. Since $P\left(\tilde{V}_{n,n}^2 \neq V_{n,n}^2\right) \to 0$ as $n\to\infty$, $\tilde{V}_{n,n}^2$ and $V_{n,n}^2$ must converge in probability to the same limit, or must both diverge. But, by Theorem 1.1 of Bai and Silverstein (2004), $\frac{1}{n}\sum_{i=1}^{n} g_j(\lambda_i) g_{j_1}(\lambda_i) - \int g_j(\lambda) g_{j_1}(\lambda) d\mathcal{F}_{\frac{n}{T}}(\lambda)$ converges in probability to zero. Therefore, since $\mathcal{F}_{\frac{n}{T}}(\lambda)$ weakly converge to $\mathcal{F}_c(\lambda)$ as $n\to\infty$, we have

$$\tilde{V}_{n,n}^2 \xrightarrow{p} \Sigma \equiv \sum_{\substack{1\leq j\leq J\\ 1\leq j_1\leq J}} \left[\left(\sum_{1\leq s\leq t\leq m} \gamma_{jst}\gamma_{j_1 st}\left(1+\delta_{st}\right)\right)\int g_j(\lambda) g_{j_1}(\lambda) d\mathcal{F}_c(\lambda)\right]. \qquad (46)$$

Hence, $V_{n,n}^2$ also converges in probability to Σ. In particular, $V_{n,n}^2$ is tight and Hall and Heyde's proposition applies. From Hall and Heyde's proposition, we know that $\sum_{i=1}^{n} X_{n,i}^2$ must converge to the same limit as $V_{n,n}^2$. Therefore, using McLeish's result, we get $\sum_{i=1}^{n} X_{n,i} \xrightarrow{d} N(0,\Sigma)$.

Let us now define $Y_{n,i} = \sum_{(j,s,t)\in\Theta_1} \gamma_{jst} g_j(\lambda_i)\frac{\varsigma_{is}\varsigma_{it}-\delta_{st}}{\sqrt{n}}$. Since $P\left(\sum_{i=1}^{n} Y_{n,i} \neq \sum_{i=1}^{n} X_{n,i}\right) \to 0$ as $n\to\infty$, we have $\sum_{i=1}^{n} Y_{n,i} \xrightarrow{d} N(0,\Sigma)$. Finally, Lemma 13 follows from the latter convergence, the Cramer-Wold result (see White (1999), p.114), and definition of Σ (46).□

Now let us formally establish the asymptotic behavior of $M_n^{(1)}(x)$, $M_n^{(2)}(x)$ and $M_n^{(3)}(x)$. By Lemma1, for any fixed k, $\lambda_1, ..., \lambda_k$ almost surely converge to b, defined in (45). This result implies that, with high probability, $M_n^{(1)}(x)$ belongs to the space $C\left[\theta_1,\theta_2\right]^{k^2}$ of continuous $k\times k$-matrix-valued functions on $x\in\left[\theta_1,\theta_2\right]$, where $\theta_2 > \theta_1 > b$. Since the weak convergence in $C\left[\theta_1,\theta_2\right]$ is well-studied, it will be convenient to modify $M_n^{(1)}(x)$ on a small probability set so that the modification is a random element of $C\left[\theta_1,\theta_2\right]^{k^2}$ equipped with the max sup norm. To construct such a modification, define $h(x, \lambda_i) = \max\left(x - \lambda_i, \frac{\theta_1 - b}{2}\right)$ and let

$$\begin{aligned}\hat{M}_n^{(1)}(x) &\equiv \sum_{i=1}^{n} \frac{\Psi_{i\cdot}' \Psi_{i\cdot}}{h(x, \lambda_i)}, \\ \hat{M}_n^{(2)}(x) &= \sum_{i=1}^{n} \frac{\Psi_{i\cdot}' \Psi_{i\cdot}}{h^2(x, \lambda_i)} \text{ and} \\ \hat{M}_n^{(3)}(x) &= \sum_{i=1}^{n} \frac{O_{i,1:k}' \Psi_{i\cdot}}{h(x, \lambda_i)}.\end{aligned}$$

We will study the asymptotic properties of $\hat{M}_n^{(j)}(x)$ keeping in mind that they are equivalent to the asymptotic properties of $M_n^{(j)}(x)$ because $\Pr\left(M_n^{(j)}(x) = \hat{M}_n^{(j)}(x), \forall x \in [\theta_1, \theta_2]\right) = \Pr\left(\lambda_1 < \frac{\theta_1 + b}{2}\right) \to 1$ as $n \to \infty$.

Let us define

$$\begin{aligned}M_0^{(1)}(x) &= \left(D + \sigma^2 c I_k\right) \int \frac{d\mathcal{F}_c(\lambda)}{x - \lambda}, \\ M_0^{(2)}(x) &= \left(D + \sigma^2 c I_k\right) \int \frac{d\mathcal{F}_c(\lambda)}{(x - \lambda)^2} \text{ and} \\ M_0^{(3)}(x) &= D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x - \lambda}.\end{aligned}$$

We have the following

Lemma 14: *Let Assumptions 1a and 2a hold. Then, for the random elements of* $C^{k^2}[\theta_1, \theta_2]$ *defined as* $N_n^{(p)}(x) = \sqrt{n}\left(\hat{M}_n^{(p)}(x) - M_0^{(p)}(x)\right), p = 1, 2, 3,$ *we have:*

$$\left\{N_n^{(p)}(x), p = 1, 2, 3\right\} \xrightarrow{d} \left\{N^{(p)}(x), p = 1, 2, 3\right\}, \tag{47}$$

where, for any $\{x_1, ..., x_J\} \in [\theta_1, \theta_2]$, *the joint distribution of entries of* $\left\{N^{(p)}(x_j); p = 1, 2, 3, j = 1, ..., J\right\}$ *is a* $3Jk^2$*-dimensional normal distribution with covariance between entry in row* s *and column* t *of* $N^{(p)}(x_j)$ *and entry in row* s_1 *and column* t_1 *of* $N^{(r)}(x_{j_1})$ *equal to* $\Omega^{(p,r)}(\tau, \tau_1)$, *where* $\tau = (s, t, j)$ *and* $\tau_1 = (s_1, t_1, j_1)$, *and* $\Omega^{(p,r)}(\tau, \tau_1)$ *is defined as follows:*

For $\tau=(s,t,j)$, $\tau_1=(s_1,t_1,j_1)$, *and integers* p_1 *and* p_2 *such that* $1\le p_1\le p_2\le 2$,

$$\begin{aligned}\Omega^{(p_1,p_2)}(\tau,\tau_1) &= \frac{c}{4}(d_s+d_t)(d_{s_1}+d_{t_1})\phi_{sts_1t_1}\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)^{p_1}}\int\frac{d\mathcal{F}_c(\lambda)}{(x_{j_1}-\lambda)^{p_2}}\\ \Omega^{(p_1,3)}(\tau,\tau_1) &= \frac{c}{4}(d_s+d_t)\sqrt{d_{s_1}}\phi_{sts_1t_1}\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)^{p_1}}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda}\\ \Omega^{(3,3)}(\tau,\tau_1) &= \frac{c}{4}\sqrt{d_sd_{s_1}}\phi_{sts_1t_1}\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda}\end{aligned}$$

if $(s_1,t_1)\neq(s,t)$ *and* $(s_1,t_1)\neq(t,s)$;

$$\begin{aligned}\Omega^{(p_1,p_2)}(\tau,\tau_1) &= \left[\frac{c}{4}(d_s+d_t)^2\phi_{stst}-(1+\delta_{st})d_sd_t\right]\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)^{p_1}}\int\frac{d\mathcal{F}_c(\lambda)}{(x_{j_1}-\lambda)^{p_2}}\\ &\quad+\left[(1+\delta_{st})\left(\sigma^4c^2+d_sd_t\right)+\sigma^2c\left(d_s+d_t+2\delta_{st}\sqrt{d_sd_t}\right)\right]\\ &\quad\cdot\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)^{p_1}(x_{j_1}-\lambda)^{p_2}}\\ \Omega^{(p_1,3)}(\tau,\tau_1) &= \left[\frac{c}{4}(d_s+d_t)\sqrt{d_s}\phi_{stst}-(1+\delta_{st})\sqrt{d_s}d_t\right]\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)^{p_1}}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda}\\ &\quad+\left[(1+\delta_{st})\sqrt{d_s}d_t+\sigma^2c\left(\sqrt{d_s}+\delta_{st}\sqrt{d_t}\right)\right]\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)^{p_1}(x_{j_1}-\lambda)}\\ \Omega^{(3,3)}(\tau,\tau_1) &= \left(\frac{c}{4}d_s\phi_{stst}-(1+\delta_{st})d_t\right)\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda}\\ &\quad+\left((1+\delta_{st})d_t+\sigma^2c\right)\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)(x_{j_1}-\lambda)}\end{aligned}$$

if $(s_1,t_1)=(s,t)$; *and*

$$\begin{aligned}\Omega^{(p_1,p_2)}(\tau,\tau_1) &= \Omega^{(p_1,p_2)}((t,s,j),(s_1,t_1,j_1))\\ \Omega^{(p_1,3)}(\tau,\tau_1) &= \Omega^{(p_1,3)}((t,s,j),(s_1,t_1,j_1))\\ \Omega^{(3,3)}(\tau,\tau_1) &= \left(\frac{c}{4}\phi_{stst}-(1+\delta_{st})\right)\sqrt{d_sd_t}\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1}-\lambda}\\ &\quad+\left((1+\delta_{st})\sqrt{d_sd_t}+\delta_{st}\sigma^2c\right)\int\frac{d\mathcal{F}_c(\lambda)}{(x_j-\lambda)(x_{j_1}-\lambda)}\end{aligned}$$

if $(s_1,t_1)=(t,s)$.

Proof of Lemma 14): To save space, we will only study the convergence of $N_n^{(1)}(x)$. The joint convergence of $\left\{N_n^{(p)}(x);p=1,2,3\right\}$ can be demonstrated using similar ideas. We will prove the convergence of $N_n^{(1)}(x)$ by first checking the convergence of the finite dimensional distributions $\left\{N_{n,st}^{(1)}(x_j),(s,t,j)\in\Theta\right\}\xrightarrow{d}\left\{N_{st}^{(1)}(x_j),(s,t,j)\in\Theta\right\}$, where Θ denotes the set of all integer triples (s,t,j) satisfying $1\le s,t\le k$ and $1\le j\le J$, and, second, by demonstrating the tightness of all entries of $N_n^{(1)}(x)$.

Note that the distribution of $N_n^{(1)}(x)$ will not change if we substitute $O_{1:k}$ and $O\varepsilon_1$ in the definition of Ψ by $\xi(\xi'\xi)^{-1/2}$ and $\sigma\eta$, where ξ and η are two independent $n \times k$ matrix with i.i.d. standard normal entries independent from η, F, and $\lambda_1, ..., \lambda_n$. Indeed, the substitution of $O\varepsilon$ by $\sigma\eta$ is justified by Assumption 2a. As to the other substitution, note that the columns of $\xi(\xi'\xi)^{-1/2}$ are orthogonal and of unit length. Further, the joint distribution of elements of $\xi(\xi'\xi)^{-1/2}$ is invariant with respect to multiplication from the left by any orthogonal matrix. Hence, this distribution coincides with the joint distribution of the elements of the first k columns of random orthogonal matrix having Haar invariant distribution. But the latter is the joint distribution of elements of $O_{1:k}$. In the rest of the proof, we, therefore, will make the substitutions and redefine $N_n^{(1)}(x)$ accordingly.

It is straightforward to check that $N_n^{(1)}(x) = \sum_{v=1}^{10} S^{(v)}(x)$, where

$$S^{(1)}(x) = \left(\tfrac{F'F}{T}\right)^{1/2} (L'L)^{1/2} \left(\tfrac{\xi'\xi}{n}\right)^{-1/2} \left(\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{\xi_{i\cdot}'\xi_{i\cdot} - I_k}{h(x,\lambda_i)}\right) \left(\tfrac{\xi'\xi}{n}\right)^{-1/2} (L'L)^{1/2} \left(\tfrac{F'F}{T}\right)^{1/2},$$

$$S^{(2)}(x) = \left(\tfrac{F'F}{T}\right)^{1/2} (L'L)^{1/2} \left(\tfrac{\xi'\xi}{n}\right)^{-1} (L'L)^{1/2} \sqrt{\tfrac{n}{T}}\sqrt{T} \left(\left(\tfrac{F'F}{T}\right)^{1/2} - I_k\right) \sum_{i=1}^{n} \tfrac{1}{nh(x,\lambda_i)},$$

$$S^{(3)}(x) = \sqrt{\tfrac{n}{T}}\sqrt{T} \left(\left(\tfrac{F'F}{T}\right)^{1/2} - I_k\right) (L'L)^{1/2} \left(\tfrac{\xi'\xi}{n}\right)^{-1} (L'L)^{1/2} \sum_{i=1}^{n} \tfrac{1}{nh(x,\lambda_i)},$$

$$S^{(4)}(x) = (L'L)^{1/2} \sqrt{n} \left(I_k - \left(\tfrac{\xi'\xi}{n}\right)\right) \left(\tfrac{\xi'\xi}{n}\right)^{-1} (L'L)^{1/2} \sum_{i=1}^{n} \tfrac{1}{nh(x,\lambda_i)},$$

$$S^{(5)}(x) = \sqrt{n}\,(L'L - D) \sum_{i=1}^{n} \tfrac{1}{nh(x,\lambda_i)},$$

$$S^{(6)}(x) = \sigma\sqrt{\tfrac{n}{T}} \left(\tfrac{F'F}{T}\right)^{1/2} (L'L)^{1/2} \left(\tfrac{\xi'\xi}{n}\right)^{-1/2} \left(\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{\xi_{i\cdot}'\eta_{i\cdot}}{h(x,\lambda_i)}\right),$$

$$S^{(7)}(x) = \sigma\sqrt{\tfrac{n}{T}} \left(\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{\eta_{i\cdot}'\xi_{i\cdot}}{h(x,\lambda_i)}\right) \left(\tfrac{\xi'\xi}{n}\right)^{-1/2} (L'L)^{1/2} \left(\tfrac{F'F}{T}\right)^{1/2},$$

$$S^{(8)}(x) = \sigma^2 \left(\tfrac{n}{T}\right) \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{\eta_{i\cdot}'\eta_{i\cdot} - I_k}{h(x,\lambda_i)},$$

$$S^{(9)}(x) = \sigma^2\sqrt{n} \left(\tfrac{n}{T} - c\right) I_k \sum_{i=1}^{n} \tfrac{1}{nh(x,\lambda_i)},$$

$$S^{(10)}(x) = -\left(D + \sigma^2 c I_k\right) \sqrt{n} \left(\int \tfrac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \tfrac{1}{nh(x,\lambda_i)}\right)$$

By Theorem 1 of Bai and Silverstein (2004), $\sqrt{n}\left(\int \frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{n}\frac{1}{x-\lambda_i}\right) \xrightarrow{p} 0$ for any $x \in [\theta_1, \theta_2]$. Our assumption that $n/T - c = o\left(1/\sqrt{n}\right)$ and the definition of Marcenko-Pastur law imply that $\sqrt{n}\left(\int \frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda} - \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}\right) \xrightarrow{p} 0$, and hence $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right) \xrightarrow{p} 0$. The latter convergence result together with the facts that $F'F/T \xrightarrow{p} I_k$, $\xi'\xi/n \xrightarrow{p} I_k$, $L'L - D = o_p\left(n^{-1/2}\right)$, and $n/T - c = o\left(n^{-1/2}\right)$ imply that $\left\{\sum_{v=1}^{10} S_{st}^{(v)}(x_j); (s,t,j) \in \Theta\right\}$ and $\left\{\sum_{v=1}^{10} \tilde{S}_{st}^{(v)}(x_j); (s,t,j) \in \Theta\right\}$ weakly converge to the same limit or do not converge together, where

$$\tilde{S}^{(1)}(x) = D^{1/2} \left(\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \tfrac{\xi_{i\cdot}'\xi_{i\cdot} - I_k}{h(x,\lambda_i)}\right) D^{1/2},$$

$$\tilde{S}^{(2)}(x) = D\sqrt{c}\sqrt{T} \left(\left(\tfrac{F'F}{T}\right)^{1/2} - I_k\right) \int \tfrac{d\mathcal{F}_c(\lambda)}{x-\lambda},$$

$\tilde{S}^{(3)}(x) = \sqrt{c}\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2} - I_k\right) D \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda},$
$\tilde{S}^{(4)}(x) = D^{1/2}\sqrt{n}\left(I_k - \left(\frac{\xi'\xi}{n}\right)\right) D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda},$
$\tilde{S}^{(5)}(x) = 0,$
$\tilde{S}^{(6)}(x) = \sigma\sqrt{c}D^{1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi'_{i\cdot}\eta_{i\cdot}}{h(x,\lambda_i)}\right),$
$\tilde{S}^{(7)}(x) = \sigma\sqrt{c}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta'_{i\cdot}\xi_{i\cdot}}{h(x,\lambda_i)}\right) D^{1/2},$
$\tilde{S}^{(8)}(x) = \sigma^2 c\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta'_{i\cdot}\eta_{i\cdot} - I_k}{h(x,\lambda_i)},$
$\tilde{S}^{(9)}(x) = \tilde{S}^{(10)}(x) = 0.$

Let us, first, consider the limit of $\left\{\tilde{S}^{(2)}_{st}(x_j) + \tilde{S}^{(3)}_{st}(x_j), (s,t,j) \in \Theta\right\}$. Since $\left(\frac{F'F}{T}\right)^{1/2} = I + \frac{1}{2}\left(\frac{F'F}{T} - I\right) + o_p\left(\frac{1}{\sqrt{T}}\right)$, using Assumption 1a, we get $\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2} - I_k\right) \xrightarrow{d} \frac{1}{2}\Phi$. The latter convergence and the definition of $\tilde{S}^{(2)}(x)$, $\tilde{S}^{(3)}(x)$, and Φ imply that $\left\{\tilde{S}^{(2)}(x_j) + \tilde{S}^{(3)}(x_j), 1 \le j \le J\right\} \xrightarrow{d} \left\{\frac{\sqrt{c}}{2}\left(D\Phi + \Phi D\right)\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}, 1 \le j \le J\right\}$, and, hence, $\left\{\tilde{S}^{(2)}_{st}(x_j) + \tilde{S}^{(3)}_{st}(x_j), (s,t,j) \in \Theta\right\}$ weakly converge to $\left\{Z^{(1)}_{stj}, (s,t,j) \in \Theta\right\}$ having joint zero-mean Gaussian distribution such that

$$\operatorname{cov}\left(Z^{(1)}_{stj}, Z^{(1)}_{s_1t_1j_1}\right) = \frac{c}{4}\left(d_s + d_t\right)\left(d_{s_1} + d_{t_1}\right)\phi_{sts_1t_1}\int\frac{d\mathcal{F}_c(\lambda)}{x_j - \lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda}. \tag{48}$$

Now, let us consider the limit of $\left\{\sum_{v\neq 2,3}\tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\}$. By definition, we have: $\sum_{v\neq 2,3}\tilde{S}^{(v)}_{st}(x_j) = \sqrt{d_sd_t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it} - \delta_{st}}{h(x_j,\lambda_i)} - \sqrt{d_sd_t}\int\frac{d\mathcal{F}_c(\lambda)}{x_j-\lambda}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it} - \delta_{st}}{\sqrt{n}} + \sigma\sqrt{cd_s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\eta_{it}}{h(x_j,\lambda_i)} + \sigma\sqrt{cd_t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{it}\eta_{is}}{h(x_j,\lambda_i)} + \sigma^2 c\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{is}\eta_{it} - \delta_{st}}{h(x_j,\lambda_i)}$. Since $[\xi, \eta]$ is an $n \times 2k$ matrix with i.i.d. standard normal entries, Lemma 13 and the above decomposition imply that
$\left\{\sum_{v\neq 2,3}\tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\}$ weakly converge to $\left\{Z^{(2)}_{stj}, (s,t,j) \in \Theta\right\}$ having joint normal distribution such that $\operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right) = 0$ if $(s,t) \neq (s_1,t_1)$ and $(s,t) \neq (t_1,s_1)$ and $\operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right)$ is equal to

$$\begin{aligned}\operatorname{cov}\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right) &= \left[\left(1 + \delta_{st}\right)\left(\sigma^4c^2 + d_sd_t\right) + \sigma^2c\left(d_s + d_t + 2\delta_{st}\sqrt{d_sd_t}\right)\right] \times \\ &\quad \times \int\frac{d\mathcal{F}_c(\lambda)}{\left(x_j - \lambda\right)\left(x_{j_1} - \lambda\right)} - \left(1 + \delta_{st}\right)d_sd_t\int\frac{d\mathcal{F}_c(\lambda)}{x_j - \lambda}\int\frac{d\mathcal{F}_c(\lambda)}{x_{j_1} - \lambda}\end{aligned} \tag{49}$$

otherwise.

Finally, since $\left\{\tilde{S}^{(2)}_{st}(x_j) + \tilde{S}^{(3)}_{st}(x_j), (s,t,j) \in \Theta\right\}$ are, by definition, independent from $\left\{\sum_{v\neq 2,3}\tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\}$, $\left\{Z^{(1)}_{stj}, (s,t,j) \in \Theta\right\}$ must be independent from $\left\{Z^{(2)}_{stj}, (s,t,j) \in \Theta\right\}$ and $\left\{\sum_{v=1}^{10}\tilde{S}^{(v)}_{st}(x_j); (s,t,j) \in \Theta\right\} \xrightarrow{d} \left\{Z^{(1)}_{stj} + Z^{(2)}_{stj}; (s,t,j) \in \Theta\right\}$, having joint zero-mean Gaussian distribution such that cov $\left(Z^{(1)}_{stj}+Z^{(2)}_{stj}, Z^{(1)}_{s_1t_1j_1}+Z^{(2)}_{s_1t_1j_1}\right) =$ cov $\left(Z^{(1)}_{stj}, Z^{(1)}_{s_1t_1j_1}\right) +$ cov $\left(Z^{(2)}_{stj}, Z^{(2)}_{s_1t_1j_1}\right)$. (48) and (49) imply that the joint distribution of $Z^{(1)}_{stj}+Z^{(2)}_{stj}$ is equal to that of $\left\{N^{(1)}_{st}(x_j);(s,t,j)\in\Theta\right\}$.

Now we have to prove the tightness of all entries of $N^{(1)}_n(x) = \sum_{v=1}^{10} S^{(v)}(x)$. Since product and sum are continuous mappings from $C\left[\theta_1,\theta_2\right]^2$ to $C\left[\theta_1,\theta_2\right]$, it is enough to prove the tightness of every entry of each matrix entering definition of $S^{(v)}(x)$, $v=1,...,10$. Assumption 1a and the facts that $F'F/T \xrightarrow{p} I_k$, $\xi'\xi/n \xrightarrow{p} I_k$, $L'L-D = o\left(n^{-1/2}\right)$, and $n/T - c = o\left(n^{-1/2}\right)$ imply the tightness of every entry of each of the matrices $\left(\frac{F'F}{T}\right)^{1/2}$, $\left(L'L\right)^{1/2}$, $\sqrt{n}\left(L'L-D\right)$, $\left(\frac{\xi'\xi}{n}\right)^{-1/2}$, $\left(\frac{\xi'\xi}{n}\right)^{-1}$, $\sqrt{\frac{n}{T}}I$, $\sqrt{n}\left(\frac{n}{T}-c\right)I$, $\sqrt{T}\left(\left(\frac{F'F}{T}\right)^{1/2}-I_k\right)$, and $\sqrt{n}\left(I_k-\left(\frac{\xi'\xi}{n}\right)\right)$ considered as (constant) elements of $C\left[\theta_1,\theta_2\right]$. Therefore, we only need to prove the tightness of entries of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it}-\delta_{st}}{h(x,\lambda_i)},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\eta_{it}}{h(x,\lambda_i)},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{is}\eta_{it}-\delta_{st}}{h(x,\lambda_i)} \tag{50}$$

of $\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)}$ and of $\sqrt{n}\left(\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)}\right)$.

Since ξ and η are, by definition, two independent $n\times k$ matrices with i.i.d. standard normal entries, to prove the tightness of the sequences of sums in (50), it is enough to prove the tightness of the first sum for all $1\leq s\leq t\leq k$. We will use Theorem 12.3 of Billingsley (1968), p. 95. Condition i) of the theorem is equivalent in our context to the assumption of the tightness of the sum at $x=\theta_1$. Lemma 5 implies that this assumption is satisfied. We will verify condition ii) of Theorem 12.3 by proving the moment condition (12.51) of Billingsley (1968). We have $\frac{E\left(\sum_{i=1}^{n}\left(h(x_1,\lambda_i)^{-1}-h(x_2,\lambda_i)^{-1}\right)\left(\xi_{is}\xi_{it}-\delta_{st}\right)\right)^2}{n(x_1-x_2)^2} \leq E\left(\sum_{i=1}^{n}\left(h(x_1,\lambda_i)h(x_2,\lambda_i)\right)^{-1}\left(\xi_{is}\xi_{it}-\delta_{st}\right)\right)^2/n \leq \frac{16}{n(\theta_1-b)^4}E\left(\sum_{i=1}^{n}\left(\xi_{is}\xi_{it}-\delta_{st}\right)\right)^2 = \frac{16}{(\theta_1-b)^4}\left(1+\delta_{st}\right)$, where the first inequality follows from the fact that $\left|\frac{1}{h(x_1,\lambda_i)}-\frac{1}{h(x_2,\lambda_i)}\right| \leq \frac{|x_2-x_1|}{h(x_1,\lambda_i)h(x_2,\lambda_i)}$. Hence, $\sup_{n;x_1,x_2\in[\theta_1,\theta_2]} E\left(\sum_{i=1}^{n}\left(h(x_1,\lambda_i)^{-1}-h(x_2,\lambda_i)^{-1}\right)\left(\xi_{is}\xi_{it}-\delta_{st}\right)\right)^2/n\left(x_1-x_2\right)^2$ is finite and the moment condition (12.51) of Billingsley (1968) is satisfied. In a more complete proof (in which the tightness of the elements of $N^{(2)}_n(x)$ is demonstrated), we also need to check Billingsley's moment condition when $h\left(\cdot,\cdot\right)$ is replaced by $h^2\left(\cdot,\cdot\right)$. We can use the above reasoning and inequality $\left|\frac{1}{h^2(x_1,\lambda_i)}-\frac{1}{h^2(x_2,\lambda_i)}\right| \leq \frac{|x_2-x_1|\left(h(x_1,\lambda_i)+h(x_2,\lambda_i)\right)}{h^2(x_1,\lambda_i)h^2(x_2,\lambda_i)} \leq \frac{32\theta_2|x_2-x_1|}{(\theta_1-b)^4}$ to perform such a check.

Similarly, conditions of Theorem 12.3 of Billingsley (1968) are satisfied for $\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)}$. Condition i) is satisfied because, as has been shown above, $\sqrt{n}\left(\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\lambda_i)}\right)\xrightarrow{p}$

0 for any $x \in [\theta_1, \theta_2]$. Condition ii) is satisfied because $E\left(\sum_{i=1}^{n} \frac{1}{nh(x_1,\lambda_i)h(x_2,\lambda_i)}\right)^2 \leq \frac{16}{(\theta_1-b)^4}$.

To prove the tightness of $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right)$, we adopt the argument on page 563 of Bai and Silverstein (2004). In notations of Bai and Silverstein (2004), $\hat{M}_n(\cdot) \to -\frac{1}{2\pi i}\int \frac{1}{x-z}\hat{M}_n(z)dz$ is a continuous mapping of $C\left(\mathcal{C}, R^2\right)$ into $C[\theta_1, \theta_2]$. Since, $\hat{M}_n(\cdot)$ is tight, $-\frac{1}{2\pi i}\int \frac{1}{x-z}\hat{M}_n(z)dz$, and subsequently $n\left(\int \frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{n}\frac{1}{x-\lambda_i}\right)$, form a tight sequence. But $\sup_{x\in[\theta_1,\theta_2]} \sqrt{n}\left(\int \frac{d\mathcal{F}_{n/T}(\lambda)}{x-\lambda} - \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}\right) \xrightarrow{p} 0$ because, by assumption, $n/T - c = o\left(1/\sqrt{n}\right)$. Therefore, $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{n}\frac{1}{x-\lambda_i}\right)$ is tight too. Finally, the latter tightness and the fact that $P\left\{\sum_{i=1}^{n} \frac{1}{\sqrt{n}}\left(\frac{1}{x-\lambda_i} - \frac{1}{h(x,\lambda_i)}\right) \neq 0\right\} \to 0$ imply that sequence $\sqrt{n}\left(\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} - \sum_{i=1}^{n} \frac{1}{nh(x,\lambda_i)}\right)$ must be tight.□

Lemma 15: *Let* $A(\varkappa) = A + \varkappa A^{(1)}$, *where* $A^{(1)}$ *is a symmetric* $k \times k$ *matrix and* $A = \operatorname{diag}(a_1, a_2, ..., a_k)$, $a_1 > a_2 > ... > a_k > 0$. *Further, let* $r_0 = \frac{1}{2}\min_{j=1,...,k}|a_j - a_{j+1}|$, *where we define* a_{k+1} *as zero. Then, for any real* $\varkappa$ *such that* $|\varkappa| < r_0/\left\|A^{(1)}\right\|$, *the following two statements hold:*
i) Exactly one eigenvalue of $A(\varkappa)$ *belongs to the segment* $(a_j - r_0, a_j + r_0)$. *Denoting this eigenvalue as* $a_j(\varkappa)$, *we have:*[4] $\left|\frac{1}{\varkappa}(a_j(\varkappa) - a_j) - A^{(1)}_{jj}\right| \leq |\varkappa|\left\|A^{(1)}\right\|\left(r_0 - |\varkappa|\left\|A^{(1)}\right\|\right)^{-1}$.
ii) Let $P_j(\varkappa)$ *be the orthogonal projection on the invariant subspace of* $A(\varkappa)$ *corresponding to eigenvalue* $a_j(\varkappa)$ *and let*
$S_j = \operatorname{diag}\left((a_1 - a_j)^{-1}, ..., (a_{j-1} - a_j)^{-1}, 0, (a_{j+1} - a_j)^{-1}, ..., (a_k - a_j)^{-1}\right)$. *Then* $e_j(\varkappa) \equiv P_j(\varkappa)e_j/\|P_j(\varkappa)e_j\|$ *is an eigenvector of* $A(\varkappa)$ *corresponding to eigenvalue* $a_j(\varkappa)$, *and* $\left\|\frac{1}{\varkappa}(e_j(\varkappa) - e_j) + S_jA^{(1)}e_j\right\| \leq 2|\varkappa|\left\|A^{(1)}\right\|^2\left(r_0 - |\varkappa|\left\|A^{(1)}\right\|\right)^{-2}$.

Proof of Lemma 15: Let $R(z,\varkappa) = (A(\varkappa) - zI_k)^{-1}$ be the resolvent of $A(\varkappa)$ defined for all complex z not equal to any of the eigenvalues of $A(\varkappa)$. We will denote $R(z,0)$ as $R(z)$. Let Γ be a positively oriented circle in the complex plane with center at a_j and radius r_0. The second Neumann series for the resolvent $R(z,\varkappa) = R(z) + \sum_{n=1}^{\infty}(-\varkappa)^n R(z)\left(A^{(1)}R(z)\right)^n$ (see Kato (1980), p.67, for a definition of the second Neumann series) is uniformly convergent on Γ for $\varkappa < \min_{z\in\Gamma}\left(\left\|A^{(1)}\right\|\|R(z)\|\right)^{-1} = r_0/\left\|A^{(1)}\right\|$, where the last equality follows from the fact that $\|R(z)\| = r_0^{-1}$ for any $z \in \Gamma$. Therefore, formula (1.19) of Kato (1980) implies that, for $|\varkappa| < r_0/\left\|A^{(1)}\right\|$, there is exactly one eigenvalue, $a_j(\varkappa)$, inside the circle Γ. Formulae (3.6)[5] and (2.32) of Kato (1980) imply the inequality stated in part i of Lemma 3.

We now turn to the proof of part ii. According to Kato (1980), p.67, projection $P_j(\varkappa)$ can be represented as $P_j(\varkappa) = -\frac{1}{2\pi i}\int_\Gamma R(z,\varkappa)\,dz$. Substituting the second Neumann series

---

[4] For any matrix (or vector) B, $\|B\| = (\max eig(B^*B))^{1/2}$, where $*$ denotes the operation of transposition and complex conjugation.

[5] Note the difference in notations. Kato's r_0 is ours $r_0/\left\|A^{(1)}\right\|$.

for the resolvent in this formula, we obtain

$$P_j(\varkappa) = P_j - \frac{1}{2\pi i} \sum_{n=1}^{\infty} (-\varkappa)^n \int_\Gamma R(z) \left(A^{(1)} R(z)\right)^n dz \qquad (51)$$

where $P_j \equiv P_j(0)$ and the series absolutely converges for $|\varkappa| < \frac{r_0}{\|A^{(1)}\|}$. Kato (1980), page 76, shows that $\frac{1}{2\pi i}\int_\Gamma R(z) A^{(1)} R(z)\, dz = -P_j A^{(1)} S_j - S_j A^{(1)} P_j$. This equality and (51) imply that $P_j(\varkappa) = P_j - \varkappa\left(P_j A^{(1)} S_j - S_j A^{(1)} P_j\right) - \frac{1}{2\pi i}\sum_{n=2}^{\infty} (-\varkappa)^n \int_\Gamma R(z) \left(A^{(1)} R(z)\right)^n dz$. Therefore, we have:

$$\left\| \frac{1}{\varkappa}\left(P_j(\varkappa) - P_j\right) + P_j A^{(1)} S_j + S_j A^{(1)} P_j \right\| \leq \frac{|\varkappa| \left\|A^{(1)}\right\|^2}{r_0\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)} \qquad (52)$$

for any $|\varkappa| < r_0 / \left\|A^{(1)}\right\|$.

Since A is diagonal with decreasing elements along the diagonal, e_j is an eigenvector of A corresponding to the eigenvalue a_j. By definition of $P_j(\varkappa)$, $e_j(\varkappa) \equiv \frac{P_j(\varkappa) e_j}{\|P_j(\varkappa) e_j\|}$ must be an eigenvector of $A(\varkappa)$ corresponding to the eigenvalue $a_j(\varkappa)$. Consider an identity $\frac{1}{\varkappa}\left(e_j(\varkappa) - e_j\right) + S_j A^{(1)} e_j = \left(\frac{1}{\varkappa}\left(P_j(\varkappa) e_j - e_j\right) + S_j A^{(1)} e_j\right) + \frac{1}{\varkappa} e_j(\varkappa)\left(1 - \|P_j(\varkappa) e_j\|\right)$. Using (52) and the fact that $S_j e_j = 0$, for the first term on right hand side of the identity we have:

$$\left\| \frac{1}{\varkappa}\left(P_j(\varkappa) e_j - e_j\right) + S_j A^{(1)} e_j \right\| \leq \frac{|\varkappa| \left\|A^{(1)}\right\|^2}{r_0\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)}. \qquad (53)$$

Using the fact that $P_j(\varkappa)$ is a projection operator so that $\|P_j(\varkappa) e_j\| \leq 1$ and $P_j(\varkappa)^2 = P_j(\varkappa)$, for the second term on right hand side of the identity we have:

$$\left\| \frac{1}{\varkappa} e_j(\varkappa)\left(1 - \|P_j(\varkappa) e_j\|\right) \right\| \leq \frac{1}{|\varkappa|}\left(1 - \|P_j(\varkappa) e_j\|^2\right) = |\varkappa| \left\| \frac{1}{\varkappa}\left(P_j(\varkappa) e_j - e_j\right) \right\|^2. \qquad (54)$$

But, form (53), $\left\|\frac{1}{\varkappa}\left(P_j(\varkappa) e_j - e_j\right)\right\|^2 \leq 2\left\|S_j A^{(1)} e_j\right\|^2 + \frac{2|\varkappa|^2 \left\|A^{(1)}\right\|^4}{r_0^2\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^2} \leq \frac{\left\|A^{(1)}\right\|^2}{2r_0^2} + \frac{2|\varkappa|^2 \left\|A^{(1)}\right\|^4}{r_0^2\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^2}$. Combining the above identity, (53), (54), and the latter inequality, we obtain: $\left\|\frac{1}{\varkappa}\left(e_j(\varkappa) - e_j\right) + S_j A^{(1)} e_j\right\| \leq \frac{|\varkappa| \left\|A^{(1)}\right\|^2 \left(3r_0^2 - 4r_0|\varkappa| \left\|A^{(1)}\right\| + 5|\varkappa|^2 \left\|A^{(1)}\right\|^2\right)}{2r_0^2\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^2} \leq \frac{2|\varkappa| \left\|A^{(1)}\right\|^2}{\left(r_0 - |\varkappa| \left\|A^{(1)}\right\|\right)^2}$, where the last inequality follows from the fact that $r_0 > |\varkappa| \left\|A^{(1)}\right\|$. This proves statement ii) of the lemma.□

Lemma 16: *Let $f_n(x)$ and $f_0(x)$ be random elements of $C[\theta_1, \theta_2]$ such that $f_n(x) \xrightarrow{d} f_0(x)$ as $n \to \infty$. And let x_n be random variables with values form $[\theta_1, \theta_2]$ and such that $x_n \xrightarrow{p} x_0$, where $x_0 \in [\theta_1, \theta_2]$ Then $f_n(x_n) - f_n(x_0) \xrightarrow{p} 0$.*

Proof of Lemma 16: Since $f_n(x) \xrightarrow{d} f_0(x)$, $\{f_n(x)\}$ is tight and, hence, for any $\varepsilon >$

0, we can choose a compact K such that $\Pr\left(f_n(x) \in K\right) > 1 - \frac{\varepsilon}{2}$ for all n. By the Arzelà-Ascoli theorem (see, for example, Billingsley (1999), p.81), for any positive ε_1, we have $K \subset \{f : |f(\theta_1)| \leq r\}$ for large enough r and $K \subset \{f : w_f(\delta(\varepsilon_1)) \leq \varepsilon_1\}$ for small enough $\delta(\varepsilon_1)$, where $w_f(\delta)$ is the modulus of continuity of function f, defined as $w_f(\delta) = \sup_{|s-t|\leq\delta} |f(s) - f(t)|, 0 < \delta \leq \theta_2 - \theta_1$. Let us choose $N(\varepsilon, \varepsilon_1)$ so that for any $n > N(\varepsilon, \varepsilon_1)$, $\Pr\left(|x_n - x_0| > \delta(\varepsilon_1)\right) < \frac{\varepsilon}{2}$. Then, for $n > N(\varepsilon, \varepsilon_1)$, we have: $\Pr\left(|f_n(x_n) - f_n(x_0)| > \varepsilon_1\right) =$ $\Pr\left(|f_n(x_n) - f_n(x_0)| > \varepsilon_1 \text{ and } |x_n - x_0| \leq \delta(\varepsilon_1)\right) + \Pr\left(|f_n(x_n) - f_n(x_0)| > \varepsilon_1 \text{ and } |x_n - x_0| > \delta(\varepsilon_1)\right) \leq$ $\Pr\left(f_n(x) \notin K\right) + \Pr\left(|x_n - x_0| > \delta(\varepsilon_1)\right) < \varepsilon$, which proves the lemma.□

2.3 Proof of Theorem 2 iii)

Using Lemma 14, it is easy to establish the probability limits of the first k eigenvalues of XX'/T. Recall that by Lemma 10, we should look at the probability limits of the solutions to $\mu_j\left(M_n^{(1)}(x)\right) = 1$. Consider, first, solutions to a related equation $\mu_j\left(M_0^{(1)}(x)\right) = 1$. Function $\mu_j\left(M_0^{(1)}(x)\right) = \left(d_j + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}$ is continuous and strictly decreasing on $(b, +\infty)$, and tends to zero as $x \to +\infty$. In addition, since, as is straightforward to check, $\lim_{x\downarrow b} \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} = \frac{1}{c\sigma^2} \frac{\sqrt{c}}{1+\sqrt{c}}$, we have: $\lim_{x\downarrow b} \mu_m(M_0^{(1)}(x)) > 1$ if and only if $d_j > \sqrt{c}\sigma^2$. Therefore, there exist unique solutions $x_{0j} \in (b, +\infty)$ to equations $\mu_j\left(M_0^{(1)}(x)\right) = 1$ for $j \leq q$, and there are no solutions to the equations on $(b, +\infty)$ for $q < j \leq k$.

Now, fix θ_1 and θ_2 so that $\theta_2 > \theta_1 > b$; $\{x_{0j} : j \leq q\} \in (\theta_1, \theta_2)$, and $\left(d_k + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{\theta_2 - \lambda} < \frac{1}{2}$. The continuous mapping theorem and Lemma 14 imply that $\mu_j\left(\hat{M}_n^{(1)}(x)\right) \xrightarrow{d} \mu_j\left(M_0^{(1)}(x)\right)$, in the sense of the weak convergence of the random elements of $C[\theta_1, \theta_2]$. Using this convergence and the monotonicity of $\mu_j\left(\hat{M}_n^{(1)}(x)\right)$ it is easy to show that with high probability there exist unique solutions $x_{nj} \in [\theta_1, \theta_2]$ to $\mu_j\left(\hat{M}_n^{(1)}(x)\right) = 1$ for $j \leq q$, and $x_{nj} \xrightarrow{p} x_{0j}$.[6] Therefore, $\mu_j(\frac{1}{T}XX') \xrightarrow{p} x_{0j}$ for $j \leq q$.

Let us show that $x_{0j} = \frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2 c\right)}{d_i}$. Recall that x_{0j} was defined as the solution to equation $\left(d_j + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} = 1$, and it is the probability limit of $\mu_j(\frac{1}{T}XX')$. Changing the roles of factors and factor loadings, it is straightforward to show that y_{0j} defined as the solution to $\left(cd_j + \sigma^2 \frac{1}{c}\right) \int \frac{d\mathcal{F}_{\frac{1}{c}}(\lambda)}{y-\lambda} = 1$ must be the probability limit of $\mu_j(\frac{1}{n}X'X)$. But $\mu_j(\frac{1}{T}XX') = \frac{n}{T}\mu_j(\frac{1}{n}X'X)$. Hence, $x_{0j} = cy_{0j}$ and $\frac{d_j+\sigma^2}{c} \int \frac{\mathcal{F}_{\frac{1}{c}}(d\lambda)}{\frac{1}{c}x_{0j}-\lambda} = 1$. Now, it is straightforward to check that $f_{\frac{1}{c}}(\lambda) = c^2 f_c(c\lambda)$ and $\mathcal{F}_{\frac{1}{c}}$ does not have mass at zero if $c > 1$ and has mass at zero equal to $1 - c$ if $c < 1$. Therefore, we have

---

[6] When there is no solution to $\mu_j\left(\hat{M}_n^{(1)}(x)\right) = 1$ on $[\theta_1, \theta_2]$, we can define $x_{nj} \in [\theta_1, \theta_2]$ arbitrarily.

$c\left(d_j+\sigma^2\right)\left(\int \frac{\mathcal{F}_c(d\lambda)}{x_{0j}-\lambda} - \frac{1-\frac{1}{c}}{x_{0j}}\right) = 1$. Substituting $\int \frac{\mathcal{F}_c(d\lambda)}{x_{0j}-\lambda}$ by $\left(d_j+\sigma^2 c\right)^{-1}$ in the latter equation, we get $1 = c\left(d_j+\sigma^2\right)\left(\left(d_j+\sigma^2 c\right)^{-1} - \frac{1-\frac{1}{c}}{x_{0j}}\right)$, which implies that $x_{0j} = \frac{\left(d_i+\sigma^2\right)\left(d_i+\sigma^2 c\right)}{d_i}$.

We now show that, for any $j \leq q$,

$$\mu_j(\hat{M}_n^{(1)}(x)) = \mu_j\left(M_0^{(1)}(x)\right) + \frac{1}{\sqrt{n}} N_{n,jj}^{(1)}(x) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{55}$$

where $o_p\left(\frac{1}{\sqrt{n}}\right)$ is understood as a random element of $C\left[\theta_1, \theta_2\right]$, which, when multiplied by $\sqrt{n}$, tends in probability to zero as $n \to \infty$. Formula (55) is an easy consequence of Lemma 14 and part i of Lemma 15.

Now, let us define function $\nu_j\left(y\right)$ for $y > 0$ so that it is equal to b if $y > \lim_{x \downarrow b} \mu_j(M_0^{(1)}(x))$ and to the inverse function to function $\mu_j(M_0^{(1)}(x))$ otherwise. Since $\frac{d}{dx}\mu_j(M_0^{(1)}(x)) = -\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{(x-\lambda)^2}$, it is easy to see that $\lim_{x\downarrow b} \frac{d}{dx}\mu_j(M_0(x)) = +\infty$, and, hence, $\nu_j\left(y\right)$ is differentiable for $y > 0$. Applying ν_j to both sides of (55) and using the first order Taylor expansion of the right hand side, we have for $x \in [\theta_1, \theta_2]$: $\nu_j\left(\mu_j(\hat{M}_n^{(1)}(x))\right) = x + \nu_j'\left(\tau_n(x)\right)\frac{1}{\sqrt{n}} N_{n,jj}^{(1)}(x) + o_p\left(\frac{1}{\sqrt{n}}\right)$, where $\tau_n(x)$ is a random element of $C\left[\theta_1, \theta_2\right]$ such that $\tau_n(x) \xrightarrow{p} \mu_j(M_0^{(1)}(x))$ as $n \to \infty$.

Note that by definition of x_{nj}, definition of $\nu_j(\cdot)$, and Lemma 10, $\mu_j\left(\hat{M}_n^{(1)}(x_{nj})\right) = 1$, $\nu_j\left(\mu_j\left(\hat{M}_n^{(1)}(x_{nj})\right)\right) = x_{0j}$, and $x_{nj} = \mu_j\left(\frac{1}{T}XX'\right)$ with probability arbitrarily close to 1 for large enough n. Substituting x by x_{nj} in the above expansion of $\nu_j\left(\mu_j(\hat{M}_n^{(1)}(x))\right)$ and using these facts, we obtain: $\sqrt{n}\left(\mu_j\left(\frac{1}{T}XX'\right) - x_{0j}\right) = -\nu_j'\left(\tau_n(x_{nj})\right) N_{n,jj}^{(1)}(x_{nm}) + o_p\left(1\right)$.

Further, since $x_{nj} \xrightarrow{p} x_{0j}$ and $\mu_j(M_0^{(1)}(x_{0j})) = 1$, we have: $\nu_j'\left(\tau_n(x_{nj})\right) \xrightarrow{p} \nu_j'\left(1\right)$. Finally, $N_{n,jj}^{(1)}(x_{nj}) - N_{n,jj}^{(1)}(x_{0j}) \xrightarrow{p} 0$, which follows from Lemma 14 and Lemma 16. Therefore, $\sqrt{n}\left(\mu_j\left(\frac{1}{T}XX'\right) - x_{0j}\right)$ has the following form

$$\sqrt{n}\left(\mu_j\left(\frac{1}{T}XX'\right) - x_{0j}\right) = -\nu_j'\left(1\right) N_{n,jj}^{(1)}(x_{0j}) + o_p\left(1\right). \tag{56}$$

Let us show that $-\nu_j'\left(1\right) = \left(d_j^2 - \sigma^4 c\right)\left(d_j+\sigma^2 c\right) d_j^{-2}$. Indeed, by definition, $\nu_j'\left(1\right) = \left(\mu_j'\left(M_0^{(1)}(x_{0j})\right)\right)^{-1} = \left(-\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{(x_{0j}-\lambda)^2}\right)^{-1}$. The latter expression can be simplified as follows. Consider x_{0j} as a function of $d_j : x_{0j} = \left(d_j+\sigma^2\right)\left(d_j+\sigma^2 c\right)/d_j$. Note that since x_{0j} is defined as the solution to equation $\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda} = 1$, we must have:

$$\left(d_j+\sigma^2 c\right)\int \frac{d\mathcal{F}_c(\lambda)}{x_{0j}-\lambda} = 1 \tag{57}$$

Differentiating both sides of (57) with respect to d_j, we get: $\left(d_j+\sigma^2c\right)^{-1}-\left(d_j+\sigma^2c\right)\int\frac{d\mathcal{F}_c(\lambda)}{(x_{0j}-\lambda)^2}\left(1-\frac{\sigma^4c}{d_j^2}\right)=0$. Solving this equation for the integral, we get:

$$\int\frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^2}=\frac{d_j^2}{\left(d_j^2-\sigma^4c\right)\left(d_j+\sigma^2c\right)^2}, \tag{58}$$

and therefore $-\nu_j'(1)=\frac{\left(d_j^2-\sigma^4c\right)\left(d_j+\sigma^2c\right)}{d_j^2}$. The latter equality, formula (56), and Lemma 14 imply statement iii of Theorem 2.□

2.4 Proof of theorem 2 ii)

First, note that representation $\hat{\mathcal{L}}_{1:q}=\mathcal{L}\cdot\tilde{P}+\mathcal{L}_q^{\perp}$, where $\mathcal{L}_q^{\perp}$ is a matrix with q columns orthogonal to $\operatorname{span}(\mathcal{L})$ is a trivial coordinate decomposition statement. The value of Theorem 2 is, therefore, contained in describing properties of $\mathcal{L}_q^{\perp}$ and $\tilde{P}$. Recall that the columns of $\hat{\mathcal{L}}_{1:q}$ are equal to the q principal eigenvectors of $\frac{1}{T}XX'$. By Assumption 2a, the joint distribution of elements of X is invariant with respect to multiplication of X from the left by any orthogonal matrix leaving columns of L unchanged. This immediately implies that the joint distribution of entries of $\mathcal{L}_q^{\perp}$ is invariant with respect to the multiplication of $\mathcal{L}_q^{\perp}$ from the left by any orthogonal matrix that has $\operatorname{span}(\mathcal{L})=\operatorname{span}(L)$ as its invariant subspace. In the rest of the proof we, therefore, focus on the properties of $\tilde{P}$.

By Lemma 10 and by definitions of $\hat{M}_n^{(1)}(x),\hat{M}_n^{(2)}(x)$ and $\hat{M}_n^{(3)}(x)$, the j-th column of $\tilde{P}$ equals

$$\tilde{P}_{\cdot j}=\left(w_{nj}'\hat{M}_n^{(2)}\left(x_{nj}\right)w_{nj}\right)^{-1/2}\hat{M}_n^{(3)}\left(x_{nj}\right)w_{nj}, \tag{59}$$

where w_{nj} is a unit-length eigenvector of $\hat{M}_n^{(1)}(x_{nj})$ with high probability for large enough n. By part ii of Lemma 15, $w_{nj}\xrightarrow{p}e_j$. Further, Lemma 14, Lemma 16 and the fact that $x_{nj}\xrightarrow{p}x_{0j}$ imply that $\hat{M}_n^{(3)}(x_{nj})\xrightarrow{p}M_0^{(3)}(x_{0j})\equiv D^{1/2}\int\frac{d\mathcal{F}_c(\lambda)}{x_{0j}-\lambda}$ and $\hat{M}_n^{(2)}(x_{nj})\xrightarrow{p}\left(D+\sigma^2cI_k\right)\int\frac{d\mathcal{F}_c(\lambda)}{(x_{0j}-\lambda)^2}$. Therefore, by (59), we get: $\tilde{P}_{\cdot j}\xrightarrow{p}d_j^{1/2}\int\frac{d\mathcal{F}_c(\lambda)}{m_j-\lambda}\left(\left(d_j+\sigma^2c\right)\int\frac{d\mathcal{F}_c(\lambda)}{(m_j-\lambda)^2}\right)^{-1/2}e_j$. Formulae (57) and (58) imply that this limit simplifies so that we get: $\tilde{P}_{\cdot j}\xrightarrow{p}\left(\frac{d_j^2-\sigma^4c}{d_j(d_j+\sigma^2c)}\right)^{1/2}e_j$ which establishes the form of $\tilde{P}^{(1)}$ in Theorem 2 ii).

Now, we will study the asymptotic behavior of $\tilde{P}$ around its probability limit $\tilde{P}^{(1)}$. Let us show that the asymptotic joint distribution of the components of q $k\times 1$ vectors $\sqrt{n}\left(\tilde{P}_{\cdot j}-\tilde{P}_{\cdot j}^{(1)}\right),j=1,...,q$ is the same as that of the components of q $k\times 1$ vectors $\sum_{s=1}^{4}\varkappa_j\tilde{A}_j^{(s)},j=1,...,q$, where $\tilde{A}_j^{(1)}=N_n^{(3)}(x_{0j})e_j$, $\tilde{A}_j^{(2)}=-0.5\left(d_j^2-c\sigma^4\right)d_j^{-3/2}N_{n,jj}^{(2)}(x_{0j})e_j$,

$\tilde{A}_j^{(3)} = \sigma^4 c \left(d_i^2 - c\sigma^4\right)^{-1} d_i^{-1/2} N_{n,jj}^{(1)}\left(x_{0j}\right) e_j$, $\tilde{A}^{(4)} = -D^{1/2} S_j N_n^{(1)}\left(x_{0j}\right) e_j$,
and $\varkappa_j = \left(d_j^2 - \sigma^4 c\right)^{1/2} \left(d_j + \sigma^2 c\right)^{1/2} d_j^{-1}$.

Representation (59) implies that $\sqrt{n}\left(\tilde{P}_{\cdot j} - \tilde{P}_{\cdot j}^{(1)}\right) = \sum_{s=1}^4 A_j^{(s)} + o_p(1)$, where
$A_j^{(1)} = N_n^{(3)}\left(x_{nj}\right) w_{nj} \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2}$,
$A_j^{(2)} = \sqrt{n}\left(D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x_{nj}-\lambda} - D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x_{0j}-\lambda}\right) w_{nj} \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2}$,
$A_j^{(3)} = D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x_{0j}-\lambda} \sqrt{n}\left(w_{nj} - e_j\right) \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2}$,
$A_j^{(4)} = D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x_{0j}-\lambda} e_j \sqrt{n}\left(\left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2} - p\lim \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2}\right)$.

Consider, first $A_j^{(4)}$ and $A_j^{(2)}$. Using the Taylor expansion of function $x^{-1/2}$ around probability limit of $w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}$, we get:

$$\begin{aligned}
&\sqrt{n}\left(\left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2} - p\lim \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2}\right) \\
= \;& -\frac{1}{2} p\lim \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-3/2} \sqrt{n}\left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj} - p\lim \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)\right) + \\
& o\left(\sqrt{n}\left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj} - p\lim \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)\right)\right)
\end{aligned}$$

As has been shown above, $\left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2} \xrightarrow{p} \left(\left(d_j + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^2}\right)^{-1/2}$. Combining this fact with formulae (57) and (58) and using the Taylor expansion of $N_n^{(2)}(x_{nj})$ around x_{0j} and Lemma 14, we get the following decomposition
$A_j^{(4)} = \varrho_j e_j \left(w_{nj} + e_j\right)' \hat{M}_n^{(2)}\left(x_{nj}\right) \sqrt{n}\left(w_{nj} - e_j\right) + \varrho_j e_j N_{n,jj}^{(2)}(x_{nj}) -$
$2\varrho_j e_j \left(d_j + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^3} \sqrt{n}\left(x_{nj} - x_{0j}\right) + o_p(1)$, where $\varrho_j = -0.5 \left(d_j^2 - \sigma^4 c\right)^{3/2} \left(d_j + \sigma^2 c\right)^{1/2} d_j^{-5/2}$.
Further, using Taylor expansion of function $\int \frac{d\mathcal{F}_c(\lambda)}{x-\lambda}$ around $x = x_{0j}$, $A_j^{(2)}$ can be transformed into
$-D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^2} \sqrt{n}\left(x_{nj} - x_{0j}\right) w_{nj} \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2} + o_p(1)$.

The formulae obtained for $A_j^{(4)}$ and $A_j^{(2)}$ imply that we have the following representation
$\sqrt{n}\left(\tilde{P}_{\cdot j} - \tilde{P}_{\cdot j}^{(1)}\right) = \sum_{s=1}^4 \hat{A}_j^{(s)} + o_p(1)$, where
$\hat{A}_j^{(1)} = N_n^{(3)}\left(x_{nj}\right) w_{nj} \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2}$,
$\hat{A}_j^{(2)} = \varrho_j e_j N_{n,jj}^{(2)}(x_{nj})$,
$\hat{A}_j^{(3)} = -\left(D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^2} w_{nj} \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2} + 2\varrho_j e_j \left(d_j + \sigma^2 c\right) \int \frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^3}\right) \sqrt{n}\left(x_{nj} - x_{0j}\right)$,
$\hat{A}_j^{(4)} = \left(D^{1/2} \int \frac{d\mathcal{F}_c(\lambda)}{x_{0j}-\lambda} \left(w_{nj}' \hat{M}_n^{(2)}\left(x_{nj}\right) w_{nj}\right)^{-1/2} + \varrho_j e_j \left(w_{nj} + e_j\right)' \hat{M}_n^{(2)}\left(x_{nj}\right)\right) \sqrt{n}\left(w_{nj} - e_j\right)$.

Statement ii) of Lemma 15 and Lemma 14 imply that

$$\sqrt{n}\left(w_{nj}-e_j\right)=-\tilde{S}\left(x_{nj}\right)N_n^{(1)}\left(x_{nj}\right)e_j+o_p(1), \tag{60}$$

where $\tilde{S}\left(x\right)=\left(\int\frac{d\mathcal{F}_c(\lambda)}{x-\lambda}\right)^{-1}\operatorname{diag}\left(\left(d_1-d_j\right)^{-1},...,\underbrace{0}_{j\text{-th position}},...,\left(d_k-d_j\right)^{-1}\right)$. Using the same argument as that in the above derivation of the explicit formula for $\nu_j'\left(1\right)$, we obtain

$$\begin{aligned}\int\frac{d\mathcal{F}_c(\lambda)}{\left(m_i-\lambda\right)^3}&=\frac{\left(d_i^3+c^2\sigma^6\right)d_i^3}{\left(d_i+c\sigma^2\right)^3\left(d_i^2-c\sigma^4\right)^3}, &(61)\\ \int\frac{d\mathcal{F}_c(\lambda)}{\left(m_i-\lambda\right)^4}&=\frac{\left(d_i^6+c^4\sigma^{12}+c\sigma^4d_i^4+4c^2\sigma^6d_i^3+c^3\sigma^8d_i^2\right)d_i^4}{\left(d_i+c\sigma^2\right)^4\left(d_i^2-c\sigma^4\right)^5}. &(62)\end{aligned}$$

Finally, the definitions of $\hat{A}_j^{(s)}$ and x_{nj}, the facts that $\left(w_{nj}'\hat{M}_n^{(2)}\left(x_{nj}\right)w_{nj}\right)^{-1/2}\xrightarrow{p}\left(\left(d_j+\sigma^2c\right)\int\frac{d\mathcal{F}_c(\lambda)}{\left(x_{0j}-\lambda\right)^2}\right)^{-1/2}$, $w_{jn}\xrightarrow{p}e_j$, $x_{nj}\xrightarrow{p}x_{0j}$, and $\hat{M}_n^{(2)}\left(x_{nj}\right)\xrightarrow{p}M_0^{(2)}(x_{0j})$, Lemma 16, and formulae (56), (60), (57), (58), and (61) imply that the distribution limit of $\left\{\sum_{s=1}^4\hat{A}_j^{(s)},j=1,...,q\right\}$ must be the same as that of $\left\{\sum_{s=1}^4\varkappa_j\tilde{A}_j^{(s)},j=1,...,q\right\}$, where $\varkappa_j$ and $\tilde{A}_j^{(s)}$ are as defined above.

Using Lemma 14, we conclude that the joint asymptotic distribution of the elements of $\sqrt{n}\left(\tilde{P}-\tilde{P}^{(1)}\right)$ is Gaussian. The elements of the covariance matrix of the asymptotic distribution of $\sqrt{n}\left(\tilde{P}-\tilde{P}^{(1)}\right)$ can be found[7] using the above definitions of $\tilde{A}_j^{(s)}, s=1,...,4$, the expressions for the covariance of $N_n^{(1)}\left(x_{0j}\right), N_n^{(2)}\left(x_{0j}\right)$, and $N_n^{(3)}\left(x_{0j}\right), j=1,...,q$ summarized in the definition of $\Omega^{(\cdot,\cdot)}$ given in Lemma 14, and formulae (57),(58),(61), and (62).□

3 Proof of Theorem 3

We will prove part ii) of the theorem. A proof of part i) is similar to the proof of part ii) and we omit it to save space. First, note that since the distribution of the data X does not depend on the multiplication of X from the left by any orthogonal matrix having $\operatorname{span}(L)$ as its invariant subspace, the joint distribution of the coordinates of the columns of $\hat{\mathcal{L}}$ in the basis formed by the columns of O_L, which is an arbitrary orthogonal matrix with first k columns equal to $\mathcal{L}$, does not depend on how the $k+1$-th, $k+2$-th, ..., n-th columns of O_L are chosen.

Denote an $n\times1$ unit-length vector with all entries but the j-th equal to zero as e_j. Let

---

[7] To obtain these formulas we used symbolic manipulation software of the Scientific Workplace, version 5.

the $k+1$-th column of O_L be chosen as $M(L)e_{j_1}/\left\|M(L)e_{j_1}\right\|$, where $M(L)$ denotes the operator of taking the residual from the orthogonal projection on $\operatorname{span}(L)$, the $k+2$-th column be chosen as $M([L,e_{j_1}])e_{j_2}/\left\|M([L,e_{j_1}])e_{j_2}\right\|$, ..., and the $k+r$-th column be chosen as $M(\left[L,e_{j_1},...,e_{j_{r-1}}\right])e_{j_r}/\left\|M(\left[L,e_{j_1},...,e_{j_{r-1}}\right])e_{j_r}\right\|$. For example, if $r=2$ and $j_1=1$ and $j_2=2$, then matrix O_L has the following structure

$$O_L = \left[\begin{array}{c|ccccc} & x & 0 & 0 & \cdots & 0 \\ \mathcal{L} & y & z & 0 & \cdots & 0 \\ \cline{2-6} & & & * & & \end{array}\right], \tag{63}$$

where $x=\left\|M(L)e_1\right\|$, $y=e_2'M(L)e_1/\left\|M(L)e_1\right\|$, and $z=\left\|M([L,e_1])e_2\right\|$. Note that:

$$\begin{aligned} x^2 &= e_{j_1}'M(L)e_{j_1} = 1-e_{j_1}'L\left(L'L\right)^{-1}L'e_{j_1} = 1-\sum_{i=1}^{k}\mathcal{L}_{j_1 i}^2 && (64)\\ y &= \frac{1}{x}e_{j_2}'M(L)e_{j_1} = -\frac{1}{x}\sum_{i=1}^{k}\mathcal{L}_{j_1 i}\mathcal{L}_{j_2 i}. && (65)\end{aligned}$$

Let us denote the $n-k$ coordinates of the columns of $\hat{\mathcal{L}}_{1:q}$ in the basis formed by the columns of O_L as $R^{\perp}$. That is, $R^{\perp}_{ij}$ is the scalar product of $\hat{\mathcal{L}}_{\cdot j}$ and the $k+i$-th column of O_L. Then, $\hat{\mathcal{L}}_{j_s i} = \mathcal{L}_{j_s \cdot}\cdot\tilde{P}_{\cdot i}+\sum_{t=1}^{r}O_{L,j_s t}\cdot R^{\perp}_{ti}$. Hence, we can obtain the asymptotic joint distribution of $\left\{\hat{\mathcal{L}}_{j_s i}; s=1,...,r; i=1,...,q\right\}$ from the asymptotic joint distribution of the entries of $\tilde{P}$ and the first r columns of $R^{\perp}$.

It is easy to see that matrix $\tilde{P}^{\perp}\equiv R^{\perp}\left(I_q-\tilde{P}'\tilde{P}\right)^{-1/2}$, where $\tilde{P}$ is as defined in Theorem 2, has orthonormal columns. Moreover, as a consequence of the invariance of the distribution of X with respect to the orthogonal transformations leaving L unchanged, the joint distribution of the entries of $\tilde{P}^{\perp}$ conditional on $\tilde{P}$ is invariant with respect to multiplication of $\tilde{P}^{\perp}$ from the left by any orthogonal matrix. This implies that the joint distribution of the entries of $\tilde{P}^{\perp}\alpha$ conditional on $\tilde{P}$, where α is any $q\times 1$ unit-length vector, is the same as the joint distribution of the entries of $\xi/\left\|\xi\right\|$, where ξ is an $(n-k)\times 1$ vector with i.i.d. Gaussian entries.

As a consequence of the above result, the entries of $\tilde{P}^{\perp}\alpha$ are independent from the entries of $\tilde{P}\alpha$, and their unconditional joint distribution is the same as that of the entries of $\xi/\left\|\xi\right\|$. This fact, together with Theorem 2 and Cramer-Wold theorem (see White (1999), p.114), implies that the entries of $\sqrt{n}\left(\tilde{P}-\tilde{P}^{(1)}\right)$ and of the first r rows of $\sqrt{n}R^{\perp}$, where r is any fixed positive number, are asymptotically independent and have asymptotic joint zero-mean Gaussian distribution. The covariance matrix of the asymptotic distribution of the first r raws of $\sqrt{n}R^{\perp}$ is diagonal and $\operatorname{Avar}\left(\sqrt{n}R^{\perp}_{ji}\right)=1-\left(\tilde{P}^{(1)}_{ii}\right)^2$.

The asymptotic joint Gaussianity of the entries of $\sqrt{n}\left(\tilde{P}-\tilde{P}^{(1)}\right)$ and $\sqrt{n}R^{\perp}$ implies that $\left\{\sqrt{n}\left(\hat{\mathcal{L}}_{j_g i}-\tilde{P}_{ii}^{(1)}\mathcal{L}_{j_g i}\right); g=1,...,r; i=1,...,q\right\}$ are asymptotically jointly mean-zero Gaussian. We will now find the variances and covariances of the asymptotic distribution. Consider the random variables $\sqrt{n}\left(\hat{\mathcal{L}}_{j_g i}-\tilde{P}_{ii}^{(1)}\mathcal{L}_{j_g i}\right)$ and $\sqrt{n}\left(\hat{\mathcal{L}}_{j_f p}-\tilde{P}_{pp}^{(1)}\mathcal{L}_{j_f p}\right)$. Without loss of generality assume that $g=1$, $f=2$. If $g\neq 1$ and/or $f\neq 2$, construct O_L so that its $k+1$-th column is $M(L)e_{j_g}/\left\|M(L)e_{j_g}\right\|$ and its $k+2$-th column is $M([L,e_{j_g}])e_{j_f}/\left\|M([L,e_{j_g}])e_{j_f}\right\|$. From (63), we have: $\sqrt{n}\left(\hat{\mathcal{L}}_{j_g i}-\tilde{P}_{ii}^{(1)}\mathcal{L}_{j_g i}\right)=\sum_{1\leq s\leq k}\mathcal{L}_{j_g s}\sqrt{n}\left(\tilde{P}_{si}-\tilde{P}_{si}^{(1)}\right)+x\sqrt{n}R_{1i}^{\perp}$, and $\sqrt{n}\left(\hat{\mathcal{L}}_{j_f p}-\tilde{P}_{pp}^{(1)}\mathcal{L}_{j_f p}\right)=\sum_{1\leq s\leq k}\mathcal{L}_{j_f s}\sqrt{n}\left(\tilde{P}_{sp}-\tilde{P}_{sp}^{(1)}\right)+y\sqrt{n}R_{1p}^{\perp}+z\sqrt{n}R_{2p}^{\perp}$. These two formulae together with (64), (65), and the formulae for the asymptotic covariance of entries of $\sqrt{n}\left(\tilde{P}-\tilde{P}^{(1)}\right)$ and of the first two rows of $\sqrt{n}R^{\perp}$ established above and in Theorem 2 imply the formula for the asymptotic covariance matrix claimed by Theorem 3.$\square$

References

[1] Anderson, T.W. (1984) An introduction to multivariate statistical analysis, second edition, John Wiley & Sons, New York, Chichister, Brisbane, Toronto, Singapore.

[2] Athreya, K.B. and S.N. Lahiri (2006) Measure Theory and Probability Theory, Springer New York.

[3] Bai, Z.D., and J.W. Silverstein (1998) "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large-Dimensional Sample Covariance Matrices", Annals of Probability 26, 316-345

[4] Bai, Z.D., and J.W. Silverstein (2004) "CLT of linear spectral statistics of large dimensional sample covariance matrices", *Annals of Probability* 32(1A) pp. 553-605

[5] Billingsley, P. (1968) Convergence of probability measures, John Wiley & Sons, New York, Chichester, Brisbane, Toronto.

[6] Billingsley, P. (1999) Convergence of probability measures, second edition, Wiley, New York.

[7] Hall, P., and C.C. Heyde (1980) Martingale limit theory and its application, New York : Academic Press.

[8] Horn, R.A. and C. R. Johnson (1985) Matrix Analysis, Cambridge University Press, Cambridge, New York.

[9] Kato, T (1980) Perturbation theory for linear operators, Springer-Verlag, Berlin, New York.

[10] Krantz, S. G. (1992) Function theory of several complex variables. Second Edition. AMS Chelsea Publishing. Providence, Phode Island.

[11] Marčenko, V.A., and L.A. Pastur (1967) “Distribution of eigenvalues for some sets of random matrices”, Math. USSR-Sbornik, vol. 1, no. 4, 457-483

[12] McLeish, D.L. (1974) "Dependent Central Limit Theorems and Invariance Principles", *Annals of Probability*, Vol. 2, No. 4, p.620-628.

[13] Schiff, J.L. (1993) Normal Families, Springer-Verlag New York.

[14] Silverstein, J.W. and Bai, Z.D. (1995) “On the empirical distribution of eigenvalues of large dimensional random matrices”, Journal of Multivariate Analysis 54, 175-192

[15] White, H. (1999) Asymptotic theory for econometricians, revised edition, Academic press, San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo.

[16] Yin, Y.Q., Z.D. Bai, and P.R. Krishnaiah (1988) “On the limit of the largest eigenvalue of the large dimensional sample covariance matrix”, *Probability Theory and Related Fields* 78, pp.509-521.