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

Alexei Onatski*
Economics Department, Columbia University

June 17, 2010

Abstract

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

1 Notation and a useful convention

In the proofs below, we frequently use the following notation.

$\lambda_i(M)$ is the i-th largest by absolute value eigenvalue of matrix M.

$\|M\|$ is a norm of M, equal to $\sqrt{\lambda_1(M'M)}$.

e_i is a vector with all components zero except the i-th component, which equals 1. The dimensionality of e_i may vary.

$\mathcal{A}_0$ is an $n \times n$ diagonal matrix with the i-th diagonal element $\sqrt{a_i}$.

$\mathcal{B}_0$ is a $T \times T$ diagonal matrix with the i-th diagonal element $\sqrt{b_i}$.

$\mathcal{A}$ is an $(n-k) \times (n-k)$ diagonal matrix with the i-th diagonal element $\sqrt{a_{k+i}}$.

$\mathcal{B}$ is an $(T-k) \times (T-k)$ diagonal matrix with the i-th diagonal element $\sqrt{b_{k+i}}$.

*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

Let $A = U_A \mathcal{A}_0 V_A$ and $B = U_B \mathcal{B}_0 V_B$ be singular value decompositions of A and B. We will assume that the eigenvalues and eigenvectors in these decompositions are ordered so that the matrix of the first k columns of U_A equals $L\left(L'L\right)^{-1/2}$ and the matrix of the first k rows of V_B equals $F'/\sqrt{T}$. That such an ordering is possible follows from the Assumptions 2 ii) and iii). Note that under such an ordering, $a_1 = ... = a_k = 1$ and $b_1 = ... = b_k = 1$ so that the matrices in the intersection of the first k rows and columns of $\mathcal{A}_0$ and $\mathcal{B}_0$ are the identity matrices.

Without loss of generality, we will assume that

$$
\begin{aligned}
X &= \sum_{i=1}^{k} L_i F_i' + \mathcal{A}_0 \varepsilon \mathcal{B}_0 \text{ with} &(1)\\
L_i &= e_i \sqrt{L_i' L_i} \text{ for } i = 1, ..., k \text{ and} &(2)\\
F_i &= e_i \sqrt{T} \text{ for } i = 1, ..., k. &(3)
\end{aligned}
$$

For the purpose of the proof of Theorems 1, 2 and 3, there is no loss of generality in such an assumption. Indeed, note that the objects of study of these theorems: $\hat{\beta}$, $\hat{\alpha}$ and $\hat{L}'\hat{L}$ are invariant with respect to the following transformation: $X \rightsquigarrow UXV$, $L \rightsquigarrow UL$, $F \rightsquigarrow V'F$ and $e \rightsquigarrow UeV$, where U and V are any orthogonal matrices. Choosing $U = U_A'$ and $V = V_B'$, we will satisfy conventions (2) and (3). To see that convention (1) is also satisfied, note that the distribution of $V_A \varepsilon U_B$ is the same as that of ε because ε_{it} are i.i.d. $N\left(0, \sigma^2\right)$.

2 Proof of Theorem 1

In the proof of Theorem 1, we will assume that $\operatorname{Var} \varepsilon_{it} \equiv \sigma^2 = 1$. Such an assumption is without loss of generality. In the general case, variables such as X, e, L and D in the proof below, should be replaced by X/σ, e/σ, L/σ and D/σ^2, which, although does not change the proof substantially, complicates notation by making it necessary to keep variable σ in the equations. Further, we will assume, also without loss of generality, that the eigenvalues a_i, $i = 1, ..., n$ of AA' and the eigenvalues $b_i, i = 1, ..., T$ of $B'B$ are non-zero. If some of them are exactly zero, we will change them so they become positive but decrease to zero as $n \to \infty$ so fast that the asymptotics of the principal components estimator does not change.

2.1 Truncation and re-normalization

Let $\bar{\varepsilon}_{it} = \left(\operatorname{Var} \tilde{\varepsilon}_{it}\right)^{-1/2} \left(\tilde{\varepsilon}_{it} - E\tilde{\varepsilon}_{it}\right)$ with $\tilde{\varepsilon}_{it} = \varepsilon_{it} I_{|\varepsilon_{it}| \leq \ln n}$ be a truncated, centralized and re-normalized version of ε_{it} and let $\bar{X} = LF' + \mathcal{A}_0 \bar{\varepsilon} \mathcal{B}_0$. For $\mu_i \equiv \lambda_i\left(XX'/T\right)$ and $\bar{\mu}_i \equiv \lambda_i\left(\bar{X}\bar{X}'/T\right)$, we have: $\max_{i \leq n} |\mu_i - \bar{\mu}_i| \overset{a.s.}{=} o(1)$. 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), we have:

$$\max_{i\leq n}\left|\sqrt{\mu_i}-\sqrt{\bar{\mu}_i}\right|\leq\left\|\mathcal{A}_0\left(\varepsilon-\bar{\varepsilon}\right)\mathcal{B}_0/\sqrt{T}\right\|\leq\left\|\mathcal{A}_0\right\|\left\|\mathcal{B}_0\right\|\left\|\varepsilon-\bar{\varepsilon}\right\|/\sqrt{T}.$$

By Assumption 3ii), $\|\mathcal{A}_0\|=O\left(1\right)$ and $\|\mathcal{B}_0\|=O\left(1\right).$ 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 $2E\left(\varepsilon_{it}^2I_{|\varepsilon_{it}|>\ln n}\right)$, 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: $\sqrt{\mu_1}=\left\|X/\sqrt{T}\right\|\leq\left\|LF'/\sqrt{T}\right\|+\left\|\mathcal{A}_0\right\|\left\|\varepsilon/\sqrt{T}\right\|\left\|\mathcal{B}_0\right\|.$ By Assumptions 1ii) and 1iii), $\left\|LF'/\sqrt{T}\right\|=O\left(1\right)$, and by Lemma 1, $\left\|\varepsilon/\sqrt{T}\right\|\overset{a.s.}{=}O\left(1\right).$ Hence, $\sqrt{\mu_1}\overset{a.s.}{=}O\left(1\right).$ Similarly, $\sqrt{\bar{\mu}_1}\overset{a.s.}{=}O\left(1\right)$, and therefore, $\max_{i\leq n}\left|\mu_i-\bar{\mu}_i\right|\overset{a.s.}{=}o(1)$.

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

$$\max_{i\leq n,t\leq T}\left|\varepsilon_{it}\right|\leq\ln n. \tag{4}$$

Assumption (4) is also without loss of generality for the purpose of proving the convergence of the elements of the first q columns of $\hat{\beta}$ and $\hat{\alpha}$ stated in parts i) and ii) of Theorem 1. It is because the projections on the principal q eigenspaces of XX'/T (similarly, of $X'X/T$) and those on the principal q eigenspaces of $\bar{X}\bar{X}'/T$ (similarly, of $\bar{X}'\bar{X}/T$) converge to each other in probability 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:

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

Note that $P_i(0)$ and $P_i\left(1\right)$ 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 $\Pr\left(\max_{1\leq i\leq q} h_i < \rho\right) \to 0$ as $n\to\infty$. Therefore, $\Pr\left(r=\rho/2\right)\to 1$ as $n\to\infty$. Further, since $\left\|\mathcal{T}^{(1)}\right\| = \left\|\frac{\bar{X}\bar{X}'}{T} - \frac{XX'}{T}\right\| \leq |\mu_1 - \bar{\mu}_1| \overset{a.s.}{=} o(1)$, 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: $\|P_i\left(1\right) - P_i(0)\| \leq \sum_{t=1}^{\infty}\left\|\frac{1}{2\pi}\int_\Gamma R\left(\zeta\right)\left(\mathcal{T}^{(1)}R\left(\zeta\right)\right)^t d\zeta\right\| \leq \sum_{t=1}^{\infty} r^{-t}\left\|\mathcal{T}^{(1)}\right\|^t = \frac{\left\|\mathcal{T}^{(1)}\right\|}{r-\left\|\mathcal{T}^{(1)}\right\|} = o_p(1)$, 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 (4) holds. We will explicitly relax this assumption only when proving the convergence of the elements of the last $k-q$ columns of $\hat{\beta}$ and $\hat{\alpha}$.

2.2 Key lemma

Note that, under our convention (1,2,3), the j-th columns of $\hat{\alpha}$ and $\hat{\beta}$ equal the first k components of the unit-length j-th principal eigenvectors of $\frac{1}{T}XX'$ and $\frac{1}{T}X'X$, respectively. Further, $\hat{L}'\hat{L}$ equals a diagonal matrix with the first k eigenvalues of $\frac{1}{T}XX'$ on the diagonal. Lemma 2 below relates the eigenvalues and eigenvectors of the high-dimensional matrix $\frac{1}{T}XX'$ to the unit eigenvalues and the corresponding eigenvectors of the low-dimensional matrix-valued function $M^{(1)}\left(x\right)$, defined as follows.

Let us partition matrix ε as $\left[\varepsilon_1, \varepsilon_2\right]$, where ε_1 are the first k columns of ε. We define:

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

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

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

i) *μ is an eigenvalue of $\frac{1}{T}XX'$ 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*

$$\lambda_m\left(M^{(1)}\left(x\right)\right) = 1, ..., \lambda_{m+s-1}\left(M^{(1)}\left(x\right)\right) = 1, \tag{5}$$

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

$$y = \left(v' M^{(2)}(\mu) v\right)^{-1/2} (\mu I_n - \Lambda)^{-1} \Psi v \tag{6}$$

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

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

iv) *Consider matrix $\frac{1}{T} X X' + \varkappa e_i e_i'$, where $\varkappa$ is an arbitrary positive number. We have: μ is an eigenvalue of $\frac{1}{T} X X' + \varkappa e_i e_i'$ 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*

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

where $M_{\varkappa i}^{(1)}(x) \equiv \Psi_{\varkappa i}' (x I_n - \Lambda)^{-1} \Psi_{\varkappa i}$ and $\Psi_{\varkappa i} \equiv [\Psi, \sqrt{\varkappa} e_i]$.

Proof of Lemma 2: Let μ be an eigenvalue of $\frac{1}{T} X X'$ of multiplicity larger than or equal to s and let $y_1, ..., y_s$ be orthonormal eigenvectors corresponding to μ. Since $\frac{1}{T} X X' = \Lambda + \Psi \Psi'$, we have: $(\Lambda + \Psi \Psi') y_j = \mu y_j$ for $j = 1, ..., s$. Note that vectors $\Psi' y_1, ..., \Psi' y_s$ are linearly independent. Otherwise, if $\sum_{j=1}^s \beta_j \Psi' y_j = 0$ for some β_j that are not all equal to zero, we would have: $\Lambda \sum_{j=1}^s \beta_j y_j = (\Lambda + \Psi \Psi') \sum_{j=1}^s \beta_j y_j = \mu \sum_{j=1}^s \beta_j y_j$, which violates our assumption that $\mu \neq \lambda_i(\Lambda)$, $i = 1, ..., n$. Equation $(\Lambda + \Psi \Psi') y_j = \mu y_j$ implies that $\Psi' (\mu I_n - \Lambda)^{-1} \Psi \Psi' y_j = \Psi' y_j$. Hence, the space spanned by $\Psi' y_j$, $j = 1, ..., s$ 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 (5) 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 (6) 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 $\sum_{j=1}^s \beta_j y_j = 0$ for some β_j that are not all equal to zero, we would have, for $\gamma_j = \left(v_j' M^{(2)}(\mu) v_j\right)^{-1/2} \beta_j : \sum_{j=1}^s \gamma_j v_j = \sum_{j=1}^s \gamma_j M^{(1)}(\mu) v_j = \Psi' \sum_{j=1}^s \beta_j y_j = 0$, which violates our assumption that $v_1, ..., v_s$ are orthonormal. Equation (6) implies that $\Psi' y_j = \left(v' M^{(2)}(\mu) v\right)^{-1/2} M^{(1)}(\mu) v_j = \left(v' M^{(2)}(\mu) v\right)^{-1/2} v_j$ and therefore, $y_j = (\mu I_n - \Lambda)^{-1} \Psi \Psi' y_j$ for all $j = 1, ..., s$. The latter equality implies that $(\Lambda + \Psi \Psi') y_j = \mu y_j$, which means that y_j, $j = 1, ..., s$ are linearly independent eigenvectors of $\frac{1}{T} X X'$, 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 $\frac{1}{T}XX'$. Further, by definition, the j-th column of $\hat{\alpha}$ equals the first k components of the unit-length j-th principal eigenvector of $\frac{1}{T}XX'$. This fact, part ii) of the lemma and the definition of $M^{(3)}(\mu)$ imply that the j-th column of $\hat{\alpha}$ 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 Ψ by $\Psi_{\varkappa i}$ and $M^{(2)}(x)$ by $M^{(2)}(\mu) \equiv \Psi'_{\varkappa i}(xI_n - \Lambda)^{-2}\,\Psi_{\varkappa i}$ 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 $\frac{1}{T}XX'$ and of matrix $\hat{\alpha}$. Derivations of the probability limit of $\hat{\beta}$ is very similar to the derivations of the probability limit of $\hat{\alpha}$, and we will omit them to save space.

2.3 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:*

$$E\left(\left|\xi' Z\xi - \operatorname{tr} Z\right|^p |Z\right) \;\le\; C_{1p} n^{p/2} \|Z\|^p \left(\left[E\left|\xi_1\right|^4\right]^{p/2} + E\left|\xi_1\right|^{2p}\right), \tag{7}$$

$$E\left(\left|\xi' Z\zeta\right|^p |Z\right) \;\le\; C_{2p} n^{p/2} \|Z\|^p \left(\left[E\left|\xi_1\right|^4\right]^{p/2} + E\left|\xi_1\right|^{2p}\right), \tag{8}$$

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

Proof of Lemma 3: Inequality (7) is a slightly simplified version of the statement of Lemma 2.7 in Bai and Silverstein (1998). Inequality (8) follows from (7). Indeed, consider a vector $\varphi = \left(\xi', \zeta'\right)'$ and consider matrix $\tilde{Z} = \begin{pmatrix} 0 & Z \\ Z' & 0 \end{pmatrix}$. We have: $E\left(\left|\xi' Z\zeta\right|^p |Z\right) = E\left(\left|\frac{1}{2}\varphi'\tilde{Z}\varphi\right|^p |Z\right) \le 2^{-p} C_{1p} (2n)^{p/2} \left\|\tilde{Z}\right\|^p \left(\left[E\left|\xi_1\right|^4\right]^{p/2} + E\left|\xi_1\right|^{2p}\right)$, where the latter inequality follows from (7) 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 $\|\Sigma' Z\Pi\| \xrightarrow{p} 0$. 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: $\Pr\left(|(\Sigma' Z\Sigma)_{ii} - \operatorname{tr} Z| > \delta_1 | Z\right) \leq \delta_1^{-2} E\left(|(\Sigma' Z\Sigma)_{ii} - \operatorname{tr} Z|^2 |Z\right) \leq \delta_1^{-2} 2C_{12} n \|Z\|^2 \mu_4$, 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 $n > N$, $\Pr(\delta_1^{-2} 2C_{12} n \|Z\|^2 \mu_4 < \delta_2/2) > 1 - \delta_2/2$. Therefore, for all $n > N$, $\Pr\left(|(\Sigma' Z\Sigma)_{ii} - \operatorname{tr} Z| > \delta_1\right) = E\left[\Pr\left(|(\Sigma' Z\Sigma)_{ii} - \operatorname{tr} Z| > \delta_1 | Z\right)\right] \leq \delta_2/2\,(1 - \delta_2/2) + \delta_2/2 < \delta_2$, 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 $xI_{n-k} - \frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}$ 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 $xI_{n-k} - \frac{1}{T}\mathcal{A}\varepsilon_{22,-i}\mathcal{B}_{-i}^2\varepsilon_{22,-i}'\mathcal{A}$ 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 $\theta_1 > \bar{x}$, where $\bar{x}$ is as in Theorem 1. Then, for any $x > \theta_1$, Y is a positive definite matrix with $\left\|Y^{-1}\right\| \leq (\theta_1 - \bar{x})^{-1}$ 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:*

$$\lambda_1\left(Y^{-1}\right) \geq \lambda_1\left(Y_i^{-1}\right) \geq \lambda_2\left(Y^{-1}\right) \geq \ldots \geq \lambda_n\left(Y^{-1}\right) \geq \lambda_n\left(Y_i^{-1}\right) \tag{9}$$

and similarly,

$$\lambda_1\left(\mathcal{A}Y^{-1}\mathcal{A}\right) \geq \lambda_1\left(\mathcal{A}Y_i^{-1}\mathcal{A}\right) \geq \lambda_2\left(\mathcal{A}Y^{-1}\mathcal{A}\right) \geq \ldots \geq \lambda_n\left(\mathcal{A}Y^{-1}\mathcal{A}\right) \geq \lambda_n\left(\mathcal{A}Y_i^{-1}\mathcal{A}\right) \tag{10}$$

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 (2009), 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 $\mathcal{G}$ as n goes to infinity. As follows from the proof of Lemma 3 in Onatski (2009), the upper boundary of support of $\mathcal{G}$ equals $\bar{x}$, 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 $\bar{x}$ also implies that, for any $x > \theta_1$, $\left\|Y^{-1}\right\| \leq (\theta_1 - \bar{x})^{-1}$ for large n with probability 1.

Matrix Y_i is positive definite whenever Y is because $Y_i - Y = \frac{1}{T} b_{i+k} \mathcal{A} \varepsilon_{22,i} \varepsilon_{22,i}' \mathcal{A}$ 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 $\mathcal{A}^{-1} Y_i \mathcal{A}^{-1} - \mathcal{A}^{-1} Y \mathcal{A}^{-1} = \frac{1}{T} b_{i+k} \varepsilon_{22,i} \varepsilon_{22,i}'$ is a rank-one positive semidefinite matrix. $\square$

Lemma 6. *Suppose that Assumptions 1-3 hold. Let θ_1 be any number such that $\theta_1 > \bar{x}$, where $\bar{x}$ 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.

<u>Proof of Lemma 6:</u> 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 $\bar{Y}$. Indeed, let events Ξ, Ω and Ω_i be defined as:

$$\begin{aligned}
\Xi &= \left\{ Y_i \neq \bar{Y}_i \text{ for some } i \leq T-k, \text{ or } Y \neq \bar{Y} \right\}, \\
\Omega &= \left\{ Y \text{ is positive definite and } \left\| Y^{-1} \right\| \leq (\theta_1 - \bar{x})^{-1} \right\}, \\
\Omega_i &= \left\{ Y_i \text{ is positive definite and } \left\| Y_i^{-1} \right\| \leq (\theta_1 - \bar{x})^{-1} \right\}.
\end{aligned}$$

Then, $\Xi \cap \Omega = \varnothing$ because Ω implies that $Y = \bar{Y}$, and, as follows from Lemma 5, $\Omega \subseteq \Omega_i$ so that $Y_i = \bar{Y}_i$ too. Further, by Lemma 5, $\Pr(\Omega) \to 1$ as $n \to \infty$, and therefore, $\Pr(\Xi) \to 0$.

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:

$$\Pr\left(\max_{1\le i\le T-k}|U_{rs}(i)|>\delta_1\right)\le \Pr\left(\max_{1\le i\le T-k}|U_{rs}(i)|>\delta_1 \text{ and } \Omega\right)+\delta_2/2\le \tag{11}$$
$$\sum_{i=1}^{T-k}\Pr\left(|U_{rs}(i)|>\delta_1 \text{ and } \Omega_i\right)+\delta_2/2\le \sum_{i=1}^{T-k}E\left[\Pr\left(|U_{rs}(i)|>\delta_1 \text{ and } \Omega_i|\bar{Y}_i\right)\right]+\delta_2/2$$

If either $\bar{Y}_i$ is not positive definite or $\left\|\bar{Y}_i^{-1}\right\|>(\theta_1-\bar{x})^{-1}$, 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 $\left\|\bar{Y}_i^{-1}\right\|\le(\theta_1-\bar{x})^{-1}$, 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:

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

where the second line follows from Lemma 3 and from assumption (4). If $\left\|\bar{Y}_i^{-1}\right\|\le(\theta_1-\bar{x})^{-1}$, we can make the latter expression smaller than $\delta_2/(2T)$ by choosing $p>2$ and large enough n. Therefore, $E\left[\Pr\left(|U_{rs}(i)|>\delta_1 \text{ and } \Omega_i|\bar{Y}_i\right)\right]\le\delta_2/(2T)$ for all $i=1,...,T-k$ and large enough n. Using (11), we obtain: $\Pr\left(\max_{1\le i\le T-k}|U_{rs}(i)|>\delta_1\right)<\delta_2$ for large enough n. Since δ_1 and δ_2 were arbitrary positive numbers, we have: $\max_{1\le i\le T-k}|U_{rs}(i)|\xrightarrow{p}0$.

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

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

Therefore, setting $r=1$ and using the interlacing inequalities (10), we conclude that

$$0\ge V_{1s}(i)\ge-\frac{1}{T}\lambda_1^s\left(\mathcal{A}\bar{Y}^{-1}\mathcal{A}\right)\ge-\frac{1}{T}\|\mathcal{A}\|^{2s}(\theta_1-\bar{x})^{-s} \tag{12}$$

whenever Ω holds. Since $\Pr(\Omega)\to1$ as $n\to\infty$, the latter inequalities imply that $\max_{1\le i\le T-k}|V_{1s}(i)|\xrightarrow{p}0$ for $s=1$ and $s=2$.

It remains to prove the convergence to zero of $\max_{1\le i\le T-k}|V_{21}(i)|$. 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: $0 \geq V_{21}(i) \geq -\left(\max_{j=1,\ldots,n-k} a_{j+k}\right) \frac{1}{T} \operatorname{tr}\left(\bar{Y}^{-2} - \bar{Y}_i^{-2}\right)$. But $\max_{j=1,\ldots,n-k} a_{j+k} = \|\mathcal{A}\|^2$ and $\frac{1}{T} \operatorname{tr}\left(\bar{Y}^{-2} - \bar{Y}_i^{-2}\right) \leq \frac{1}{T} \lambda_1^2 \left(\bar{Y}^{-1}\right) = \frac{1}{T} \left\|\bar{Y}^{-1}\right\|^2 \leq \frac{1}{T} (\theta_1 - \bar{x})^{-2}$ when Ω holds, where the first of the latter two inequalities can be obtained from the interlacing inequalities (9) similarly to as (12) was obtained from (10). Thus, when Ω holds, we have: $0 \geq V_{21}(i) \geq -\frac{1}{T} \|\mathcal{A}\|^2 (\theta_1 - \bar{x})^{-2}$, and therefore $\max_{1 \leq i \leq T-k} |V_{21}(i)| \xrightarrow{p} 0.\square$

Lemma 7. *Let θ_1 be any number such that $\theta_1 > \bar{x}$, where $\bar{x}$ 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 $w(z) = \int \frac{\lambda d\mathcal{G}_A(\lambda)}{x - \left(z + \int \frac{\tau d\mathcal{G}_B(\tau)}{1 - \tau c w(z)}\right)\lambda}$ has a unique solution $w(z)$ such that $\operatorname{Im} w(z) > 0$. Function $w(z)$ 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:*

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

where, if Y is not invertible, we set the left hand of the above convergence statements to an arbitrary number, which equals neither $cw(0)$ nor $cw'(0)$. 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.

<u>Proof of Lemma 7:</u> 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] $\tilde{m}(z) = \tilde{m}_A\left(z + c^{-1} \int \frac{\tau d\mathcal{G}_B(\tau)}{1 - \tau \tilde{m}(z)}\right)$ that satisfies $\operatorname{Im} \tilde{m}(z) > 0$. Here, $\tilde{m}_A(z)$ 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 $1 - \lim_{\tau \uparrow xc^{-1}\lambda^{-1}} \mathcal{G}_A(\tau)$. Hence, $\tilde{m}_A(z) = \int \frac{\tau d\mathcal{G}_A(\tau)}{xc^{-1} - z\tau}$ and $\tilde{m}(z) = \int \frac{\lambda d\mathcal{G}_A(\lambda)}{xc^{-1} - \left(z + c^{-1} \int \frac{\tau d\mathcal{G}_B(\tau)}{1 - \tau \tilde{m}(z)}\right)\lambda}$. 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 $w(z) = c^{-1} \tilde{m}\left(c^{-1} z\right)$, which is an analytic function in the $\operatorname{Im} z > 0$ domain and which is the unique solution to equation $w(z) = \int \frac{\lambda d\mathcal{G}_A(\lambda)}{x - \left(z + \int \frac{\tau d\mathcal{G}_B(\tau)}{1 - \tau c w(z)}\right)\lambda}$ that satisfies $\operatorname{Im} w(z) > 0$.

[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 $-d\mathcal{G}_B(-\tau)$.

[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.

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

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 $w(z)$, by classical Weirstrass theorem, the derivatives of $m_n(z)$ also converge to the corresponding derivatives of $w(z)$, and this convergence is uniform on the compact subsets of U_0. We therefore have: $\frac{1}{T}\operatorname{tr}\left(\mathcal{A}Y^{-1}\mathcal{A}\right)^2\overset{a.s}{\to}cw'(0)$.

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.□

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

$$\begin{cases} v=x\left(c\int\frac{au}{u-a}d\mathcal{G}_A(a)\right)^{-1} \\ u=x\left(\int\frac{bv}{v-b}d\mathcal{G}_B(b)\right)^{-1}\end{cases}. \tag{13}$$

Then for function $w(z)$ defined in Lemma 7, we have:

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

where

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

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 $f_2(z,u,v)=x-uz+uv\left(1+vm_B(v)\right)$, where $m_A(u)$ and $m_B(v)$ are the Stieltjes transforms of $\mathcal{G}_A$ and $\mathcal{G}_B$, respectively. Further, consider a system

$\begin{cases} f_1(z,u,v)=0 \\ f_2(z,u,v)=0 \end{cases}$. Note that f_1 and f_2 are holomorphic functions of z, u and v near the point $(z,u,v)=(0,u_{2,x},v_{2,x})$. This follows from the fact that $m_A(u)$ and $m_B(v)$ are holomorphic at $u_{2,x}$ and $v_{2,x}$, respectively, which, in turn, follows from the fact that $u_{2,x}>\bar{x}_A$ and $v_{2,x}>\bar{x}_B$.

According to the holomorphic implicit function theorem (see Krantz (1992), p.54), there exists a unique holomorphic solution $\{u(z),v(z)\}$ to the above system in a neighborhood of $z=0$ such that $u(0)=u_{2,x}$ and $v(0)=v_{2,x}$ as long as $\det f'_{1,2}\neq 0$ at $(z,u,v)=(0,u_{2,x},v_{2,x})$, 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 (13): $v=x\left(c\int\frac{\lambda g_1(v)}{g_1(v)-\lambda}d\mathcal{G}_A(\lambda)\right)^{-1}$ and $g_2(v)=x\left(\int\frac{\tau v}{v-\tau}d\mathcal{G}_B(\tau)\right)^{-1}$, 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 $(u,v)=(u_{2,x},v_{2,x})$. The latter inequality is equivalent to the inequality

$$\frac{d}{du}\left[x\left(c\int\frac{\lambda u}{u-\lambda}d\mathcal{G}_A(\lambda)\right)^{-1}\right]\frac{d}{dv}\left[x\left(\int\frac{\tau v}{v-\tau}d\mathcal{G}_B(\tau)\right)^{-1}\right]<1. \tag{14}$$

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

$$\begin{array}{l} c\int\frac{\lambda u}{u-\lambda}d\mathcal{G}_A(\lambda)=-\frac{\partial f_1}{\partial v},\ \int\frac{\tau v}{v-\tau}d\mathcal{G}_B(\tau)=-\frac{\partial f_2}{\partial u} \text{ and} \\ \frac{d}{du}c\int\frac{\lambda u}{u-\lambda}d\mathcal{G}_A(\lambda)=-\frac{1}{v}\frac{\partial f_1}{\partial u},\ \frac{d}{dv}\int\frac{\tau v}{v-\tau}d\mathcal{G}_B(\tau)=-\frac{1}{u}\frac{\partial f_2}{\partial v} \end{array} \tag{15}$$

at $(z,u,v)=(0,u_{2,x},v_{2,x})$. Using (15) in (14), we obtain:

$$x^2\left(-\frac{\partial f_1}{\partial v}\right)^{-2}\left(-\frac{\partial f_2}{\partial u}\right)^{-2}\left(-\frac{1}{v}\frac{\partial f_1}{\partial u}\right)\left(-\frac{1}{u}\frac{\partial f_2}{\partial v}\right)<1 \tag{16}$$

at $(z,u,v)=(0,u_{2,x},v_{2,x})$. But at $(u,v)=(u_{2,x},v_{2,x})$, the curves $u=g_1(v)$ and $u=g_2(v)$ intersect. Therefore, $v_{2,x}u_{2,x}=x^2\left(c\int\frac{\lambda u_{2,x}}{u_{2,x}-\lambda}d\mathcal{G}_A(\lambda)\right)^{-1}\left(\int\frac{\tau v_{2,x}}{v_{2,x}-\tau}d\mathcal{G}_B(\tau)\right)^{-1}$, and hence, $\left(-\frac{\partial f_1}{\partial v}\right)\left(-\frac{\partial f_2}{\partial u}\right)=x^2/(v_{2,x}u_{2,x})$, which, together with (16), implies that $\det f'_{1,2}<0$ at $(z,u,v)=(0,u_{2,x},v_{2,x})$.

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,

$$\left(\frac{du}{dz},\frac{dv}{dz}\right)=-u_{2,x}\det\left(f'_{1,2}\right)^{-1}\left(\frac{\partial f_1}{\partial v},-\frac{\partial f_1}{\partial u}\right). \tag{17}$$

Equations (15) together with the fact that $u_{2,x} > \bar{x}_A$ and $v_{2,x} > \bar{x}_B$ 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 (17) is negative,

$$\frac{du}{dz} < 0 \text{ and } \frac{dv}{dz} < 0 \tag{18}$$

at $z = 0$.

The last of the two inequalities in (18) 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, $v^{-1}(z) = c \int \frac{\lambda d\mathcal{G}_A(\lambda)}{x - \left(z + \int \frac{\tau d\mathcal{G}_B(\tau)}{1 - \tau v^{-1}(z)}\right)\lambda}$. 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 $cw(z)$. Hence, $cw(z) = v^{-1}(z)$, and $cw(0) = v_{2,x}^{-1}$.

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 (17) and (15), 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,

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

where the latter equality follows from (15). Using the definition of the determinant $\det\left(f'_{1,2}\right)^{-1}$ and, once again, equations (15), we obtain:

$$\begin{aligned} \det\left(f'_{1,2}\right) &= c u_{2,x} v_{2,x} \int \frac{\lambda^2}{\left(u_{2,x} - \lambda\right)^2} d\mathcal{G}_A(\lambda) \int \frac{\tau^2}{\left(v_{2,x} - \tau\right)^2} d\mathcal{G}_B(\tau) - \\ &\quad c u_{2,x} v_{2,x} \int \frac{\lambda}{u_{2,x} - \lambda} d\mathcal{G}_A(\lambda) \int \frac{\tau}{v_{2,x} - \tau} d\mathcal{G}_B(\tau) \end{aligned}$$

so that, finally,

$$cw'(0) = \frac{d}{dz} v^{-1}(0) = -\frac{v_{2,x}^{-2} \int \frac{\lambda^2 d\mathcal{G}_A(\lambda)}{\left(u_{2,x} - \lambda\right)^2}}{\int \frac{\lambda^2 d\mathcal{G}_A(\lambda)}{\left(u_{2,x} - \lambda\right)^2} \int \frac{\tau^2 d\mathcal{G}_B(\tau)}{\left(v_{2,x} - \tau\right)^2} - \int \frac{\lambda d\mathcal{G}_A(\lambda)}{u_{2,x} - \lambda} \int \frac{\tau d\mathcal{G}_B(\tau)}{v_{2,x} - \tau}} = \frac{r_{u,x} u_{2,x}}{\left(1 - r_{u,x} r_{v,x}\right) x v_{2,x}},$$

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 (13), satisfy $u_{2,x} = x \left(\int \frac{\tau v_{2,x} d\mathcal{G}_B(\tau)}{v_{2,x} - \tau}\right)^{-1}$.□

Lemma 9. *Under assumptions of Lemma 8:*

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

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

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

iv) $M^{(1)}_{\varkappa i}(x) \xrightarrow{p} x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1}\begin{pmatrix} D & \sqrt{\varkappa d_i}e_i \\ \sqrt{\varkappa d_i}e_i' & \varkappa \end{pmatrix} + v_{2,x}^{-1}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$..

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

$$xI_n - \Lambda \equiv \begin{pmatrix} xI_k - \frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{12}' & -\frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{22}'\mathcal{A} \\ -\frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{12}' & Y \end{pmatrix}, \tag{19}$$

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 $x > \theta_1 > \bar{x}$, 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 $xI_n - \Lambda \equiv xI_n - \frac{1}{T}\mathcal{A}_0\varepsilon_2\mathcal{B}^2\varepsilon_2'\mathcal{A}_0$ for large n with probability 1. Below, we will work with Y and $xI_n - \Lambda$ 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:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Upsilon^{-1} & -\Upsilon^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}\Upsilon^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}\Upsilon^{-1}A_{12}A_{22}^{-1} \end{bmatrix}, \tag{20}$$

where $\Upsilon = A_{11} - A_{12}A_{22}^{-1}A_{21}$ is invertible as long as A is invertible. Applying formula (20) to (19), we find that, for any $x > \theta_1 > \bar{x}$, $M^{(1)}(x)$ 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

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

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

$$\left\| xI_k - \frac{1}{T}\varepsilon_{12}\mathcal{B}^2\varepsilon_{12}' - (x-1)I_k \right\| \xrightarrow{p} 0. \tag{21}$$

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 $n\|Z\|^2 \leq \frac{n}{T^2}\|\mathcal{A}\|^4\left\|Y^{-1}\right\|^2\|\varepsilon_{22}\|^4\|\mathcal{B}\|^8$ so that by Lemmas 1 and 5, $n\|Z\|^2 \xrightarrow{a.s.} 0$. Therefore, by Lemma 4:

$$\left\|\frac{1}{T}\varepsilon_{12}Z\varepsilon_{12}' - (\operatorname{tr} Z)\, I_k\right\| \xrightarrow{p} 0. \tag{22}$$

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:

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

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 $v_{2,x} > \bar{x}_B = \lim_{n\to\infty}\max_{i=1,\dots,T} b_i$, the quantity $1 - b_{k+i}\frac{1}{T}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}$ is separated from zero with probability arbitrarily close to 1 for large enough n. Therefore, we have:

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

so that $\left|\operatorname{tr} Z - \frac{1}{T}\sum_{i=1}^{T-k}\frac{b_{k+i}^2 v_{2,x}^{-1}}{1-b_{k+i}v_{2,x}^{-1}}\right| \xrightarrow{p} 0$. Finally, note that, since $b_1 = \dots = b_k = 1$:

$$\left|\frac{1}{T}\sum_{i=1}^{T-k}\frac{b_{k+i}^2 v_{2,x}^{-1}}{1-b_{k+i}v_{2,x}^{-1}} - \frac{1}{T}\sum_{i=1}^{T}\frac{b_i^2 v_{2,x}^{-1}}{1-b_i v_{2,x}^{-1}}\right| = \frac{k}{T}\frac{v_{2,x}^{-1}}{1-v_{2,x}^{-1}} \xrightarrow{p} 0 \tag{23}$$

and that, by Assumption 3i),

$$\frac{1}{T}\sum_{i=1}^{T}\frac{b_i^2 v_{2,x}^{-1}}{1-b_i v_{2,x}^{-1}} = \int\frac{\tau^2}{v_{2,x}-\tau}dG_B(\tau) \to \int\frac{\tau^2}{v_{2,x}-\tau}d\mathcal{G}_B(\tau). \tag{24}$$

Putting the latter three convergence statements together, we obtain:

$$\left|\operatorname{tr} Z - \int\frac{\tau^2}{v_{2,x}-\tau}d\mathcal{G}_B(\tau)\right| \xrightarrow{p} 0. \tag{25}$$

Combining (21), (22) and (25), we get: $\left\|K_1 - (x - 1 - \int\frac{\tau^2}{v_{2,x}-\tau}d\mathcal{G}_B(\tau))I_k\right\| \xrightarrow{p} 0$. But

$\int \frac{\tau^2}{v_{2,x}-\tau} d\mathcal{G}_B(\tau) = -1 + \int \frac{\tau v_{2,x}}{v_{2,x}-\tau} d\mathcal{G}_B(\tau) = -1 + x u_{2,x}^{-1}$, where the last equality holds because $(u_{2,x}, v_{2,x})$ is a solution to (13). Therefore, finally,

$$\left\| K_1 - x(1-u_{2,x}^{-1}) I_k \right\| \xrightarrow{p} 0. \tag{26}$$

Note that since $u_{2,x} > \bar{x}_A$, $u_{2,x}$ 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 $\left\| \frac{1}{T}\left(x\mathcal{A}_0^{-2} - \frac{1}{T}\varepsilon_2\mathcal{B}^2\varepsilon_2'\right)^{-1} \right\| \leq \frac{1}{T}\|\mathcal{A}_0\|^2 (\theta_1 - \bar{x})^{-1}$ 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,

$$\left\| K_2 - v_{2,x}^{-1} I_k \right\| \xrightarrow{p} 0. \tag{27}$$

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 $\max\{T-n,0\}$ zero columns and $\max\{n-T,0\}$ zero rows. Similarly, let $\bar{\varepsilon}_{12}$ be obtained from ε_{12} by adding $\max\{n-T,0\}$ columns with i.i.d. entries distributed as ε_{it}, and let $\bar{\varepsilon}_{21}$ be obtained from ε_{21} by adding $\max\{T-n,0\}$ 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 $\|\bar{G}\| \leq \frac{1}{T^{3/2}}\|\mathcal{B}\|^2\|\varepsilon_{22}\|\|\mathcal{A}\|^2(\theta_1-\bar{x})^{-1}$ for large n with probability 1. Using Lemma 1 and Assumption 1i), we further get: $n\|\bar{G}\|^2 \xrightarrow{a.s.} 0$ so that, by Lemma 4: $\|\varepsilon_{12}G\varepsilon_{21}\| = \|\bar{\varepsilon}_{12}\bar{G}\bar{\varepsilon}_{21}\| \xrightarrow{p} 0$. Combining this finding with the fact that $K_3 = K_1^{-1}\varepsilon_{12}G\varepsilon_{21}$ and with (26), we obtain:

$$\|K_3\| \xrightarrow{p} 0. \tag{28}$$

The convergence facts (26), (27) and (28) established above together with the fact that, by Assumption 1iii), $\frac{1}{T}\Delta'\Delta \xrightarrow{p} D$, imply that $M^{(1)}(x) \xrightarrow{p} x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1} D + v_{2,x}^{-1} I_k$.□

Proof of Lemma 9 ii): For $M^{(2)}(x)$, using the square of the inverse of a partitioned matrix formula (20), we have:

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

where

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

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 $n\left\|\tilde{Z}\right\|^2 \leq \frac{n}{T^2}\left\|\mathcal{A}\right\|^4\left\|Y^{-1}\right\|^4\left\|\varepsilon_{22}\right\|^4\left\|\mathcal{B}\right\|^8$ so that by Lemmas 1 and 5, $n\left\|\tilde{Z}\right\|^2 \xrightarrow{a.s.} 0$. Therefore, by Lemma 4:

$$
\left\|K_4 - \left(\operatorname{tr}\tilde{Z}\right)I_k\right\| \xrightarrow{p} 0. \tag{29}
$$

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

$$
v'\left(M - vv'\right)^{-2}v = \frac{v'M^{-2}v}{\left(1 - v'M^{-1}v\right)^2}. \tag{30}
$$

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

$$
\begin{aligned}
\operatorname{tr}\tilde{Z} &= \frac{1}{T^2}\sum_{i=1}^{T-k} b_{k+i}^2\varepsilon_{22,i}'\mathcal{A}\left(Y_i - \frac{1}{T}b_{k+i}\mathcal{A}\varepsilon_{22,i}\varepsilon_{22,i}'\mathcal{A}\right)^{-2}\mathcal{A}\varepsilon_{22,i} \qquad (31)\\
&= \frac{1}{T}\sum_{i=1}^{T-k} b_{k+i}\frac{\frac{1}{T}b_{k+i}\varepsilon_{22,i}'\mathcal{A}Y_i^{-2}\mathcal{A}\varepsilon_{22,i}}{\left(1 - \frac{1}{T}b_{k+i}\varepsilon_{22,i}'\mathcal{A}Y_i^{-1}\mathcal{A}\varepsilon_{22,i}\right)^2}.
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{1}{T}\operatorname{tr}\mathcal{A}Y^{-2}\mathcal{A} &= \frac{1}{T}\operatorname{tr}Y^{-1}\mathcal{A}^2Y^{-1} \\
&= \frac{1}{T}x^{-1}\operatorname{tr}Y^{-1}\mathcal{A}^2Y^{-1}\left(xI_{n-k} - \frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A} + \frac{1}{T}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A}\right) \\
&= \frac{1}{T}x^{-1}\operatorname{tr}Y^{-1}\mathcal{A}^2 + \frac{1}{T}x^{-1}\operatorname{tr}Y^{-1}\mathcal{A}^2Y^{-1}\mathcal{A}\varepsilon_{22}\mathcal{B}^2\varepsilon_{22}'\mathcal{A} \\
&= \frac{1}{T}x^{-1}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A} + \frac{1}{T^2}x^{-1}\operatorname{tr}\mathcal{B}\varepsilon_{22}'\left[\mathcal{A}Y^{-1}\mathcal{A}\right]^2\varepsilon_{22}\mathcal{B}. \qquad (32)
\end{aligned}
$$

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

$$
\frac{1}{T}x^{-1}\operatorname{tr}\mathcal{A}Y^{-1}\mathcal{A} \xrightarrow{p} x^{-1}v_{2,x}^{-1}. \tag{33}
$$

For the second term, using (30), we obtain:

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

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 $1-b_{k+i}v_{2,x}^{-1}$ is separated from zero for large n, we have:

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

and using equations analogous to (23) and (24), we obtain:

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

Combining the latter result with (32) and (33), we obtain:

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

The latter expression can be simplified. We have:

$$\begin{aligned} & x^{-1}v_{2,x}^{-1}+\frac{v_{2,x}^{-1}x^{-2}u_{2,x}r_{u,x}}{1-r_{u,x}r_{v,x}}\int\frac{\tau d\mathcal{G}_B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2} \\ = \; & x^{-1}v_{2,x}^{-1}\left(1-r_{u,x}r_{v,x}\right)^{-1}\left(1-r_{u,x}r_{v,x}+r_{u,x}u_{2,x}x^{-1}\int\frac{\tau v_{2,x}^2 d\mathcal{G}_B(\tau)}{\left(v_{2,x}-\tau\right)^2}\right) \end{aligned} \tag{34}$$

On the other hand,

$$u_{2,x}x^{-1}=\left(\int\frac{\tau v_{2,x}}{v_{2,x}-\tau}d\mathcal{G}_B(\tau)\right)^{-1} \tag{35}$$

because $(u_{2,x}, v_{2,x})$ solve system (13). Therefore,

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

Substituting this result in (34), we obtain:

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

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 (31), using equations analogous to (23) and (24), we obtain:

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

Note that $\int\frac{\tau^2 d\mathcal{G}_B(\tau)}{\left(1-\tau v_{2,x}^{-1}\right)^2}=v_{2,x}r_{v,x}\int\frac{\tau v_{2,x}d\mathcal{G}_B(\tau)}{v_{2,x}-\tau}=\frac{xv_{2,x}r_{v,x}}{u_{2,x}}$, where the last equality follows from (35). 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 (29), we obtain:

$$\left\| K_4 - \frac{r_{v,x}\left(1+r_{u,x}\right)}{u_{2,x}\left(1-r_{u,x}r_{v,x}\right)} I_k \right\| \xrightarrow{p} 0. \tag{36}$$

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 $\frac{1+r_{u,x}}{x v_{2,x}(1-r_{u,x}r_{v,x})}$ so that

$$\left\| K_5 - \frac{1+r_{u,x}}{x v_{2,x}\left(1-r_{u,x}r_{v,x}\right)} I_k \right\| \xrightarrow{p} 0. \tag{37}$$

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 (28), substituting Y^{-1} by Y^{-2} and $(\theta_1 - \bar{x})^{-1}$ by $(\theta_1 - \bar{x})^{-2}$, we conclude that it also converges in probability to zero. Hence,

$$\|K_6\| \xrightarrow{p} 0 \tag{38}$$

Finally, combining (26), (36), (37) and (38), we get:

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

□

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

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

Therefore, using (26) and (28), we get $M^{(3)}(x) \xrightarrow{p} x^{-1} \left(1 - u_{2,x}^{-1}\right)^{-1} D^{1/2}$.□

<u>Proof of Lemma 9 iv):</u> 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 $\xi' = \sqrt{\frac{\varkappa}{T}} e_i' K_1^{-1} \left(\Delta + \varepsilon_{11}\right) + \sqrt{\varkappa} e_i' K_3$. 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.□

In the next two sections, we will prove parts ii) and iii) of Theorem 1. Part i) follows from part ii) by, essentially, flipping the cross-sectional and temporal parameters of the model. Hence, we omit the proof of part i) to save space.

2.4 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 $d_i > \bar{x}\left(1-\bar{u}^{-1}\right)\left(1-\bar{v}^{-1}\right)$. Since $x\left(1-u_{2,x}^{-1}\right)\left(1-v_{2,x}^{-1}\right)$ is a continuous strictly increasing function of $x \geq \bar{x}$, there exists a small enough $\theta_1 > \bar{x}$ such that $d_i > \theta_1\left(1-u_{2,\theta_1}^{-1}\right)\left(1-v_{2,\theta_1}^{-1}\right)$ 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

$$\theta_1^{-1}\left(1-u_{2,\theta_1}^{-1}\right)^{-1} d_i + v_{2,\theta_1}^{-1} > 1 \text{ for all } i \leq q \tag{39}$$

$$\theta_1^{-1}\left(1-u_{2,\theta_1}^{-1}\right)^{-1} d_i + v_{2,\theta_1}^{-1} < 1 \text{ for all } q < i \leq k. \tag{40}$$

Note that $x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1} d_i + v_{2,x}^{-1}$ is the probability limit of the i-th largest eigenvalue of $M^{(1)}(x)$ as $n \to \infty$. Functions $g_i(x) \equiv x^{-1}\left(1-u_{2,x}^{-1}\right)^{-1} d_i + v_{2,x}^{-1}$, $i = 1, ..., k$ 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 (39-40), we conclude that equations $g_i(x) = 1$ have unique solutions $x = x_i$ for $i \leq q$ and $x > \theta_1$, and no solutions for $i > q$ and $x > \theta_1$. Note that $x_1 > x_2 > ... > x_q > \theta_1$.

Let δ be a small positive number such that $\delta < x_q - \theta_1$, let $\delta_1 = \min_{i=1,...,q;j=1,...,k} \left|g_j\left(x_i \pm \delta\right)\right|$ and $\delta_2 = \min\left\{\left|g_q\left(\theta_1\right)\right|, \left|g_{q+1}\left(\theta_1\right)\right|\right\}$. Further, let $\Phi_i^{\pm}$ denote the events that $\max_{j=1,...,k}\left|\lambda_j\left(M^{(1)}\left(x_i \pm \delta\right)\right) - g_j\left(x_i \pm \delta\right)\right| < \delta_1$ and let Φ be the event that $\max_{j=1,...,k}\left|\lambda_j\left(M^{(1)}\left(\theta_1\right)\right) - g_j\left(\theta_1\right)\right| < \delta_2$. 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:

$$\begin{array}{c} \lambda_j\left(M^{(1)}\left(\theta_1\right)\right) > 1 \text{ if and only if } j \leq q \text{ and} \\ \lambda_j\left(M^{(1)}\left(\theta_1\right)\right) < 1 \text{ if and only if } j > q, \end{array} \tag{41}$$

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

$$\begin{array}{c} \lambda_j\left(M^{(1)}\left(x_i - \delta\right)\right) > 1 \text{ if and only if } j \leq i \text{ and} \\ \lambda_j\left(M^{(1)}\left(x_i + \delta\right)\right) < 1 \text{ if and only if } j > i \end{array} . \tag{42}$$

Now, if $\lambda_1(\Lambda) < \theta_1$ and Ψ 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 y_1 and y_2 such that $y_2 > y_1 \geq \theta_1$, matrix $M^1(y_1) - M^1(y_2) = \Psi'\left((y_1 I_n - \Lambda)^{-1} - (y_2 I_n - \Lambda)^{-1}\right)\Psi$ is positive definite (this follows from the fact that the eigenvalues of $(y_1 I_n - \Lambda)^{-1} - (y_2 I_n - \Lambda)^{-1}$ equal $\frac{y_2 - y_1}{(y_1 - \lambda_j(\Lambda))(y_2 - \lambda_j(\Lambda))}$, $j = 1,...,n$, and therefore, are positive). Note that, as was mentioned above, $\lambda_1(\Lambda) \xrightarrow{a.s.} \bar{x}$ and, as is easily verified, $\Psi'\Psi \xrightarrow{p} D + I_k$ so that with probability arbitrarily close to 1, $\lambda_1(\Lambda) < \theta_1$ and Ψ is full rank for large enough n. The strict monotonicity of $\lambda_j\left(M^1(x)\right), j = 1,...,k$ and inequalities (41) and (42) 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 $x_1,...,x_q$. Since δ was an arbitrary positive number, we conclude, using Lemma 2, that $x_1,...,x_q$ must be the probability limits of the first q eigenvalues of $\frac{1}{T}XX'$, which establishes the first probability limit of part iii) of Theorem 1.

Furthermore, as follows from above and from Lemma 2, for any $\theta_1 > \bar{x}$, there will be only q eigenvalues of $\frac{1}{T}XX'$ larger than θ_1 for large enough n. On the other hand, $\frac{1}{T}XX' = \Psi\Psi' + \Lambda$ so that the k-th eigenvalue of $\frac{1}{T}XX'$ cannot be smaller than the k-th eigenvalue of Λ, which converges to $\bar{x}$. Hence, the $q+1$-th,...,k-th eigenvalues of $\frac{1}{T}XX'$ converge to $\bar{x}$, which establishes the second probability limit of part iii) of Theorem 1.

2.5 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 $\frac{1}{T}XX'$ with $j \leq q$. Then, since, as has been just shown, $\mu_j \xrightarrow{p} x_j$ and since the probability limit of $M^{(1)}(x)$, $x^{-1}\left(1 - u_{2,x}^{-1}\right)^{-1} D + v_{2,x}^{-1} I_k$, is a continuous function of $x \geq \theta_1$, we have: $M^{(1)}(\mu_j) \xrightarrow{p} x_j^{-1}\left(1 - u_{2,x_j}^{-1}\right)^{-1} D + v_{2,x_j}^{-1} I_k$. 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. 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, let us denote r_{ux_j} as r_{uj} and r_{vx_j} as r_{vj}. By Lemma 9, and since $u_{2,x_j} = u_j$ and $v_{2,x_j} = v_j$, we have: the j,j-th element of $M^{(2)}(\mu_j)$ converges in probability to $d_j x_j^{-2}\left(1 - u_j^{-1}\right)^{-2}\left(1 + \frac{r_{vj}(1+r_{uj})}{(1-r_{uj}r_{vj})u_j}\right) + \frac{1+r_{uj}}{(1-r_{uj}r_{vj})x_j v_j}$ and the j-th column of $M^{(3)}(\mu_j)$ converges in probability to $e_j x_j^{-1}\left(1 - u_j^{-1}\right)^{-1} d_j^{1/2}$. Therefore, by Lemma 2 iii, the j-th column of $\hat{\alpha}$, where $j \leq q$, is proportional to e_j, which establishes the fact that $\operatorname{plim}\hat{\alpha}_{ij} = 0$ for

$j \leq q$ and $i \neq j$. For the coefficient of the proportionality $\hat{\alpha}_{jj}$, we have:

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

where the last equality follows from the fact that $d_j = x_j\left(1-u_j^{-1}\right)\left(1-v_j^{-1}\right)$. Continuing algebraic manipulations a little further, we get:

$$\begin{aligned}
\hat{\alpha}_{jj}^{-2} &= 1+\frac{\left(1+r_{uj}\right)\left(1+r_{vj}\right)}{\left(1-r_{uj}r_{vj}\right)u_j v_j}\left(\frac{v_j r_{vj}}{1+r_{vj}}+\frac{v_j\left(u_j-1\right)}{\left(1+r_{vj}\right)\left(v_j-1\right)}\right) \\
&= 1+\frac{\left(1+r_{uj}\right)\left(1+r_{vj}\right)}{\left(1-r_{uj}r_{vj}\right)u_j v_j}\left(v_j-\frac{v_j}{1+r_{vj}}\frac{v_j-u_j}{v_j-1}\right).
\end{aligned}$$

To establish the first probability limit of part ii) of Theorem 1, it remains to show that

$$c\left(m\left(x_j\right)+x_j m'\left(x_j\right)\right) = \frac{\left(1+r_{uj}\right)\left(1+r_{vj}\right)}{\left(1-r_{uj}r_{vj}\right)u_j v_j} \text{ and} \tag{43}$$

$$\frac{1+v_j m_B\left(v_j\right)}{m_B\left(v_j\right)+v_j m_B'\left(v_j\right)} = -\frac{v_j}{1+r_{vj}}. \tag{44}$$

The latter equality follows directly from the definition of r_{vj}. To establish (43), note that (44) implies that

$$1+r_{vj} = -v_j\frac{d}{dv}\log\left(-1-v_j m_B\left(v_j\right)\right). \tag{45}$$

Similarly, we can show that

$$1+r_{uj} = -u_j\frac{d}{du}\log\left(-1-u_j m_A\left(u_j\right)\right). \tag{46}$$

Now, recall that, according to Zhang (2006), $m\left(z\right)$, $u\left(z\right)$ and $v\left(z\right)$ solve the system

$$\begin{cases}
-zm(z)-1 = -u\left(z\right)m_A\left(u(z)\right)-1 \\
-zm\left(z\right)-1 = c^{-1}\left[-v\left(z\right)m_B\left(v\left(z\right)\right)-1\right] \\
-zm\left(z\right)-1 = c^{-1}\frac{z}{u(z)v(z)}
\end{cases}.$$

Taking logarithms of both sides of the three equations of the above system, differentiating with respect to z, and noting that $\frac{d}{dz}\log\left(-zm(z)-1\right)=\frac{m(z)+zm'(z)}{1+zm(z)}=-\left(m(z)+zm'(z)\right)\frac{cu(z)v(z)}{z}$, we get:

$$\begin{cases} -\left(m(z)+zm'(z)\right)\frac{cu(z)v(z)}{z} = \frac{d}{dz}u(z)\frac{d}{du}\log\left(-u\left(z\right)m_A\left(u(z)\right)-1\right) \\ -\left(m(z)+zm'(z)\right)\frac{cu(z)v(z)}{z} = \frac{d}{dz}v(z)\frac{d}{dv}\log\left(-v\left(z\right)m_B\left(v(z)\right)-1\right) \\ -\left(m(z)+zm'(z)\right)\frac{cu(z)v(z)}{z} = \frac{1}{z}-\frac{1}{u(z)}\frac{d}{dz}u(z)-\frac{1}{v(z)}\frac{d}{dz}v(z) \end{cases}.$$

Solving for $\frac{d}{dz}u(z)$ and $\frac{d}{dz}v(z)$ from the first two equations and dividing both sides of the third equation by $-\left(m(z)+zm'(z)\right)\frac{cu(z)v(z)}{z}$, we get:

$$\begin{aligned} 1 \;=\;& -\left[cu(z)v(z)\left(m(z)+zm'(z)\right)\right]^{-1}-\left[u(z)\frac{d}{du}\log\left(-u\left(z\right)m_A\left(u(z)\right)-1\right)\right]^{-1} \qquad (47)\\ & -\left[v(z)\frac{d}{dv}\log\left(-v\left(z\right)m_B\left(v(z)\right)-1\right)\right]^{-1}. \end{aligned}$$

As was mentioned above, $m\left(z\right)$, $u(z)$ and $v(z)$ can be analytically continued from the complex area $\operatorname{Im} z>0$ to the real segment $z\in(\bar{x},\infty)$ so that $u(x_j)=u_j$ and $v(x_j)=v_j$. Substituting $z=x_j$ in (47), and using (45) and (46), we obtain:

$$1=-\left[cu_jv_j\left(m(x_j)+x_jm'(x_j)\right)\right]^{-1}+\frac{1}{1+r_{uj}}+\frac{1}{1+r_{vj}},$$

which implies (43).

Now, let us prove that $\operatorname{plim}\hat{\alpha}_{ij}=0$ for $j>q$. We no longer assume that (4) is satisfied (this assumption was innocuous for the proof of the first convergence statement as have been explained above). Let $y_1,y_2,...,y_k$ be the unit-length eigenvectors of $\frac{1}{T}XX'$ corresponding to the k of the largest eigenvalues. Note that $\hat{\alpha}_{ij}^2=y_{ij}^2$, where y_{ij} is the i-th component of y_j. Define $Q_j=\sum_{r=1}^{q}y_r'\left(\frac{1}{T}XX'+\varkappa e_ie_i'\right)y_r+y_j'\left(\frac{1}{T}XX'+\varkappa e_ie_i'\right)y_j$, where $\varkappa$ is an arbitrary positive number. Since $y_1,...,y_k$ are orthonormal,

$$Q_j\leq\sum\nolimits_{r=1}^{q+1}\lambda_r\left(\frac{1}{T}XX'+\varkappa e_ie_i'\right). \qquad (48)$$

Consider first the case when $j>q$ and $i\leq q$. If $\varkappa_1$ is so small that, for any $0<\varkappa<\varkappa_1$, the smallest eigenvalue of $\bar{x}^{-1}\left(1-\bar{u}^{-1}\right)^{-1}\begin{pmatrix} d_i & \sqrt{\varkappa d_i}\\ \sqrt{\varkappa d_i} & \varkappa\end{pmatrix}-\begin{pmatrix}\bar{v}^{-1} & 0\\ 0 & 0\end{pmatrix}$ is less than 1, then eigenvalues $\lambda_r\left(\frac{1}{T}XX'+\varkappa e_ie_i'\right)$ converge to x_r for $r\leq q$ and $r\neq i$, and to $\bar{x}$ for $q<r\leq k$. For the i-th eigenvalue of $\frac{1}{T}XX'+\varkappa e_ie_i'$, 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 $\left|\lambda_i\left(\frac{1}{T}XX'+\varkappa e_ie_i'\right)-\lambda_i\left(\frac{1}{T}XX'\right)-\varkappa y_{ii}^2\right|<$

$C\varkappa^2$, for all $n > N$ and $\varkappa < \varkappa_0$ with probability no smaller than $1-\delta$.

Such a behavior of the eigenvalues $\lambda_r\left(\frac{1}{T}XX' + \varkappa e_i e_i'\right)$, together with (48), imply that:

$$Q_j \leq \sum\nolimits_{r=1}^{q} x_r + \varkappa y_{ii}^2 + \bar{x} + \delta + C\varkappa^2 \tag{49}$$

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

$$Q_j \geq \sum\nolimits_{r=1}^{q} x_r + \varkappa y_{ii}^2 + \bar{x} + \varkappa y_{ij}^2 - \delta \tag{50}$$

with probability no smaller than $1-\delta$ for large enough n. Combining (49) and (50), we have: $\varkappa y_{ij}^2 \leq 2\delta + C\varkappa^2$ with probability no smaller than $1-2\delta$ for large enough n. Let us take $\delta = \varkappa^2$. Then, we have: $y_{ij}^2 \leq (2+C)\varkappa$ with probability no smaller than $1-2\varkappa^2$ for large enough n. Since $\varkappa$ is an arbitrary positive number, smaller than $\min(\varkappa_0, \varkappa_1)$, we have: $y_{ij}^2 \xrightarrow{p} 0$, and therefore, $\hat{\alpha}_{ij} \xrightarrow{p} 0$.

Now, let us consider the case when $j > q$ and $i > q$. If $\varkappa_2$ is so small that, for any $0 < \varkappa < \varkappa_2$, the largest eigenvalue of $\bar{x}^{-1}\left(1-\bar{u}^{-1}\right)^{-1}\begin{pmatrix} d_i & \sqrt{\varkappa d_i} \\ \sqrt{\varkappa d_i} & \varkappa \end{pmatrix} - \begin{pmatrix} \bar{v}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$ is less than 1, then eigenvalues $\lambda_r\left(\frac{1}{T}XX' + \varkappa e_i e_i'\right)$ converge to x_r for $r \leq q$ and to $\bar{x}$ for $q < r \leq k$. Therefore, we can replace inequalities (49) and (50) by $Q_j \leq \sum_{r=1}^{q} x_r + \bar{x} + \delta$ and $Q_j \geq \sum_{r=1}^{q} x_r + \bar{x} + \varkappa y_{ij}^2 - \delta$, respectively. So, with $\delta = \varkappa^2$, we have: $y_{ij}^2 \leq 2\varkappa$ with probability no smaller than $1-2\varkappa^2$ for large enough n. Since $\varkappa$ is an arbitrary positive number, smaller than $\varkappa_2$, we again have: $y_{ij}^2 \xrightarrow{p} 0$, and therefore, $\hat{\alpha}_{ij} \xrightarrow{p} 0$. This completes the proof.$\square$

3 Proof of Theorem 2

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. As before, we will consider only the case when $\mathrm{Var}\,\varepsilon_{it} \equiv \sigma^2 = 1$. The general-case formulae reported in Theorem 2 can be obtained from the formulae derived below by replacing $\hat{L}$ by $\hat{L}/\sigma$ and D by D/σ^2. We will use notation introduced in the proof of Theorem 1. In addition, 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.

3.1 A key lemma

Let $\Lambda = O'\tilde{\Lambda}O$ be a spectral decomposition of $\Lambda \equiv \frac{1}{T}\mathcal{A}_0\varepsilon_2\mathcal{B}^2\varepsilon_2'\mathcal{A}_0 = \frac{1}{T}\varepsilon_2\mathcal{B}^2\varepsilon_2'$, where the latter equality follows from the assumption of Theorem 2 ii) that $A = I_n$. Note that, since the entries of ε_2 are i.i.d. Gaussian, the spectral decomposition can be chosen so that O has the Haar invariant distribution (see Anderson (1984, p.536)).[3] Define $\tilde{X} = OX$ and $\tilde{\Psi} = O\Psi = O_{1:k}\left(L'L\right)^{1/2} + O\varepsilon_1/\sqrt{T}$. Then, matrix $\tilde{X}\tilde{X}'/T$ has a convenient representation $\tilde{X}\tilde{X}'/T = \tilde{\Psi}\tilde{\Psi}' + \tilde{\Lambda}$ and the same eigenvalues as matrix XX'/T.

Let y_{ij} denote the i-th component of an eigenvector of $\tilde{X}\tilde{X}'/T$, corresponding to eigenvalue $\lambda_j\left(XX'/T\right)$, and let $\tilde{\lambda}_i$ denote the i-th largest diagonal element of $\tilde{\Lambda}$. Let us define

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

The following Lemma is a straightforward consequence of Lemma 2:

Lemma 10: *Let $\mu \neq \tilde{\lambda}_i$, $i = 1, ..., n$ 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*

$$\lambda_m\left(M_n^{(1)}(x)\right) = 1, ..., \lambda_{m+s-1}\left(M_n^{(1)}(x)\right) = 1, \tag{51}$$

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

$$y = \left(v' M_n^{(2)}(\mu)\, v\right)^{-1/2}\left(\mu I_n - \tilde{\Lambda}\right)^{-1}\tilde{\Psi} v \tag{52}$$

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 $M_n^{(1)}(\mu)$, 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 $M_n^{(1)}(\mu)$, then the j-th column of matrix $\hat{\alpha}$ 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 Silverstein (1995), who generalized previous results of Yin (1986) and Marchenko and Pastur (1967). Silverstein

[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.

showed that the empirical distribution of the elements along the diagonal of $\tilde{\Lambda}$ defined as $G_{\tilde{\Lambda}}(x) \equiv \frac{1}{n}\sum_{i=1}^{n} 1\left\{\tilde{\lambda}_i \leq x\right\}$ almost surely converges to a non-random cumulative distribution function $\mathcal{G}$, which is fully characterized by the limiting spectral distribution $\mathcal{G}_B$ of matrix $B'B$. Silverstein's result was later generalized by Zhang (2006) to the case when $A \neq I_n$. It is this generalization that was used in the proof of Theorem 1.

To see the significance of Silverstein's 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 $\tilde{\Psi}_i^2$ with weights $\left(x - \tilde{\lambda}_i\right)^{-1}$. Now, by definition, $\tilde{\Psi}_i = O_{i,1}(L'L)^{1/2} + O_{i\cdot}\varepsilon_1/\sqrt{T}$. The second element in this sum is independent of the first and, by Assumption 2 i), is $N(0, 1/T)$. 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 $\frac{1}{n}(d_1 + c)\left(x - \tilde{\lambda}_i\right)^{-1}$. Intuitively, such a sum should converge to $(d_1 + c)\int (x - \lambda)^{-1} d\mathcal{G}(\lambda)$, 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.

3.2 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 $\varepsilon > 0$, $\sum_i E\left[X_{n,i}^2 I(|X_{n,i}| > \varepsilon)|\mathcal{F}_{n,i-1}\right] \xrightarrow{p} 0$. 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 $\left(\bar{l}, \bar{u}\right)$

containing the support of the distribution $\mathcal{G}$ and such that $\bar{l} < 0$. Note that the boundedness of the support of $\mathcal{G}_B$ implies the boundedness of the support of $\mathcal{G}$ (see Silverstein and Choi (1995) for an analysis of the support of $\mathcal{G}$). Further, let $\varsigma^{(n)}$ be an array of $n \times m$ matrices with i.i.d. standard normal entries independent of $\tilde{\lambda}_1, ..., \tilde{\lambda}_n$. 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 1, 2, 3 and 4 hold, and let $A = I_n$. Then, the joint distribution of random variables $\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_j(\tilde{\lambda}_i) \left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right); (j, s, t) \in \Theta_1 \right\}$ 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 $(1 + \delta_{st}) \int g_j(\lambda) g_{j_1}(\lambda) d\mathcal{G}(\lambda)$ when $(s, t) = (s_1, t_1)$.*

Proof of Lemma 13: Let real numbers l_1 and u_1 be such that $l_1 < 0$ and $[l_1, u_1]$ is included in $\left(\bar{l}, \bar{u}\right)$, but itself includes the support of the $\mathcal{G}$ law. Define functions $h_j(\lambda), j = 1, ..., J$, so that $h_j(\lambda) = g_j(\lambda)$ for $\lambda \in [l_1, u_1]$, and $h_j(\lambda) = 0$ otherwise. Note that $|h_j(\lambda)| < R$ for any $j = 1, ..., J$ and any λ, where R is a constant larger than $\max_{j=1,...,J} \sup_{\lambda \in [l_1, u_1]} |g_j(\lambda)|$. Note also that since, by Lemma 3 of Onatski (2009), almost surely for all large n, all $\tilde{\lambda}_i, i \leq n$ belong to $[l_1, u_1]$, $\Pr\left\{ \exists j \leq J, i \leq n \text{ such that } h_j(\tilde{\lambda}_i) \neq g_j(\tilde{\lambda}_i) \right\} \to 0$ as $n \to \infty$.

Consider random variables $X_{n,i} = \frac{1}{\sqrt{n}} \sum_{(j,s,t) \in \Theta_1} \gamma_{jst} h_j(\tilde{\lambda}_i) \left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right)$, where γ_{jst} are some constants. Let $\mathcal{F}_{n,i}$ be sigma-algebra generated by $\tilde{\lambda}_1, ..., \tilde{\lambda}_n$ 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 $\sum_{r=1}^{K} y_r z_r \leq \left(\sum_{r=1}^{K} (y_r)^p \right)^{1/p} \left(\sum_{r=1}^{K} (z_r)^q \right)^{1/q}$, which holds for $y_r > 0, z_r > 0, p > 1, q > 1$, and $(1/p) + (1/q) = 1$, take $y_r = \left| \frac{1}{\sqrt{n}} \gamma_{jst} h_j(\tilde{\lambda}_i) \left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right) \right|$, where (j, s, t) is the r-th triple in Θ_1, $z_r = 1$, and $p = 2 + \delta$ for some $\delta > 0$. Then, the inequality implies that $|X_{n,i}|^{2+\delta} \leq K^{1+\delta} R^{2+\delta} \sum_{(j,s,t) \in \Theta_1} \left| \gamma_{jst} \frac{\varsigma_{is}\varsigma_{it} - \delta_{st}}{\sqrt{n}} \right|^{2+\delta}$. Recalling that ς_{is} are i.i.d. $N(0, 1)$, 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

$$V_{n,n}^2 = \frac{1}{n} \sum_{i=1}^{n} E\Big(\sum_{\substack{(j,s,t) \in \Theta_1, \\ (j_1,s_1,t_1) \in \Theta_1}} \gamma_{jst}\gamma_{j_1 s_1 t_1} h_j(\tilde{\lambda}_i) h_{j_1}(\tilde{\lambda}_i) \left(\varsigma_{is}\varsigma_{it} - \delta_{st}\right) \left(\varsigma_{is_1}\varsigma_{it_1} - \delta_{s_1 t_1}\right) |\mathcal{F}_{n,i-1}\Big).$$

It is straightforward to check that the latter expression is equal to

$\sum_{\substack{1\leq j\leq J\\ 1\leq j_1\leq J}}\left[\left(\sum_{1\leq s\leq t\leq m}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\frac{1}{n}\sum_{i=1}^{n}h_j(\tilde{\lambda}_i)h_{j_1}(\tilde{\lambda}_i)\right]$.

Consider now $\tilde{V}^2_{n,n}=\sum_{\substack{1\leq j\leq J\\ 1\leq j_1\leq J}}\left[\left(\sum_{1\leq s\leq t\leq m}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\frac{1}{n}\sum_{i=1}^{n}g_j(\tilde{\lambda}_i)g_{j_1}(\tilde{\lambda}_i)\right]$. Since $P\left(\tilde{V}^2_{n,n}\neq V^2_{n,n}\right)\to 0$ as $n\to\infty$, $\tilde{V}^2_{n,n}$ and $V^2_{n,n}$ must converge in probability to the same limit, or must both diverge. But, by Theorem 1.1 of Silverstein (1995), the empirical distribution of $\tilde{\lambda}_1,...,\tilde{\lambda}_n$ almost surely weakly converges to $\mathcal{G}$. In addition, by Lemma 3 of Onatski (2009), almost surely for all large n, all $\tilde{\lambda}_i, i\leq n$ belong to $[l_1,u_1]$. Therefore, since by assumption, $g_j\left(\lambda\right)$ with $j=1,...,J$ are continuous on $[l_1,u_1]$, $\frac{1}{n}\sum_{i=1}^{n}g_j(\tilde{\lambda}_i)g_{j_1}(\tilde{\lambda}_i)-\int g_j(\lambda)g_{j_1}(\lambda)d\mathcal{G}(\lambda)$ converges in probability to zero, and we have:

$$\tilde{V}^2_{n,n}\xrightarrow{p}\Sigma\equiv\sum_{\substack{1\leq j\leq J\\ 1\leq j_1\leq J}}\left[\left(\sum_{1\leq s\leq t\leq m}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\int g_j(\lambda)g_{j_1}(\lambda)d\mathcal{G}(\lambda)\right]. \qquad (53)$$

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

Let us now define $Y_{n,i}=\sum_{(j,s,t)\in\Theta_1}\gamma_{jst}g_j(\tilde{\lambda}_i)\frac{\varsigma_{is}\varsigma_{it}-\delta_{st}}{\sqrt{n}}$. Since $\Pr\left(\sum_{i=1}^{n}Y_{n,i}\neq\sum_{i=1}^{n}X_{n,i}\right)\to 0$ 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 (53).□

Now let us formally establish the asymptotic behavior of $M_n^{(1)}(x)$, $M_n^{(2)}(x)$ and $M_n^{(3)}(x)$. By Lemma3 of Onatski (2009), for any fixed k, $\tilde{\lambda}_1,...,\tilde{\lambda}_k$ almost surely converge to $\bar{x}$, the upper boundary of support of $\mathcal{G}$. This result implies that, with high probability, $M_n^{(1)}\left(x\right)$ 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 $\theta_2>\theta_1>\bar{x}$. Since the weak convergence in $C\left[\theta_1,\theta_2\right]$ is well-studied, it will be convenient to modify $M_n^{(1)}\left(x\right)$ 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 $h(x,\tilde{\lambda}_i)=\max\left(x-\tilde{\lambda}_i,\frac{\theta_1-\bar{x}}{2}\right)$ and let

$$\begin{aligned}\hat{M}_n^{(1)}\left(x\right) &\equiv \sum_{i=1}^{n}\frac{\tilde{\Psi}'_{i\cdot}\tilde{\Psi}_{i\cdot}}{h\left(x,\tilde{\lambda}_i\right)},\\ \hat{M}_n^{(2)}(x) &= \sum_{i=1}^{n}\frac{\tilde{\Psi}'_{i\cdot}\tilde{\Psi}_{i\cdot}}{h^2(x,\tilde{\lambda}_i)} \text{ and}\\ \hat{M}_n^{(3)}(x) &= \sum_{i=1}^{n}\frac{O'_{i,1:k}\tilde{\Psi}_{i\cdot}}{h(x,\tilde{\lambda}_i)}.\end{aligned}$$

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 $\Pr\left(M_n^{(j)}(x) = \hat{M}_n^{(j)}(x), \forall x \in [\theta_1, \theta_2]\right) = \Pr\left(\tilde{\lambda}_1 < \frac{\theta_1 + \bar{x}}{2}\right) \to 1$ as $n \to \infty$.

Define

$$\begin{aligned}
M_0^{(1)}(x) &= (D + cI_k) \int \frac{d\mathcal{G}(\lambda)}{x - \lambda}, \\
M_0^{(2)}(x) &= (D + cI_k) \int \frac{d\mathcal{G}(\lambda)}{(x - \lambda)^2} \text{ and} \\
M_0^{(3)}(x) &= D^{1/2} \int \frac{d\mathcal{G}(\lambda)}{x - \lambda}.
\end{aligned}$$

We have the following

Lemma 14: *Let Assumptions 1,2,3 and 4 hold, and let $A = I_n$. 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:*

$$\left\{N_n^{(p)}(x), p = 1, 2, 3\right\} \xrightarrow{d} \left\{N^{(p)}(x), p = 1, 2, 3\right\}, \tag{54}$$

where, for any $\{y_1, ..., y_J\} \in [\theta_1, \theta_2]$, the joint distribution of entries of $\left\{N^{(p)}(y_j); p = 1, 2, 3, j = 1, ..., J\right\}$ is a $3Jk^2$-dimensional normal distribution with covariance between entry in row s and column t of $N^{(p)}(y_j)$ and entry in row s_1 and column t_1 of $N^{(r)}(y_{j_1})$ 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 $1 \le p_1 \le p_2 \le 2$,

$$\Omega^{(p_1,p_2)}(\tau, \tau_1) = \Omega^{(p_1,3)}(\tau, \tau_1) = \Omega^{(3,3)}(\tau, \tau_1) = 0$$

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

$$\begin{aligned}
\Omega^{(p_1,p_2)}(\tau, \tau_1) &= -(1 + \delta_{st}) d_s d_t \int \frac{d\mathcal{G}(\lambda)}{(y_j - \lambda)^{p_1}} \int \frac{d\mathcal{G}(\lambda)}{(y_{j_1} - \lambda)^{p_2}} \\
&\quad + \left[(1 + \delta_{st})\left(c^2 + d_s d_t\right) + c\left(d_s + d_t + 2\delta_{st}\sqrt{d_s d_t}\right)\right] \int \frac{d\mathcal{G}(\lambda)}{(y_j - \lambda)^{p_1}(y_{j_1} - \lambda)^{p_2}}, \\
\Omega^{(p_1,3)}(\tau, \tau_1) &= -(1 + \delta_{st})\sqrt{d_s} d_t \int \frac{d\mathcal{G}(\lambda)}{(y_j - \lambda)^{p_1}} \int \frac{d\mathcal{G}(\lambda)}{y_{j_1} - \lambda} \\
&\quad + \left[(1 + \delta_{st})\sqrt{d_s} d_t + c\left(\sqrt{d_s} + \delta_{st}\sqrt{d_t}\right)\right] \int \frac{d\mathcal{G}(\lambda)}{(y_j - \lambda)^{p_1}(y_{j_1} - \lambda)}, \\
\Omega^{(3,3)}(\tau, \tau_1) &= -(1 + \delta_{st}) d_t \int \frac{d\mathcal{G}(\lambda)}{y_j - \lambda} \int \frac{d\mathcal{G}(\lambda)}{y_{j_1} - \lambda} + ((1 + \delta_{st}) d_t + c) \int \frac{d\mathcal{G}(\lambda)}{(y_j - \lambda)(y_{j_1} - \lambda)}
\end{aligned}$$

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

$$\begin{aligned}\Omega^{(p_1,p_2)}(\tau,\tau_1) &= \Omega^{(p_1,p_2)}((t,s,j),(s_1,t_1,j_1)) \\ \Omega^{(p_1,3)}(\tau,\tau_1) &= \Omega^{(p_1,3)}((t,s,j),(s_1,t_1,j_1)) \\ \Omega^{(3,3)}(\tau,\tau_1) &= -(1+\delta_{st})\sqrt{d_s d_t}\int\frac{d\mathcal{G}(\lambda)}{y_j-\lambda}\int\frac{d\mathcal{G}(\lambda)}{y_{j_1}-\lambda} \\ &\quad+\left((1+\delta_{st})\sqrt{d_s d_t}+\delta_{st}c\right)\int\frac{d\mathcal{G}(\lambda)}{(y_j-\lambda)(y_{j_1}-\lambda)}\end{aligned}$$

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 $\left\{N_{n,st}^{(1)}(y_j), (s,t,j) \in \Theta\right\} \xrightarrow{d} \left\{N_{st}^{(1)}(y_j), (s,t,j) \in \Theta\right\}$, 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 $\tilde{\Psi}$ by $\xi(\xi'\xi)^{-1/2}$ and η, where ξ and η are two independent $n\times k$ matrix with i.i.d. standard normal entries independent from η and $\tilde{\lambda}_1, ..., \tilde{\lambda}_n$. Indeed, the substitution of $O\varepsilon_1$ by η is justified by the Assumption 2 i). 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 $N_n^{(1)}(x) = \sum_{v=1}^{8} S^{(v)}(x)$, where

$S^{(1)}(x) = (L'L)^{1/2}\left(\frac{\xi'\xi}{n}\right)^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\xi'_{i\cdot}\xi_{i\cdot}-I_k}{h(x,\tilde{\lambda}_i)}\right)\left(\frac{\xi'\xi}{n}\right)^{-1/2}(L'L)^{1/2}$,

$S^{(2)}(x) = (L'L)^{1/2}\sqrt{n}\left(I_k - \left(\frac{\xi'\xi}{n}\right)\right)\left(\frac{\xi'\xi}{n}\right)^{-1}(L'L)^{1/2}\sum_{i=1}^n \frac{1}{nh(x,\tilde{\lambda}_i)}$,

$S^{(3)}(x) = \sqrt{n}\,(L'L - D)\sum_{i=1}^n \frac{1}{nh(x,\tilde{\lambda}_i)}$,

$S^{(4)}(x) = \sqrt{\frac{n}{T}}\,(L'L)^{1/2}\left(\frac{\xi'\xi}{n}\right)^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\xi'_{i\cdot}\eta_{i\cdot}}{h(x,\tilde{\lambda}_i)}\right)$,

$S^{(5)}(x) = \sqrt{\frac{n}{T}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\eta'_{i\cdot}\xi_{i\cdot}}{h(x,\tilde{\lambda}_i)}\right)\left(\frac{\xi'\xi}{n}\right)^{-1/2}(L'L)^{1/2}$,

$S^{(6)}(x) = \left(\frac{n}{T}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\eta'_{i\cdot}\eta_{i\cdot}-I_k}{h(x,\tilde{\lambda}_i)}$,

$S^{(7)}(x) = \sqrt{n}\left(\frac{n}{T}-c\right)I_k\sum_{i=1}^n \frac{1}{nh(x,\tilde{\lambda}_i)}$,

$S^{(8)}(x) = -(D + cI_k)\sqrt{n}\left(\int\frac{d\mathcal{G}(\lambda)}{x-\lambda} - \sum_{i=1}^n \frac{1}{nh(x,\tilde{\lambda}_i)}\right)$

By Theorem 1 of Bai and Silverstein (2004), $\sqrt{n}\left(\int \frac{d\mathcal{G}^{(n)}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{n}\frac{1}{x-\tilde{\lambda}_i}\right)\xrightarrow{p}0$ for any $x\in[\theta_1,\theta_2]$, where $\mathcal{G}^{(n)}(\lambda)$ is defined as follows. Let $B^{(n,m)}=I_m\otimes B^{(n)}$ and let $\varepsilon^{(n,m)}$ be an $nm\times T^{(n)}m$ matrix with i.i.d. $N(0,1)$ elements. Then, $\mathcal{G}^{(n)}(\lambda)$ is the weak limit of the empirical eigenvalue distribution of $\frac{1}{T^{(n)}m}\varepsilon^{(n,m)}B^{(n,m)}B^{(n,m)\prime}\varepsilon^{(n,m)\prime}$ as m goes to infinity. Lemma 17 below shows that our Assumptions 1, 2, 3 and 4 imply that $\sqrt{n}\left(\int\frac{d\mathcal{G}^{(n)}(\lambda)}{x-\lambda}-\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}\right)\to 0$, and hence $\sqrt{n}\left(\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\tilde{\lambda}_i)}\right)\xrightarrow{p}0$. The latter convergence result together with the facts that $\xi'\xi/n\xrightarrow{p}I_k$, $L'L-D=o\left(n^{-1/2}\right)$, and $n/T-c=o\left(n^{-1/2}\right)$ imply that $\left\{\sum_{v=1}^{8}S_{st}^{(v)}(y_j);(s,t,j)\in\Theta\right\}$ and $\left\{\sum_{v=1}^{8}\tilde{S}_{st}^{(v)}(y_j);(s,t,j)\in\Theta\right\}$ weakly converge to the same limit or do not converge together, where

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

$\tilde{S}^{(2)}(x)=D^{1/2}\sqrt{n}\left(I_k-\left(\frac{\xi'\xi}{n}\right)\right)D^{1/2}\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}$,

$\tilde{S}^{(3)}(x)=0$,

$\tilde{S}^{(4)}(x)=\sqrt{c}D^{1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{i\cdot}'\eta_{i\cdot}}{h(x,\tilde{\lambda}_i)}\right)$,

$\tilde{S}^{(5)}(x)=\sqrt{c}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{i\cdot}'\xi_{i\cdot}}{h(x,\tilde{\lambda}_i)}\right)D^{1/2}$,

$\tilde{S}^{(6)}(x)=c\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{i\cdot}'\eta_{i\cdot}-I_k}{h(x,\tilde{\lambda}_i)}$,

$\tilde{S}^{(7)}(x)=\tilde{S}^{(8)}(x)=0$.

By definition, we have:

$$\begin{aligned}\sum_{v=1}^{8}\tilde{S}_{st}^{(v)}(y_j) &= \sqrt{d_sd_t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it}-\delta_{st}}{h(y_j,\tilde{\lambda}_i)}-\sqrt{d_sd_t}\int\frac{d\mathcal{G}(\lambda)}{y_j-\lambda}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it}-\delta_{st}}{\sqrt{n}}+\\ &\quad \sqrt{cd_s}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\eta_{it}}{h(y_j,\tilde{\lambda}_i)}+\sqrt{cd_t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{it}\eta_{is}}{h(y_j,\tilde{\lambda}_i)}+c\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{is}\eta_{it}-\delta_{st}}{h(y_j,\tilde{\lambda}_i)}.\end{aligned}$$

Since $[\xi,\eta]$ is an $n\times 2k$ matrix with i.i.d. standard normal entries, Lemma 13 and the above decomposition imply that $\left\{\sum_{v=1}^{8}\tilde{S}_{st}^{(v)}(y_j);(s,t,j)\in\Theta\right\}$ weakly converge to $\{Z_{stj},(s,t,j)\in\Theta\}$ having joint normal distribution such that $\operatorname{cov}\left(Z_{stj},Z_{s_1t_1j_1}\right)=0$ if $(s,t)\neq(s_1,t_1)$ and $(s,t)\neq(t_1,s_1)$ and $\operatorname{cov}\left(Z_{stj},Z_{s_1t_1j_1}\right)$ is equal to

$$\begin{aligned}\operatorname{cov}\left(Z_{stj},Z_{s_1t_1j_1}\right) &= \left[(1+\delta_{st})\left(c^2+d_sd_t\right)+c\left(d_s+d_t+2\delta_{st}\sqrt{d_sd_t}\right)\right]\times \\ &\quad\times\int\frac{d\mathcal{G}(\lambda)}{(y_j-\lambda)(y_{j_1}-\lambda)}-(1+\delta_{st})\,d_sd_t\int\frac{d\mathcal{G}(\lambda)}{y_j-\lambda}\int\frac{d\mathcal{G}(\lambda)}{y_{j_1}-\lambda}\end{aligned}\tag{55}$$

otherwise, which establishes the limit of the joint distribution of $\left\{N_{st}^{(1)}(y_j);(s,t,j)\in\Theta\right\}$.

Now we have to prove the tightness of all entries of $N_n^{(1)}(x)=\sum_{v=1}^{8}S^{(v)}(x)$. Since product and sum are continuous mappings from $C[\theta_1,\theta_2]^2$ to $C[\theta_1,\theta_2]$, it is enough to prove the tightness of every entry of each matrix entering definition of $S^{(v)}(x)$, $v=1,...,8$. The facts that $\xi'\xi/n \xrightarrow{p} I_k$, $L'L-D=o\left(n^{-1/2}\right)$, and $n/T-c=o\left(n^{-1/2}\right)$ imply the tightness of every entry of each of the matrices $(L'L)^{1/2}$, $\sqrt{n}\,(L'L-D)$, $\left(\frac{\xi'\xi}{n}\right)^{-1/2}$, $\left(\frac{\xi'\xi}{n}\right)^{-1}$, $\sqrt{\frac{n}{T}}I$, $\sqrt{n}\left(\frac{n}{T}-c\right)I$, 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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\xi_{it}-\delta_{st}}{h(x,\tilde{\lambda}_i)},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_{is}\eta_{it}}{h(x,\tilde{\lambda}_i)},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\eta_{is}\eta_{it}-\delta_{st}}{h(x,\tilde{\lambda}_i)}, \tag{56}$$

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

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 (56), 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 $\frac{E\left(\sum_{i=1}^{n}\left(h(y_1,\tilde{\lambda}_i)^{-1}-h(y_2,\tilde{\lambda}_i)^{-1}\right)(\xi_{is}\xi_{it}-\delta_{st})\right)^2}{n(y_1-y_2)^2}\leq E\left(\sum_{i=1}^{n}\left(h(y_1,\tilde{\lambda}_i)h(y_2,\tilde{\lambda}_i)\right)^{-1}(\xi_{is}\xi_{it}-\delta_{st})\right)^2/n\leq\frac{16}{n(\theta_1-\bar{x})^4}E\left(\sum_{i=1}^{n}(\xi_{is}\xi_{it}-\delta_{st})\right)^2=\frac{16}{(\theta_1-\bar{x})^4}(1+\delta_{st})$, where the first inequality follows from the fact that $\left|\frac{1}{h(y_1,\tilde{\lambda}_i)}-\frac{1}{h(y_2,\tilde{\lambda}_i)}\right|\leq\frac{|y_2-y_1|}{h(y_1,\tilde{\lambda}_i)h(y_2,\tilde{\lambda}_i)}$. Hence, $\sup_{n;y_1,y_2\in[\theta_1,\theta_2]}\frac{E\left(\sum_{i=1}^{n}\left(h(y_1,\tilde{\lambda}_i)^{-1}-h(y_2,\tilde{\lambda}_i)^{-1}\right)(\xi_{is}\xi_{it}-\delta_{st})\right)^2}{n(y_1-y_2)^2}$ 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 $N_n^{(2)}(x)$ is demonstrated), we also need to check Billingsley's moment condition when $h(\cdot,\cdot)$ is replaced by $h^2(\cdot,\cdot)$. We can use the above reasoning and inequality $\left|\frac{1}{h^2(y_1,\tilde{\lambda}_i)}-\frac{1}{h^2(y_2,\tilde{\lambda}_i)}\right|\leq\frac{|y_2-y_1|\left(h(y_1,\tilde{\lambda}_i)+h(y_2,\tilde{\lambda}_i)\right)}{h^2(y_1,\tilde{\lambda}_i)h^2(y_2,\tilde{\lambda}_i)}\leq\frac{32\theta_2|y_2-y_1|}{(\theta_1-\bar{x})^4}$ to perform such a check.

Similarly, conditions of Theorem 12.3 of Billingsley (1968) are satisfied for $\sum_{i=1}^{n}\frac{1}{nh(x,\tilde{\lambda}_i)}$. Condition i) is satisfied because, as has been shown above, $\sqrt{n}\left(\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\tilde{\lambda}_i)}\right)\xrightarrow{p}0$ for any $x\in\left[\theta_1,\theta_2\right]$. Condition ii) is satisfied because $E\left(\sum_{i=1}^{n}\frac{1}{nh(y_1,\tilde{\lambda}_i)h(y_2,\tilde{\lambda}_i)}\right)^2\leq\frac{16}{(\theta_1-\bar{x})^4}$ for any $y_1,y_2\in\left[\theta_1,\theta_2\right]$.

To prove the tightness of $\sqrt{n}\left(\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\tilde{\lambda}_i)}\right)$, 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 $n\left(\int\frac{d\mathcal{G}^{(n)}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{n}\frac{1}{x-\tilde{\lambda}_i}\right)$, form a tight sequence. But by Lemma 17, $\sup_{x\in[\theta_1,\theta_2]}\sqrt{n}\left(\int\frac{d\mathcal{G}^{(n)}(\lambda)}{x-\lambda}-\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}\right)\xrightarrow{p}0$. Therefore, $\sqrt{n}\left(\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{n}\frac{1}{x-\tilde{\lambda}_i}\right)$ is tight too. Finally, the latter tightness and the fact that $\Pr\left\{\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\left(\frac{1}{x-\tilde{\lambda}_i}-\frac{1}{h(x,\tilde{\lambda}_i)}\right)\neq 0\right\}\to 0$ imply that sequence $\sqrt{n}\left(\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}-\sum_{i=1}^{n}\frac{1}{nh(x,\tilde{\lambda}_i)}\right)$ 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 $R(z,\varkappa)=R(z)+\sum_{n=1}^{\infty}(-\varkappa)^nR(z)\left(A^{(1)}R(z)\right)^n$ (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 for the resolvent in this formula, we obtain

$$P_j(\varkappa)=P_j-\frac{1}{2\pi i}\sum_{n=1}^{\infty}(-\varkappa)^n\int_\Gamma R(z)\left(A^{(1)}R(z)\right)^n dz \tag{57}$$

where $P_j\equiv P_j(0)$ and the series absolutely converges for $|\varkappa|<\frac{r_0}{\left\|A^{(1)}\right\|}$. Kato (1980), page 76, shows that $\frac{1}{2\pi i}\int_\Gamma R(z)A^{(1)}R(z)\,dz=-P_jA^{(1)}S_j-S_jA^{(1)}P_j$. This equality and (57) im-

[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 $r_0/\left\|A^{(1)}\right\|$.

ply that $P_j(\varkappa) = P_j - \varkappa\left(P_jA^{(1)}S_j - S_jA^{(1)}P_j\right) - \frac{1}{2\pi i}\sum_{n=2}^{\infty}(-\varkappa)^n \int_\Gamma R(z)\left(A^{(1)}R(z)\right)^n dz$. Therefore, we have:

$$\left\|\frac{1}{\varkappa}\left(P_j(\varkappa) - P_j\right) + P_jA^{(1)}S_j + S_jA^{(1)}P_j\right\| \le \frac{|\varkappa|\left\|A^{(1)}\right\|^2}{r_0\left(r_0 - |\varkappa|\left\|A^{(1)}\right\|\right)} \tag{58}$$

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 $\frac{1}{\varkappa}\left(e_j(\varkappa) - e_j\right) + S_jA^{(1)}e_j = \left(\frac{1}{\varkappa}\left(P_j(\varkappa)e_j - e_j\right) + S_jA^{(1)}e_j\right) + \frac{1}{\varkappa}e_j(\varkappa)\left(1 - \|P_j(\varkappa)e_j\|\right)$. Using (58) and the fact that $S_je_j = 0$, for the first term on right hand side of the identity we have:

$$\left\|\frac{1}{\varkappa}\left(P_j(\varkappa)e_j - e_j\right) + S_jA^{(1)}e_j\right\| \le \frac{|\varkappa|\left\|A^{(1)}\right\|^2}{r_0\left(r_0 - |\varkappa|\left\|A^{(1)}\right\|\right)}. \tag{59}$$

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

$$\left\|\frac{1}{\varkappa}e_j(\varkappa)\left(1 - \|P_j(\varkappa)e_j\|\right)\right\| \le \frac{1}{|\varkappa|}\left(1 - \|P_j(\varkappa)e_j\|^2\right) = |\varkappa|\left\|\frac{1}{\varkappa}\left(P_j(\varkappa)e_j - e_j\right)\right\|^2. \tag{60}$$

But, form (59), $\left\|\frac{1}{\varkappa}\left(P_j(\varkappa)e_j - e_j\right)\right\|^2 \le 2\left\|S_jA^{(1)}e_j\right\|^2 + \frac{2|\varkappa|^2\left\|A^{(1)}\right\|^4}{r_0^2\left(r_0-|\varkappa|\left\|A^{(1)}\right\|\right)^2} \le \frac{\left\|A^{(1)}\right\|^2}{2r_0^2} + \frac{2|\varkappa|^2\left\|A^{(1)}\right\|^4}{r_0^2\left(r_0-|\varkappa|\left\|A^{(1)}\right\|\right)^2}$. Combining the above identity, (59), (60), and the latter inequality, we obtain: $\left\|\frac{1}{\varkappa}\left(e_j(\varkappa) - e_j\right) + S_jA^{(1)}e_j\right\| \le \frac{|\varkappa|\left\|A^{(1)}\right\|^2\left(3r_0^2 - 4r_0|\varkappa|\left\|A^{(1)}\right\| + 5|\varkappa|^2\left\|A^{(1)}\right\|^2\right)}{2r_0^2\left(r_0-|\varkappa|\left\|A^{(1)}\right\|\right)^2} \le \frac{2|\varkappa|\left\|A^{(1)}\right\|^2}{\left(r_0-|\varkappa|\left\|A^{(1)}\right\|\right)^2}$, 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 $K \subset \{f : |f(\theta_1)| \le r\}$ for large enough r and $K \subset \{f : w_f(\delta(\varepsilon_1)) \le \varepsilon_1\}$ for small enough $\delta(\varepsilon_1)$, where $w_f(\delta)$ is the modulus of continuity of function f, defined as $w_f(\delta) = \sup_{|s-t|\le\delta}|f(s) - f(t)|$, $0 < \delta \le \theta_2 - \theta_1$. Let us choose $N(\varepsilon, \varepsilon_1)$ so that for any

$n > N(\varepsilon, \varepsilon_1)$, $\Pr(|x_n - x_0| > \delta(\varepsilon_1)) < \frac{\varepsilon}{2}$. Then, for $n > N(\varepsilon, \varepsilon_1)$, we have:

$$\begin{aligned}\Pr(|f_n(x_n) - f_n(x_0)| > \varepsilon_1) &= \Pr(|f_n(x_n) - f_n(x_0)| > \varepsilon_1 \text{ and } |x_n - x_0| \leq \delta(\varepsilon_1)) \\ &\quad + \Pr(|f_n(x_n) - f_n(x_0)| > \varepsilon_1 \text{ and } |x_n - x_0| > \delta(\varepsilon_1)) \\ &\leq \Pr(f_n(x) \notin K) + \Pr(|x_n - x_0| > \delta(\varepsilon_1)) < \varepsilon,\end{aligned}$$

which proves the lemma.□

Lemma 17: *Let (a, b) be such that $a < 0$ and $b > \bar{x}$. Suppose that $g(x, \lambda)$ is a continuous function on $(x, \lambda) \in [\theta_1, \theta_2] \times (a, b)$ and that it is monotone increasing and has bounded derivative with respect to λ on $\lambda \in (a, b)$ for any $x \in [\theta_1, \theta_2]$. Then, under Assumptions 1, 2, 3 and 4, as $n \to \infty$*

$$\sup_{x \in [\theta_1, \theta_2]} \sqrt{n} \left| \int g(x, \lambda) \, d\mathcal{G}(\lambda) - \int g(x, \lambda) \, d\mathcal{G}^{(n)}(\lambda) \right| \to 0,$$

where $\mathcal{G}^{(n)}$ is as defined in the proof of Lemma 14.

Proof of Lemma 17: As has been shown in the proof of Lemma 3 of Onatski (2009), the upper boundary of the support of $\mathcal{G}^{(n)}$ converges to $\bar{x}$ as $n \to \infty$. Therefore, (a, b) contains the supports of $\mathcal{G}^{(n)}$ for all large enough n. Hence, the function

$$\Delta_n(x) \equiv \sqrt{n} \left| \int g(x, \lambda) \, d\mathcal{G}(\lambda) - \int g(x, \lambda) \, d\mathcal{G}^{(n)}(\lambda) \right|$$

is continuous on $x \in [\theta_1, \theta_2]$, and it is enough to show the convergence of $\Delta_n(x)$ to zero pointwise for any $x \in [\theta_1, \theta_2]$.

Let $\nu_{1,j}$ and $\nu_{2,j}$ be the j-th largest eigenvalues of $B^{(nm)} B^{(nm)\prime}$ and $I_m \otimes B^{(n)} B^{(n)\prime}$, respectively. Here $B^{(s)}$ denotes the s-th element in the sequence of matrices B (which introduces the temporal correlation to the idiosyncratic terms $e = A\varepsilon B$), satisfying our Assumptions 2, 3 and 4. Further, let $\varepsilon_{\cdot j} = (\varepsilon_{1,j}, ..., \varepsilon_{nm,j})'$ where $\{\varepsilon_{i,j}, i \in \mathbb{N} \text{ and } j \in \mathbb{N}\}$ are i.i.d. $N(0, 1)$ random variables. By Theorem 1.1 of Silverstein (1995), $\mathcal{G}(\lambda)$ and $\mathcal{G}^{(n)}(\lambda)$ are the weak limits of the empirical distributions of the eigenvalues μ_{1j} and μ_{2j} of $R_1 \equiv \frac{1}{T^{(nm)}} \sum_{j=1}^{T^{(nm)}} \nu_{1,j} \varepsilon_{\cdot j} \varepsilon_{\cdot j}'$ and $R_2 \equiv \frac{1}{T^{(n)}m} \sum_{j=1}^{T^{(n)}m} \nu_{2,j} \varepsilon_{\cdot j} \varepsilon_{\cdot j}'$, respectively, as $m \to \infty$, almost surely. Therefore, for any n, with probability 1, $|\Delta_n(x) - \Delta_{n,m}(x)| \to 0$ as $m \to \infty$, where

$$\Delta_{n,m}(x) \equiv \sqrt{n} \left| \frac{1}{nm} \sum_{j=1}^{nm} g\left(x, \mu_{1,j}\right) - \frac{1}{nm} \sum_{j=1}^{nm} g\left(x, \mu_{2,j}\right) \right|.$$

Hence, to establish the convergence $\Delta_n(x) \to 0$, it is enough to show that with positive probability, $\Delta_{n,m}(x) \to 0$ as $n \to \infty$ uniformly in $m \geq 1$.

We will prove the latter convergence in the case when $T^{(n)}m \leq T^{(nm)}$. The extension to the case when $T^{(n)}m$ is larger than $T^{(nm)}$ is straightforward. Let us define a set $J_{n,m}$ as the set of positive integers j, such that $j \leq T^{(n)}m$, and $\{\nu_{1,j}, \nu_{2,j}\} \in [\underline{x}_B, \bar{x}_B]$. Consider matrices $\tilde{R}_1 \equiv \frac{1}{T^{(nm)}} \sum_{j=1}^{T^{(nm)}} \tilde{\nu}_{1,j} \varepsilon_{\cdot j} \varepsilon_{\cdot j}'$ and $\tilde{R}_2 \equiv \frac{1}{T^{(n)}m} \sum_{j=1}^{T^{(n)}m} \tilde{\nu}_{2,j} \varepsilon_{\cdot j} \varepsilon_{\cdot j}'$, where $\tilde{\nu}_{1,j} = \tilde{\nu}_{2,j} = 0$ if $j \notin J_{n,m}$ and $\tilde{\nu}_{1,j} = \nu_{1,j}$ and $\tilde{\nu}_{2,j} = \nu_{2,j}$ otherwise. The ranks of $R_1 - \tilde{R}_1$ and of $R_2 - \tilde{R}_2$ are no larger than $o\left(n^{1/2}\right) m$, where $o\left(n^{1/2}\right)$ is uniform in $m \geq 1$. It is because the number of integers $j \leq T^{(nm)}$ such that $j \notin J_{n,m}$ is $o\left(n^{1/2}\right) m$.

Indeed, for any positive integer s, let us denote the empirical distribution of the eigenvalues of $B^{(s)}B^{(s)\prime}$ as $G_B^{(s)}(x)$, and let

$$l_s = \sup_x \left| G_B^{(s)}(x) - \mathcal{G}_B(x) \right|.$$

Then, the number of integers $j \leq T^{(nm)}$ such that $\nu_{1,j} \notin [\underline{x}_B, \bar{x}_B]$ cannot be larger than $2l_{nm}T^{(nm)}$. Similarly, the number of integers $j \leq T^{(n)}m$ such that $\nu_{2,j} \notin [\underline{x}_B, \bar{x}_B]$ cannot be larger than $2l_nT^{(n)}m$. Hence, the number of $j \leq T^{(nm)}$ such that $j \notin J_{n,m}$ cannot exceed $2l_{nm}T^{(nm)} + 2l_nT^{(n)}m + T^{(nm)} - T^{(n)}m$, which is $o\left(n^{1/2}\right) m$, as stated above, because $s/T^{(s)} - c = o\left(s^{-1/2}\right)$ and $l_s = o\left(s^{-1/2}\right)$ as $s \to \infty$.

The fact that the ranks of $R_1 - \tilde{R}_1$ and of $R_2 - \tilde{R}_2$ are no larger than $o\left(n^{1/2}\right) m$ together with Weyl's theorem (see Horn and Johnson, 1985, p.184) imply that the j-th largest eigenvalues of $\tilde{R}_1$ and $\tilde{R}_2$ are no smaller than the $j + o\left(n^{1/2}\right) m$-th largest eigenvalues of R_1 and R_2, respectively. Further, by construction, $R_1 - \tilde{R}_1$ and $R_2 - \tilde{R}_2$ are positive semi-definite matrices, and therefore, the j-th largest eigenvalues of $\tilde{R}_1$ and $\tilde{R}_2$ are no larger than the corresponding eigenvalues of R_1 and R_2. These eigenvalue bounds together with the fact that $g(x, \lambda)$ is monotone increasing imply that

$$\begin{aligned}
\sqrt{n} \left| \frac{1}{nm} \sum_{j=1}^{nm} g\left(x, \mu_{1,j}\right) - \frac{1}{nm} \sum_{j=1}^{nm} g\left(x, \tilde{\mu}_{1,j}\right) \right| &\leq \frac{1}{n^{1/2}m} \sum_{j=1}^{o\left(n^{1/2}\right)m} g\left(x, \mu_{1,j}\right) \text{ and} \\
\sqrt{n} \left| \int \frac{1}{nm} \sum_{j=1}^{nm} g\left(x, \mu_{2,j}\right) - \frac{1}{nm} \sum_{j=1}^{nm} g\left(x, \tilde{\mu}_{2,j}\right) \right| &\leq \frac{1}{n^{1/2}m} \sum_{j=1}^{o\left(n^{1/2}\right)m} g\left(x, \mu_{2,j}\right),
\end{aligned}$$

where $\tilde{\mu}_{1,j}$ and $\tilde{\mu}_{2,j}$ are the j-th largest eigenvalues of $\tilde{R}_1$ and $\tilde{R}_2$. Since $g(x, \lambda)$ is bounded, the right hand sides of the above two inequalities are $o(1)$ with probability 1 as $n \to \infty$, uniformly in $m \geq 1$. Therefore, to establish the lemma, we only need to show that, with positive probability,

$$\sqrt{n}\left|\frac{1}{nm}\sum_{j=1}^{nm}g\left(x,\tilde{\mu}_{1,j}\right)-\frac{1}{nm}\sum_{j=1}^{nm}g\left(x,\tilde{\mu}_{2,j}\right)\right|\to 0$$

as $n\to\infty$ uniformly in m. Since, by assumption, $g\left(x,\lambda\right)$ has bounded derivative, the latter convergence would follow from the fact that, with probability 1, $\max_{j=1,...,nm}\left|\tilde{\mu}_{1,j}-\tilde{\mu}_{2,j}\right|=o\left(n^{-1/2}\right)$ as $n\to\infty$, uniformly in $m\geq 1$.

Now, $\max_{j=1,...,nm}\left|\tilde{\mu}_{1,j}-\tilde{\mu}_{2,j}\right|\leq\frac{1}{T^{(n)}m}\left\|\sum_{j=1}^{T^{(n)}m}\varepsilon_{\cdot j}\varepsilon_{\cdot j}'\right\|\max_{j=1,...,T^{(n)}m}\left(\left|\frac{T^{(n)}m}{T^{(nm)}}\tilde{\nu}_{1,j}-\tilde{\nu}_{2,j}\right|\right).$ By Lemma 1, the term $\frac{1}{T^{(n)}m}\left\|\sum_{j=1}^{T^{(n)}m}\varepsilon_{\cdot j}\varepsilon_{\cdot j}'\right\|$ almost surely converges to $\left(1+\sqrt{c}\right)^2$ as $n\to\infty$. Hence, it is enough to prove that $\max_{j=1,...,T^{(n)}m}\left(\left|\frac{T^{(n)}m}{T^{(nm)}}\tilde{\nu}_{1,j}-\tilde{\nu}_{2,j}\right|\right)=o\left(n^{-1/2}\right)$ as $n\to\infty$, uniformly in $m\geq 1$. Further, since $n/T^{(n)}=c+o\left(n^{-1/2}\right)$, we have: $\frac{T^{(n)}m}{T^{(nm)}}=1+o\left(n^{-1/2}\right)$, and hence, the lemma will be proven if we show that $\max_{j=1,...,T^{(n)}m}\left(\left|\tilde{\nu}_{1,j}-\tilde{\nu}_{2,j}\right|\right)$ is $o\left(n^{-1/2}\right)$ as $n\to\infty$, uniformly in $m\geq 1$. Finally, for j which does not belong to $J_{n,m}$, $\left|\tilde{\nu}_{1,j}-\tilde{\nu}_{2,j}\right|=0$. Hence, we only need to show that $\max_{j\in J_{n,m}}\left(\left|\nu_{1,j}-\nu_{2,j}\right|\right)$ is $o\left(n^{-1/2}\right)$ as $n\to\infty$, uniformly in $m\geq 1$.

Let us assume that $\nu_{1,j}\geq\nu_{2,j}$. The analysis in the case when $\nu_{2,j}\geq\nu_{1,j}$ is similar. For $j\in J_{n,m}$, several cases are possible. First, both $\nu_{1,j}$ and $\nu_{2,j}$ may be equal to $\underline{x}_B$, or both $\nu_{1,j}$ and $\nu_{2,j}$ may be equal to $\bar{x}_B$ (the case when $\underline{x}_B=\bar{x}_B$ is not excluded). Such j would not contribute to $\max_{j\in J_{n,m}}\left(\left|\nu_{1,j}-\nu_{2,j}\right|\right)$, so we will not consider it. Next, in principle, it may be that $\nu_{1,j}=\bar{x}_B$ and $\nu_{2,j}=\underline{x}_B$, while $\underline{x}_B\neq\bar{x}_B$. Such a case is not possible asymptotically. Three other cases are as follows.

Case A: $\nu_{1,j}\in\left(\underline{x}_B,\bar{x}_B\right)$ and $\nu_{2,j}\in\left(\underline{x}_B,\bar{x}_B\right).$

Case B: $\nu_{1,j}=\bar{x}_B$ and $\nu_{2,j}\in\left(\underline{x}_B,\bar{x}_B\right);$

Case C: $\nu_{2,j}=\underline{x}_B$ and $\nu_{1,j}\in\left(\underline{x}_B,\bar{x}_B\right);$

Suppose that Case A holds. Let $\inf_{x\in\left(\underline{x}_B,\bar{x}_B\right)}\left(\frac{d\mathcal{G}_B(x)}{dx}\right)=\delta$, which, by Assumption 3, must be larger than zero. We have:

$$\begin{aligned}\left|\nu_{1,j}-\nu_{2,j}\right| &\leq \delta^{-1}\left|\mathcal{G}_B\left(\nu_{1,j}\right)-\mathcal{G}_B\left(\nu_{2,j}\right)\right| \\ &\leq \delta^{-1}\left|\mathcal{G}_B\left(\nu_{1,j}\right)-G_B^{(nm)}\left(\nu_{1,j}\right)\right|+\delta^{-1}\left|G_B^{(nm)}\left(\nu_{1,j}\right)-G_B^{(n)}\left(\nu_{2,j}\right)\right| \\ &\quad +\delta^{-1}\left|G_B^{(n)}\left(\nu_{2,j}\right)-\mathcal{G}_B\left(\nu_{2,j}\right)\right| \\ &\leq \delta^{-1}\left(l_{nm}+l_n\right)+\delta^{-1}\left|G_B^{(nm)}\left(\nu_{1,j}\right)-G_B^{(n)}\left(\nu_{2,j}\right)\right|. \end{aligned} \tag{61}$$

Note that if Case A holds, the multiplicity of eigenvalues $\nu_{1,j}$ and $\nu_{2,j}$ must be at most $o\left(n^{1/2}\right)m$. Otherwise $\sup_{x\in\left(\underline{x}_B,\bar{x}_B\right)}\left|G_B^{(nm)}\left(x\right)-\mathcal{G}_B(x)\right|\neq o\left(n^{-1/2}\right)$ and $\sup_{x\in\left(\underline{x}_B,\bar{x}_B\right)}\left|G_B^{(n)}\left(x\right)-\mathcal{G}_B(x)\right|\neq o\left(n^{-1/2}\right)$, which would violate Assumption 4. Therefore, we must have: $\left|G_B^{(nm)}(\nu_{1,j}) - 1 + \frac{j-1}{T^{(nm)}}\right| = o\left(n^{1/2}\right)$ and $\left|G_B^{(n)}(\nu_{1,j}) - 1 + \frac{j-1}{T^{(n)}m}\right| = o\left(n^{1/2}\right)$. Hence, inequality (61) implies that the maximum of $|\nu_{1,j} - \nu_{2,j}|$ over all j satisfying Case A is $o\left(n^{1/2}\right)$.

Suppose now that Case B holds. Then, since $\sup_x \left|G_B^{(nm)}(x) - \mathcal{G}_B(x)\right| = o\left(n^{-1/2}\right)m^{-1/2}$ and since, by definition, $G_B^{(nm)}(\nu_{1,j}) \le 1 - \frac{j-1}{T^{(nm)}}$, we must have:

$$j \le T^{(nm)}\left(1 - \lim_{x\uparrow\bar{x}_B} \mathcal{G}_B(x)\right) + o\left(n^{1/2}\right). \tag{62}$$

On the other hand, since $\nu_{2,j} < \bar{x}_B$ and $\sup_x \left|G_B^{(n)}(x) - \mathcal{G}_B(x)\right| = o\left(n^{-1/2}\right)$, it must be that:

$$j \ge T^{(n)}m\left(1 - \lim_{x\uparrow\bar{x}_B} \mathcal{G}_B(x)\right) + o\left(n^{1/2}\right)m. \tag{63}$$

Combining (62) and (63), and using the fact that $T^{(n)}m - T^{(nm)} = o\left(n^{1/2}\right)m$, we obtain:

$$j = T^{(n)}m\left(1 - \lim_{x\uparrow\bar{x}_B} \mathcal{G}_B(x)\right) + o\left(n^{1/2}\right)m. \tag{64}$$

Now, since $\nu_{2,j} \in (\underline{x}_B, \bar{x}_B)$, the multiplicity of the eigenvalue $\nu_{2,j}$ must be at most $o\left(n^{1/2}\right)m$ as in Case A, and therefore, (64) and the definition of $G_B^{(n)}(\nu_{2,j})$ imply:

$$G_B^{(n)}(\nu_{2,j}) = \lim_{x\uparrow\bar{x}_B} \mathcal{G}_B(x) + o\left(n^{-1/2}\right). \tag{65}$$

Finally, we have:

$$\begin{aligned}
|\nu_{1,j} - \nu_{2,j}| &\le \delta^{-1}\left|\lim_{x\uparrow\bar{x}_B} \mathcal{G}_B(x) - \mathcal{G}_B(\nu_{2,j})\right| \\
&\le \delta^{-1}\left|\lim_{x\uparrow\bar{x}_B} \mathcal{G}_B(x) - G_B^{(n)}(\nu_{2,j})\right| + \delta^{-1}\left|G_B^{(n)}(\nu_{2,j}) - \mathcal{G}_B(\nu_{2,j})\right| \\
&= \delta^{-1}o\left(n^{-1/2}\right) + \delta^{-1}l_n = o\left(n^{-1/2}\right).
\end{aligned}$$

Case C is analyzed similarly to Case B.□

3.3 Proof of Theorem 2 ii)

Let us first prove the first convergence statement of Theorem 2 ii). By definition, $\hat{L}'_{1:q}\hat{L}_{1:q}$ is a diagonal matrix with the diagonal elements equal to the first q eigenvalues of XX'/T. By Lemma 10, the j-th largest of these eigenvalues must solve $\lambda_j\left(M_n^{(1)}(x)\right) = 1$. Recall that, by Theorem 1, for any $j \le q$, plim $\hat{L}'_j\hat{L}_j$ must be equal to $x_j > \bar{x}$. Let us fix θ_1 and θ_2 so that $\theta_2 > x_1$ and $\bar{x} < \theta_1 < x_q$. Since, by Lemma 14, $M_n^{(1)}(x)$, considered as a random element of $C^{k^2}[\theta_1,\theta_2]$, weakly converges to $M_0^{(1)}(x)$, the solution to $\lambda_j\left(M_n^{(1)}(x)\right)=1$ must converge in probability to the solution of $\lambda_j\left(M_0^{(1)}(x)\right)=1$ by continuous mapping theorem. We conclude that, for any $j\le q$, x_j must be a solution to $\lambda_j\left(M_0^{(1)}(x)\right)=1$ on $[\theta_1,\theta_2]$. But $\lambda_j\left(M_0^{(1)}(x)\right)=(d_j+c)\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}$ by definition.Hence, x_j must be a solution to $(d_j+c)\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}=1$ on $[\theta_1,\theta_2]$. For $x>\bar{x}$, function $\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}$ is a decreasing function of x, so that x_j is the only solution of $(d_j+c)\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}=1$ on $x>\bar{x}$, and hence, the largest solution to $(d_j+c)\int\frac{d\mathcal{G}(\lambda)}{x-\lambda}=1$ on $x\in\mathbb{R}$, as stated in Theorem 2 ii).

Let us denote the solution of $\lambda_j\left(\hat{M}_n^{(1)}(x)\right)=1$ on $x\in[\theta_1,\theta_2]$ as x_{nj}.[6] Since the probability that $\hat{M}_n^{(1)}(x)\neq M_n^{(1)}(x)$ for some $x\in[\theta_1,\theta_2]$ converges to zero as $n\to\infty$, and since $M_n^{(1)}(x)$ converges to $M_0^{(1)}(x)$ as $n\to\infty$, such a solution exists for $j\le q$ and converges in probability to x_j as $n\to\infty$. Moreover, with probability arbitrarily close to 1, x_{nj} is identically equal to $\hat{L}_j'\hat{L}_j$ (by Lemma 10) for large enough n. Therefore, by Lemmas 10 and 14, the asymptotic distribution of $\hat{L}_j'\hat{L}_j$ around its probability limit must be the same as that of x_{nj} around x_j.

Lemma 14 and part i of Lemma 15 imply that, for any $j\le q$,

$$\lambda_j(\hat{M}_n^{(1)}(x))=\lambda_j\left(M_0^{(1)}(x)\right)+\frac{1}{\sqrt{n}}N_{n,jj}^{(1)}(x)+o_p\left(\frac{1}{\sqrt{n}}\right),\tag{66}$$

where $o_p\left(\frac{1}{\sqrt{n}}\right)$ is understood as a random element of $C[\theta_1,\theta_2]$, which, when multiplied by $\sqrt{n}$, tends in probability to zero as $n\to\infty$.

Now, let us define function $\nu_j(y)$ for $y>0$ so that it is equal to $\bar{x}$ if $y>\lim_{x\downarrow\bar{x}}\lambda_j(M_0^{(1)}(x))$ and to the inverse function to function $\lambda_j(M_0^{(1)}(x))$ otherwise. Since $\frac{d}{dx}\lambda_j(M_0^{(1)}(x))=-(d_j+c)\int\frac{d\mathcal{G}(\lambda)}{(x-\lambda)^2}$, Silverstein and Choi's (1995) result that the density of $\mathcal{G}$ has form $f(\lambda)=\mathit{const}\cdot(\bar{x}-\lambda)^{1/2}(1+o(1))$ for $\lambda\to\bar{x}$ implies that $\lim_{x\downarrow\bar{x}}\frac{d}{dx}\lambda_j(M_0(x))=+\infty$, and, hence, $\nu_j(y)$ is continuously differentiable for $y>0$. Applying ν_j to both sides of (66) and using the first order Taylor expansion of the right hand side, we have for $x\in[\theta_1,\theta_2]$: $\nu_j\left(\lambda_j(\hat{M}_n^{(1)}(x))\right)=x+\nu_j'(\tau_n(x))\frac{1}{\sqrt{n}}N_{n,jj}^{(1)}(x)+o_p\left(\frac{1}{\sqrt{n}}\right)$, where $\tau_n(x)$ is a random element of $C[\theta_1,\theta_2]$ such that $\tau_n(x)\xrightarrow{p}\lambda_j(M_0^{(1)}(x))$ as $n\to\infty$. Substituting x by x_{nj} in the above expansion of $\nu_j\left(\lambda_j(\hat{M}_n^{(1)}(x))\right)$ and using the facts that $\lambda_j(\hat{M}_n^{(1)}(x_{nj}))=1$, that $\nu_j(1)=x_j$, and that $x_{nj}=\hat{L}_j'\hat{L}_j$ with probability arbitrarily close to 1 for large enough n, we obtain: $\sqrt{n}\left(\hat{L}_j'\hat{L}_j-x_j\right)=-\nu_j'(\tau_n(x_{nj}))N_{n,jj}^{(1)}(x_{nj})+o_p(1)$.

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

Further, since $x_{nj} \xrightarrow{p} x_j$ and $\lambda_j(M_0^{(1)}(x_j)) = 1$, we have: $\nu_j'(\tau_n(x_{nj})) \xrightarrow{p} \nu_j'(1)$. Finally, $N_{n,jj}^{(1)}(x_{nj}) - N_{n,jj}^{(1)}(x_j) \xrightarrow{p} 0$, which follows from Lemma 14 and Lemma 16. Therefore, $\sqrt{n}\left(\hat{L}_j'\hat{L}_j - x_j\right)$ has the following form

$$\sqrt{n}\left(\hat{L}_j'\hat{L}_j - x_j\right) = -\nu_j'(1)\, N_{n,jj}^{(1)}(x_j) + o_p(1). \tag{67}$$

Finally, by definition, $\nu_j'(1) = \left(\lambda_j'\left(M_0^{(1)}(x_j)\right)\right)^{-1} = \left(-(d_j + c)\int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}\right)^{-1}$. The asymptotic normality of $\hat{L}_j'\hat{L}_j$ and the form of its asymptotic variance Ω_{jj} stated in Theorem 2 ii) now follow from (67) and Lemma 14.

Let us turn to the proof of the second convergence statement of Theorem 2 ii). By Lemma 10 and by the definitions of $\hat{M}_n^{(1)}(x)$, $\hat{M}_n^{(2)}(x)$ and $\hat{M}_n^{(3)}(x)$, for any $j \leq q$, the j-th column of $\hat{\alpha}$ equals

$$\hat{\alpha}_{\cdot j} = \left(w_{nj}'\hat{M}_n^{(2)}(x_{nj})\, w_{nj}\right)^{-1/2} \hat{M}_n^{(3)}(x_{nj})\, w_{nj}, \tag{68}$$

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 $x_{nj} \xrightarrow{p} x_j$ imply that $\hat{M}_n^{(3)}(x_{nj}) \xrightarrow{p} M_0^{(3)}(x_j) \equiv D^{1/2}\displaystyle\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda}$ and $\hat{M}_n^{(2)}(x_{nj}) \xrightarrow{p}$ $(D + cI_k)\displaystyle\int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}$. Therefore, by (68), we get: $\hat{\alpha}_{\cdot j} \xrightarrow{p} d_j^{1/2}\displaystyle\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda}\left((d_j + c)\int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}\right)^{-1/2} e_j$, which establishes the form of $\operatorname{plim}\hat{\alpha}_{jj}$ stated in Theorem 2 ii).

Now, we will study the asymptotic behavior of $\hat{\alpha}_{1:q}$ around its probability limit. Let us denote $(d_j + c)\displaystyle\int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}$ as ρ_j. Representation (68) and the facts that $w_{nj}'\hat{M}_n^{(2)}(x_{nj})\, w_{nj} \xrightarrow{p} \rho_j$ and $w_{nj} \xrightarrow{p} e_j$ imply that, for any $j \leq q$, $\sqrt{n}\left(\hat{\alpha}_{\cdot j} - \operatorname{plim}\hat{\alpha}_{\cdot j}\right) = \sum_{s=1}^4 A_j^{(s)} + o_p(1)$, where
$A_j^{(1)} = \rho_j^{-1/2} N_n^{(3)}(x_{nj})\, e_j$,
$A_j^{(2)} = \rho_j^{-1/2}\sqrt{n}\left(D^{1/2}\int \frac{d\mathcal{G}(\lambda)}{x_{nj}-\lambda} - D^{1/2}\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda}\right) e_j$,
$A_j^{(3)} = \rho_j^{-1/2} D^{1/2}\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda}\sqrt{n}\left(w_{nj} - e_j\right)$,
$A_j^{(4)} = D^{1/2}\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} e_j \sqrt{n}\left(\left(w_{nj}'\hat{M}_n^{(2)}(x_{nj})\, w_{nj}\right)^{-1/2} - \rho_j^{-1/2}\right)$.

Using the Taylor expansion of function $x^{-1/2}$ around $x = \rho_j$, we get:

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

Therefore,

$$\begin{aligned}
A_j^{(4)} &= -\frac{1}{2}\rho_j^{-3/2} D^{1/2} \int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} e_j \sqrt{n}\left(w_{nj}' \hat{M}_n^{(2)}(x_{nj}) w_{nj} - \rho_j\right) + o_p(1) \\
&= -\frac{1}{2}\rho_j^{-3/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} e_j \Big[(w_{nj}+e_j)' \hat{M}_n^{(2)}(x_{nj}) \sqrt{n}(w_{nj}-e_j) \\
&\qquad + N_{n,jj}^{(2)}(x_{nj}) - 2(d_j+c) \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^3} \sqrt{n}(x_{nj}-x_j) \Big] + o_p(1),
\end{aligned}$$

where, to obtain the second equality, we used the Taylor expansion of $N_n^{(2)}(x_{nj})$ around x_j. Similarly, using the Taylor expansion of function $\int \frac{d\mathcal{G}(\lambda)}{x-\lambda}$ around $x = x_j$, we obtain:

$$A_j^{(2)} = -\rho_j^{-1/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2} \sqrt{n}(x_{nj}-x_j) e_j + o_p(1)$$

The formulae obtained for $A_j^{(4)}$ and $A_j^{(2)}$ and the facts that $(d_j+c)\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} = 1$, that $\hat{M}_n^{(2)}(x_{nj}) \xrightarrow{p} (D + cI_k)\int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}$, and that $w_{nj} \xrightarrow{p} e_j$, imply that we have the following representation $\sqrt{n}(\hat{\alpha}_{\cdot j} - \operatorname{plim}\hat{\alpha}_{\cdot j}) = \sum_{s=1}^4 \hat{A}_j^{(s)} + o_p(1)$, where
$\hat{A}_j^{(1)} = \rho_j^{-1/2} N_n^{(3)}(x_{nj}) e_j$,
$\hat{A}_j^{(2)} = -\frac{1}{2}\rho_j^{-3/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} e_j N_{n,jj}^{(2)}(x_{nj})$,
$\hat{A}_j^{(3)} = \left(\rho_j^{-3/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^3} - \rho_j^{-1/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}\right) e_j \sqrt{n}(x_{nj}-x_j)$,
$\hat{A}_j^{(4)} = \left(\rho_j^{-1/2} D^{1/2} \int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} - \rho_j^{-3/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2} e_j e_j'\right) \sqrt{n}(w_{nj}-e_j)$.

Statement ii) of Lemma 15 and Lemma 14 imply that

$$\sqrt{n}(w_{nj}-e_j) = -\tilde{S}\left(\int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda}\right)^{-1} N_n^{(1)}(x_{nj}) e_j + o_p(1), \tag{69}$$

where $\tilde{S} = \operatorname{diag}\left((d_1-d_j)^{-1}, ..., \underbrace{0}_{j\text{-th position}}, ..., (d_k-d_j)^{-1}\right)$. Further, by the definition of x_{nj} and by (67),

$$\sqrt{n}(x_{nj}-x_j) = \rho_j^{-1} N_{n,jj}^{(1)}(x_j) + o_p(1). \tag{70}$$

Now, formulas (69) and (70), the definitions of $\hat{A}_j^{(s)}$, the fact that $x_{nj} \xrightarrow{p} x_j$ and Lemma 16 imply that we have the following final representation $\sqrt{n}(\hat{\alpha}_{\cdot j} - \operatorname{plim}\hat{\alpha}_{\cdot j}) = \sum_{s=1}^3 \tilde{A}_j^{(s)} + o_p(1)$, where
$\tilde{A}_j^{(1)} = \rho_j^{-1/2} N_n^{(3)}(x_j) e_j$,
$\hat{A}_j^{(2)} = -\frac{1}{2}\rho_j^{-3/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{x_j-\lambda} e_j N_{n,jj}^{(2)}(x_j)$,
$\hat{A}_j^{(3)} = \left(\rho_j^{-5/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^3} - \rho_j^{-3/2} d_j^{1/2} \int \frac{d\mathcal{G}(\lambda)}{(x_j-\lambda)^2}\right) e_j N_{n,jj}^{(1)}(x_j) - \rho_j^{-1/2} D^{1/2} \tilde{S} N_n^{(1)}(x_j) e_j$.

Using Lemma 14, we conclude that the joint asymptotic distribution of the elements of $\sqrt{n}\left(\hat{\alpha}_{1:q} - \text{plim}\,\hat{\alpha}_{1:q}\right)$ is Gaussian. The explicit expressions for the elements of the asymptotic covariance matrix follow[7] from the above definitions of $\tilde{A}_j^{(s)}, s = 1, ..., 3,$ from the expressions for the covariance of $N_n^{(1)}(x_j), N_n^{(2)}(x_j),$ and $N_n^{(3)}(x_j), j = 1, ..., q$ summarized in the definition of $\Omega^{(\cdot,\cdot)}$ given in Lemma 14, and from the fact that $(d_j + c)\int \frac{d\mathcal{G}(\lambda)}{x_j - \lambda} = 1$ for any $j = 1, ..., q$.$\square$

4 Proof of Theorem 3

In the proof of Theorem 3, we will not make the assumption, made above, that $\sigma^2 = 1$. Also, we will denote $\lambda_j(XX'/T)$ as λ_j. First, let us show that $\hat{c}$, $\hat{\sigma}^2$, $\hat{m}_i(r)$, $\hat{m}_{is}(1,1)$, $\widehat{\tilde{m}}_i(r)$ and $\widehat{\tilde{m}}_{is}(1,1)$ are consistent estimators of c, σ^2, $m_i(r)$, $m_{is}(1,1)$, $\tilde{m}_i(r)$ and $\tilde{m}_{is}(1,1)$, respectively. For the reader's convenience, we repeat here the definitions of $\hat{c}$, $\hat{\sigma}^2$, $\hat{m}_i(r)$, $\hat{m}_{is}(1,1)$, $\widehat{\tilde{m}}_i(r)$ and $\widehat{\tilde{m}}_{is}(1,1)$. For any $i, s \leq q$ and any non-negative integer r,

$$\begin{aligned}
\hat{c} &= n/T, \\
\hat{\sigma}^2 &= \sum_{j=\hat{q}+1}^{T} \lambda_j / (n - \hat{q}), \\
\widehat{\tilde{m}}_i(r) &= \frac{\hat{\sigma}^{2r}}{T - \hat{q}} \sum_{j=\hat{q}+1}^{T} (\lambda_i - \lambda_j)^{-r}, \\
\widehat{\tilde{m}}_{is}(1,1) &= \frac{\hat{\sigma}^4}{T - \hat{q}} \sum_{j=\hat{q}+1}^{T} (\lambda_i - \lambda_j)^{-1} (\lambda_s - \lambda_j)^{-1} \\
\hat{m}_i(r) &= \frac{\hat{\sigma}^{2r}}{n - \hat{q}} \sum_{j=\hat{q}+1}^{n} (\lambda_i - \lambda_j)^{-r} \text{ and} \\
\hat{m}_{is}(1,1) &= \frac{\hat{\sigma}^4}{n - \hat{q}} \sum_{j=\hat{q}+1}^{n} (\lambda_i - \lambda_j)^{-1} (\lambda_s - \lambda_j)^{-1}.
\end{aligned}$$

The consistency of $\hat{c}$ follows from Assumption 1 i). For $\hat{\sigma}^2$, note that it converges to the same limit as $\frac{1}{n}\sum_{j=1}^{n} \lambda_j$ because $\hat{q} \xrightarrow{p} q$ and $\lambda_q \leq ... \leq \lambda_1 = \left(\hat{L}'\hat{L}\right)_{11}$, which is bounded in probability by Theorem 1 iii). Further, $\frac{1}{n}\sum_{j=1}^{n} \lambda_j \equiv \frac{1}{nT}\operatorname{tr} XX'$ converges to the same limit as $\frac{1}{nT}\operatorname{tr} A\varepsilon BB'\varepsilon' A'$. Indeed, $XX' - A\varepsilon BB'\varepsilon' A'$ equals $LF'FL' + LF'B'\varepsilon' A' + A\varepsilon BFL'$, which is a matrix of rank $3k$ at most. Hence, by Theorem 4.3.6 in Horn and Johnson (1985), $\frac{1}{nT}\left|\operatorname{tr} XX' - \operatorname{tr} A\varepsilon BB'\varepsilon' A'\right|$ is no larger than $\frac{3k}{n}\left(\lambda_1 + \|A\varepsilon BB'\varepsilon' A'/T\|\right) \leq \frac{3k}{n}\left(\lambda_1 + \frac{1}{T}\|A\|^2\|B\|^2\|\varepsilon\|^2\right)$. But the expression in the brackets is bounded in probability

[7] To obtain and to check the explicit expressions we used the symbolic manipulation software of the Scientific Workplace, version 5.

by Theorem 1 iii), Assumption 2 and Lemma 1, whereas $\frac{3k}{n} \to 0$. It remains to note that $\frac{1}{nT}\operatorname{tr} A\varepsilon BB'\varepsilon' A' \overset{d}{=} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^{2}a_i b_t \xrightarrow{p} \sigma^2$ by Assumption 2 and the law of large numbers.

To establish such the consistency of $\hat{m}_i(r)$ and $\widehat{\tilde{m}}_i(r)$, we need to show that, for any integer r,

$$\frac{1}{T-\hat{q}}\sum_{j=\hat{q}+1}^{T}\frac{1}{(\lambda_i-\lambda_j)^r} \xrightarrow{p} \sigma^{-2r}\int\frac{d\mathcal{G}(\lambda)}{(x_i-\lambda)^r}, \text{ when } B = I_T, \text{ and} \tag{71}$$

$$\frac{1}{n-\hat{q}}\sum_{j=\hat{q}+1}^{n}\frac{1}{(\lambda_i-\lambda_j)^r} \xrightarrow{p} \sigma^{-2r}\int\frac{d\tilde{\mathcal{G}}(\lambda)}{(x_i-\lambda)^r} \text{ when } A = I_n. \tag{72}$$

Then, the consistency of $\hat{m}_i(r)$ and $\widehat{\tilde{m}}_i(r)$ would follow from the consistency of $\hat{\sigma}^2$ and from the continuous mapping theorem. Showing the convergences (71) and (72) is also sufficient for establishing the consistency of $\hat{d}_i$. Indeed, by Theorem 2, $d_i = \sigma^2\left(\int (x_i-\lambda)^{-1}d\tilde{\mathcal{G}}(\lambda)\right)^{-1} - \sigma^2$ if $B = I_T$ and $d_i = \sigma^2\left(\int (x_i-\lambda)^{-1}d\mathcal{G}(\lambda)\right)^{-1} - c\sigma^2$ if $A = I_n$. Hence, if the validity of (71) and (72) is established, the consistency of $\hat{d}_i$ would follow from (71) and (72), from the consistency of $\hat{\sigma}^2$ and $\hat{c}$, and from the continuous mapping theorem.

Let us denote the empirical distribution of $\lambda_{\hat{q}+1}/\sigma^2, ..., \lambda_n/\sigma^2$ as $\widehat{\tilde{\mathcal{G}}}(\lambda)$ and the empirical distribution of $\lambda_{\hat{q}+1}/\sigma^2, ..., \lambda_T/\sigma^2$ as $\widehat{\mathcal{G}}(\lambda)$. In this notation, we have:

$$\begin{aligned}\frac{1}{T-\hat{q}}\sum_{j=\hat{q}+1}^{T}\frac{1}{(\lambda_i-\lambda_j)^r} &= \sigma^{-2r}\int\frac{d\widehat{\mathcal{G}}(\lambda)}{(\lambda_i/\sigma^2-\lambda)^r} \text{ and}\\ \frac{1}{n-\hat{q}}\sum_{j=\hat{q}+1}^{n}\frac{1}{(\lambda_i-\lambda_j)^r} &= \sigma^{-2r}\int\frac{d\widehat{\tilde{\mathcal{G}}}(\lambda)}{(\lambda_i/\sigma^2-\lambda)^r}.\end{aligned}$$

We need to show that $\int\frac{d\widehat{\mathcal{G}}(\lambda)}{(\lambda_i/\sigma^2-\lambda)^r}$ converges in probability to $\int\frac{d\mathcal{G}(\lambda)}{(x_i-\lambda)^r}$ and that $\int\frac{d\widehat{\tilde{\mathcal{G}}}(\lambda)}{(\lambda_i/\sigma^2-\lambda)^r}$ converges in probability to $\int\frac{d\tilde{\mathcal{G}}(\lambda)}{(x_i-\lambda)^r}$. The latter convergence can be established similarly to the former one. Hence, we will focus on proving that $\int\frac{d\widehat{\mathcal{G}}(\lambda)}{(\lambda_i/\sigma^2-\lambda)^r} \xrightarrow{p} \int\frac{d\mathcal{G}(\lambda)}{(x_i-\lambda)^r}$.

For any $i \leq q$, let us define $\bar{\lambda}_i = \frac{x_i+\bar{x}}{2}$. Note that $\bar{\lambda}_i$ is outside of the support of $\mathcal{G}$, and it is outside of the support of $\widehat{\mathcal{G}}$ with probability arbitrarily close to 1 for large enough n. Therefore, it is enough to prove that $\int h_n(\lambda)d\widehat{\mathcal{G}}(\lambda)$ converges in probability to $\int h(\lambda)d\mathcal{G}(\lambda)$, where

$$h(\lambda) = \begin{cases}(x_i-\lambda)^{-r} \text{ for } 0\leq\lambda\leq\bar{\lambda}_i\\ \left(x_i-\bar{\lambda}_i\right)^{-r} \text{ for } \lambda>\bar{\lambda}_i\\ 0 \text{ for } \lambda<0\end{cases} \quad\text{and } h_n(\lambda) = \begin{cases}\left(\lambda_i/\sigma^2-\lambda\right)^{-r} \text{ for } 0\leq\lambda\leq\bar{\lambda}_i\\ \left(\lambda_i/\sigma^2-\bar{\lambda}_i\right)^{-r} \text{ for } \lambda>\bar{\lambda}\\ 0 \text{ for } \lambda<0\end{cases}.$$

Note that, with high probability, for large enough n, both $h(\lambda)$ and $h_n(\lambda)$ are continuous bounded functions on $\mathbb{R}$. It is because, by Theorem 1 iii) $\lambda_i/\sigma^2 \xrightarrow{p} x_i$.

Finally, the absolute value of the difference between $\int h_n(\lambda)d\widehat{\mathcal{G}}(\lambda)$ and $\int h(\lambda)d\mathcal{G}(\lambda)$ is no larger than A_1+A_2, where $A_1 = \int |h_n(\lambda) - h(\lambda)| \, d\widehat{\mathcal{G}}(\lambda)$ and $A_2 = \left|\int h(\lambda)d\widehat{\mathcal{G}}(\lambda) - \int h(\lambda)d\mathcal{G}(\lambda)\right|$. The term A_1 converges to zero in probability because $\lambda_i/\sigma^2 \xrightarrow{p} x_i$ and because $\lambda_{\hat{q}+1}/\sigma^2$, which is the upper boundary of the support of $\widehat{\mathcal{G}}(\lambda)$, converges in probability to $\bar{x} < \bar{\lambda}_i < x_i$. The term A_2 converges almost surely to zero because, according to Zhang (2006), $\widehat{\mathcal{G}}(\lambda)$ weakly converges to $\mathcal{G}(\lambda)$ almost surely, and because $h(\lambda)$ is a continuous bounded function on $\mathbb{R}$. Hence, $A_1 + A_2$ converges in probability to zero, which establishes the consistency of $\hat{m}_i\left(r\right)$. The consistency of $\widehat{\widetilde{m}}_i\left(r\right)$, $\widehat{\widetilde{m}}_{i,s}\left(1,1\right)$ and $\hat{m}_{i,s}\left(1,1\right)$ can be shown using similar arguments.

Theorem 3 now follows from the consistency of $\hat{c}$, $\hat{\sigma}^2$, $\hat{m}_i\left(r\right)$, $\widehat{\widetilde{m}}_i\left(r\right)$, $\widehat{\widetilde{m}}_{i,s}\left(1,1\right)$ and $\hat{m}_{i,s}\left(1,1\right)$ by continuous mapping theorem.□

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, Rhode Island.

[11] Marchenko, 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] Onatski, A. (2009) "Determining the number of factors from the empirical distribution of eigenvalues", forthcoming in the Review of Economics and Statistics.

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

[15] Silverstein, J.W. (1995) "Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices", Journal of Multivariate Analysis 55, 331-339

[16] 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

[17] Silverstein, J.W. and S. Choi (1995) "Analysis of the Limiting Spectral Distribution of Large Dimensional Random Matrices", Journal of Multivariate Analysis, 54, 295-309

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

[19] Yin, Y.Q. (1986) "Limiting Spectral Distribution of a Class of Random Matrices", Journal of Multivariate Analysis 20, 50-68.

[20] 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.

[21] Zhang, L. (2006). Spectral Analysis of Large Dimensional Random Matrices. Ph. D. Thesis. National University of Singapore.