Supplementary Material for "Estimating Number of Factors by Adjusted Eigenvalues Thresholding"

July 22, 2020

The supplementary material includes nine lemmas and their proofs, and Theorems 1, 2, 3 in the main text.

S.1 Some Lemmas

We first collect and establish some lemmas that will be used in the subsequent proofs. Let $\hat{x}_{ji} = x_{ji}1(|x_{ji}| \leq \eta_n\sqrt{n})$ and $\tilde{x}_{ji} = (\hat{x}_{ji} - \mathrm{E}\hat{x}_{ji})/\sqrt{\mathrm{Var}(\hat{x}_{ji})}$ with $K\eta_n \to 0$, $K\eta_n \log n \to +\infty$, $\mathbf{x}_i = (x_{1i}, \cdots, x_{p+K,i})^T$, $\hat{\mathbf{x}}_i = (\hat{x}_{1i}, \cdots, \hat{x}_{p+K,i})^T$ and $\tilde{\mathbf{x}}_i = (\tilde{x}_{1i}, \cdots, \tilde{x}_{p+K,i})^T$. Let the empirical spectral distribution of a $p \times p$ non-negative definite matrix $\mathbf{A}$ as

$$F^A(t) = p^{-1}\sum_{j=1}^{p} 1(\lambda_j(\mathbf{R}) \leq t). \tag{S.1}$$

Lemma S.1 *(Weyl theorem) Letting* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times p$ *Hermitian matrices, then we* *have for $\ell, j \in [p]$,*

$$\lambda_\ell(\mathbf{A}+\mathbf{B}) \leq \lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B}),\ \ell \geq j+k-1,$$
$$\lambda_j(\mathbf{A}) + \lambda_\ell(\mathbf{B}) \leq \lambda_{j+\ell-p}(\mathbf{A}+\mathbf{B}),\ j+\ell \geq p.$$

Lemma S.2 *Letting $\mathbf{A}$ and $\mathbf{B}$ be two $p \times p$ Hermitian matrices, then we have*

$$\lambda_j(\mathbf{A})\lambda_p(\mathbf{B}) \leq \lambda_j(\mathbf{AB}) \leq \lambda_j(\mathbf{A})\lambda_1(\mathbf{B}),\ j = 1, \cdots p.$$

Lemma S.3 *(Theorem A.43 of Bai and Silverstein (2010)) Letting $\mathbf{A}$ and $\mathbf{B}$ be two $p \times p$ Hermitian matrices, then we have*

$$\sup_t |F^{\mathbf{A}}(t) - F^{\mathbf{B}}(t)| \leq p^{-1}\operatorname{rank}(\mathbf{A}-\mathbf{B}).$$

Lemma S.4 *((9.9.6) of Bai and Silverstein (2010)) Suppose that the real-valued random variables $x_i, i = 1, \cdots, p$ are independent, with $\mathrm{E}x_i = 0$, $\mathrm{E}x_i^2 = 1$, $\sup_i \mathrm{E}x_i^4 = \beta_x < \infty$ and $|x_i| \leq \eta_n\sqrt{n}$, we have*

$$|\mathrm{E}\prod_{\ell=1}^{q}(\boldsymbol{\alpha}^T\mathbf{A}_\ell\boldsymbol{\alpha} - \operatorname{tr}(\mathbf{A}_\ell))| \leq Cn^{q-1}\eta_n^{2q-4}\prod_{\ell=1}^{q}\|\mathbf{A}_\ell\|,$$

where $\boldsymbol{\alpha} = (x_1, \cdots, x_p)^T$, q is a positive integer greater than 1, C is a positive constant and $\|\mathbf{A}_\ell\|$ is the spectral norm of the matrix $\mathbf{A}_\ell$.

Lemma S.5 *(Lemma 5.9 of Bai and Silverstein (2010)) Suppose that the entries of the array $\{x_{ji}, j \in [p+K], i \in [n]\}$ are independent (not necessarily identically distributed) and satisfy*

- $\mathrm{E}x_{ji} = 0,\ \ \mathrm{E}(|x_{ji}|^\ell) \leq b(\sqrt{n}\eta_n)^{\ell-3}$ *for all $\ell \geq 3$,*

- $|x_{ji}| \le \sqrt{n}\eta_n,$, $\max_{j,i} |\mathrm{E}x_{ji}^2 - 1| \to 0$ *as* $n \to \infty$,

where $\eta_n \to 0$ *and* $b > 0$. *Then for any* $x > \epsilon_0 > 0$ *and integers* $j, k \ge 2$, *we have*

$$P(\lambda_1(n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T) \ge (1+\sqrt{(p+K)/n})^2 + x) \le Cn^{-k}[(1+\sqrt{(p+K)/n})^2 + x - \epsilon_0]^{-k},$$

for some constant $C > 0$ *where* p *and* n *tend to infinity proportionally.*

Lemma S.6 *Suppose that* $\mathbf{x}_1, \cdots, \mathbf{x}_n$ *are i.i.d. and the entries of the array* $\{x_{j1}, j \in [p+K]\}$ *are independent (not necessarily identically distributed) and satisfy*

- $\mathrm{E}(x_{j1}) = 0$, $\mathrm{Var}(x_{j1}) = 1$ *and* $\sup_{j\in[p+K]} \mathrm{E}(x_{j1}^{4+\delta_0})$ *is bounded for all* p *and a positive constant* $\delta_0 > 0$.

Then we have

$$\begin{aligned}
&\lambda_1(n^{-1}\sum_{i=1}^n \hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T) \le (1+\sqrt{\rho_n})^2 + \epsilon_0, \ a.s.,\\
&\lambda_1(n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T) \le (1+\sqrt{\rho_n})^2 + \epsilon_0, \ a.s.,\\
&\lambda_1(n^{-1}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T) \le (1+\sqrt{\rho_n})^2 + \epsilon_0, \ a.s.,\\
&\lambda_1(n^{-1}\sum_{i=1}^n (\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T) \le (1+\sqrt{\rho_n})^2 + \epsilon_0, \ a.s.,\\
&\lambda_1(n^{-1}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathrm{diag}(\mathbf{0}_K, \lambda_{K+1}(\mathbf{R}), ..., \lambda_p(\mathbf{R})))\\
&\le \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})) + o_p(1),
\end{aligned}$$

for any positive constant ϵ_0.

Proof. It follows from $\mathrm{E}x_{j1} = 0$ that

$$|\mathrm{E}\hat{x}_{j1}| = |\mathrm{E}x_{j1}1(|x_{j1}| > \sqrt{n}\eta_n)| \le \frac{\mathrm{E}x_{j1}^4}{\eta_n^3 n^{3/2}} = O(\eta_n^{-3}n^{-3/2}),$$

leading to $\|\mathrm{E}\hat{\mathbf{x}}_i\|^2 = \sum_{j=1}^p (\mathrm{E}\hat{x}_{j1})^2 \le O(\eta_n^{-6}n^{-3}p)$. Moreover, we have

$$\begin{aligned} 1 \ge \mathrm{Var}(\hat{x}_{j1}) &= \mathrm{E}(\hat{x}_{j1}^2) - (\mathrm{E}\hat{x}_{j1})^2 \\ &= 1 - \mathrm{E}(x_{j1}^2 1(|x_{j1}| \ge \sqrt{n}\eta_n)) + O(\eta_n^{-6}n^{-3}) \\ &\ge 1 - \frac{\max_{j\in[p+K]} \mathrm{E}(x_{j1}^4)}{n\eta_n^2} + O(\eta_n^{-6}n^{-3}) = 1 + O(\eta_n^{-2}n^{-1}). \end{aligned}$$

Thus, we have

$$\max_{j\in[p+K]} |\mathrm{Var}(\hat{x}_{j1}) - 1| = o(1), \quad \max_{j\in[p+K]} |\mathrm{E}\hat{x}_{j1}| = O(\eta_n^{-3}n^{-3/2}). \tag{S.2}$$

Step 1. Proving $n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T = n^{-1}\sum_{i=1}^n \hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T, a.s.$ Following a similar proof of Yin et al. (2013), we have

$$\begin{aligned} & P(\{\mathbf{x}_1, ..., \mathbf{x}_n\} \ne \{\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_n\}, i.o.) \\ \le\ & \lim_{k\to\infty} \sum_{m=k}^{\infty} P\left(\bigcup_{2^{m-1}\le n<2^m} \bigcup_{i=1}^{2^m} \bigcup_{j=1}^{2\rho 2^m} \{|x_{ji}| \ge \eta_n\sqrt{n}\} \right) \\ =\ & \lim_{k\to\infty} \sum_{m=k}^{\infty} P\left(\bigcup_{i=1}^{2^m} \bigcup_{j=1}^{2\rho 2^m} \{|x_{ji}| \ge \eta_n 2^{m/2}\} \right) \\ \le\ & \lim_{k\to\infty} \sum_{m=k}^{\infty} \sum_{i=1}^{2^m} \sum_{j=1}^{2\rho 2^m} \mathrm{E}(|x_{ji}|^{4+\delta_0})\eta_n^{-(4+\delta_0)} 2^{-(4+\delta_0)m/2} \\ \le\ & \mathrm{E}(|x_{ji}|^{4+\delta_0})\eta_n^{-(4+\delta_0)} \lim_{k\to\infty} \sum_{m=k}^{\infty} \frac{1}{(2^{(\delta_0/2)})^m} = 0. \end{aligned} \tag{S.3}$$

In other words,

$$n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T = n^{-1}\sum_{i=1}^n \hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T, \ a.s. \tag{S.4}$$

Step 2. Using (S.2), we have

$$\begin{aligned}\lambda_1(n^{-1}\sum_{i=1}^n \hat{\mathbf{x}}_i\hat{\mathbf{x}}_i)^T &\le \lambda_1(n^{-1}\sum_{i=1}^n(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T) + \|\mathrm{E}(\hat{\mathbf{x}}_i)\|^2 \\ &\le \lambda_1(n^{-1}\sum_{i=1}^n(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T) + O(\eta_n^{-6}n^{-3}p). \qquad (S.5)\end{aligned}$$

Step 3. Letting $\mathbf{\Lambda} = \mathrm{diag}(\mathrm{Var}(\hat{x}_{11}), \cdots, \mathrm{Var}(\hat{x}_{p+K,1}))$, then by Lemma S.2

$$\begin{aligned}&\lambda_1(n^{-1}\sum_{i=1}^n(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T) = \lambda_1(n^{-1}\sum_{i=1}^n \mathbf{\Lambda}^{1/2}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\mathbf{\Lambda}^{1/2}) \\ \le\ &\lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T)\max_{j\in[p+K]}\mathrm{Var}(\hat{x}_{j1}) \le \lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T)[1+o(1)],\end{aligned}$$

where $\tilde{\mathbf{x}}_i = (\tilde{x}_{1i}, \cdots, \tilde{x}_{pi})^T$ and the second inequality is from (S.2). Thus, we conclude that

$$\lambda_1(n^{-1}\sum_{i=1}^n(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T) \le \lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T)[1+o(1)]. \qquad (S.6)$$

Step 4. It follows from $n^{-1}\sum_{i=1}^n(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \le n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T$ that

$$\lambda_1(n^{-1}\sum_{i=1}^n(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T) \le \lambda_1(n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T). \qquad (S.7)$$

Combination of (S.4)-(S.7) leads to

$$\lambda_1(n^{-1}\sum_{i=1}^n(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T) \le \lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T)[1+o(1)] + o_{a.s.}(1).$$

By Lemma S.5, for $0 < \epsilon_0 < 1$, we have

$$\sum_{n=1}^\infty P\left(\lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T) \ge (1+\sqrt{\rho_n})^2 + \epsilon_0\right) \le C\sum_{n=1}^\infty n^{-2} < \infty.$$

That is, $\lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T) \le (1+\sqrt{\rho_n})^2+\epsilon_0,\ a.s.$ Similarly, we have

$$\begin{aligned}
& \lambda_1(n^{-1}\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T\mathrm{diag}(\mathbf{0}_K,\lambda_{K+1}(\mathbf{R}),...,\lambda_p(\mathbf{R}))) \\
\le\ & \lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\mathrm{diag}(\mathbf{0}_K,\lambda_{K+1}(\mathbf{R}),...,\lambda_p(\mathbf{R})))[1+o(1)]+o_{a.s.}(1). \qquad \text{(S.8)}
\end{aligned}$$

By (1.2) and (1.6) of Bao et al. (2015), we have

$$\begin{aligned}
& \lambda_1(n^{-1}\sum_{i=1}^n \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\mathrm{diag}(\mathbf{0}_K,\lambda_{K+1}(\mathbf{R}),...,\lambda_p(\mathbf{R}))) \\
=\ & \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho}))+o_p(1). \qquad \text{(S.9)}
\end{aligned}$$

By (S.8)-(S.9), we have

$$\begin{aligned}
& \lambda_1(n^{-1}\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T\mathrm{diag}(\mathbf{0}_K,\lambda_{K+1}(\mathbf{R}),...,\lambda_p(\mathbf{R}))) \\
\le\ & \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho}))+o_p(1).
\end{aligned}$$

Therefore, all the desire conclusions follow.

Lemma S.7 *Suppose that the real-valued random variables $x_{hi}, h\in[p+K], i\in[n]$ are independent, with $\mathrm{E}x_{hi}=0$, $\mathrm{E}x_{hi}^2=1$, $\sup_{h,i}\mathrm{E}x_{hi}^6<\infty$ and $|x_{hi}|\le\eta_n\sqrt{n}$ satisfying $\eta_n\to 0$ and $\eta_n\log n\to+\infty$. Then, for matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ satisfying $\mathbf{U}_1\mathbf{U}_1^T=\mathbf{I}_K$, $\mathbf{U}_1\mathbf{U}_2^T=\mathbf{0}_{K\times(p-K)}$, bounded $\|\mathbf{U}_2\mathbf{U}_2^T\|$ and $\mathbf{X}=(x_{hi})_{h\in[p+K],i\in[n]}$, we have*

$$\sum_{\ell=1}^K[\lambda f(\ell,\ell,\mathbf{X})-\lambda s_n(\lambda)]^2+\sum_{1\le\ell_1\ne\ell_2\le K}[\lambda f(\ell_1,\ell_2,\mathbf{X})]^2=o_p(1), \qquad \text{(S.10)}$$

where $\mathbf{e}_\ell$ is the ℓth column of $\mathbf{I}_K$, $s_n(\lambda)=n^{-1}\mathrm{Etr}[(n^{-1}\mathbf{X}^T\mathbf{U_2}^T\mathbf{U}_2\mathbf{X}-\lambda\mathbf{I}_n)^{-1}]$,

$$f(\ell_1,\ell_2,\mathbf{X})=n^{-1}\mathbf{e}_{\ell_1}^T\mathbf{U}_1\mathbf{X}(n^{-1}\mathbf{X}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{X}-\lambda\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{U}_1^T\mathbf{e}_{\ell_2},$$

and $\lambda\ge\|\mathbf{U}_2\|^2(1+\sqrt{\rho})^2+\epsilon_0$ for some $\epsilon_0>0$.

Proof. Let $\{x_{hi}, h \in [p+K], i \in [n]\}$ be independent of the array $\{y_{hi}, h \in [p+K], i \in [n]\}$ with $y_{hi} \overset{i.i.d.}{\sim} N(0,1)$. Let $\hat{y}_{hi} = y_{hi} I_{\{|y_{hi}| \le \sqrt{n}\eta_n\}}$, $\tilde{y}_{hi} = \sigma_n^{-1} y_{hi} I_{\{|y_{hi}| \le \sqrt{n}\eta_n\}}$, where $\sigma_n^2 = \mathrm{E}\hat{y}_{hi}^2$, $\mathrm{E}\hat{y}_{hi} = \mathrm{E}\tilde{y}_{hi} = 0$ and $\mathrm{E}\tilde{y}_{hi}^2 = 1$. For $i \in [n]$, let

$$\mathbf{y}_i = (y_{1i}, ..., y_{p+K,i})^T, \quad \mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)^T, \quad \mathbf{w}_i = n^{-1/2}\mathbf{U}_1\mathbf{x}_i,$$
$$\hat{\mathbf{y}}_i = (\hat{y}_{1i}, ..., \hat{y}_{p+K,i})^T, \quad \hat{\mathbf{Y}} = (\hat{\mathbf{y}}_1, ..., \hat{\mathbf{y}}_n)^T, \quad \mathbf{z}_i = n^{-1/2}\mathbf{U}_2\mathbf{x}_i,$$
$$\tilde{\mathbf{Y}} = \sigma_n^{-1}\hat{\mathbf{Y}}, \quad \tilde{\mathbf{Y}} = (\tilde{\mathbf{y}}_1, ..., \tilde{\mathbf{y}}_n)^T, \quad \hat{\mathbf{w}}_i = n^{-1/2}\mathbf{U}_1\tilde{\mathbf{y}}_i, \ \hat{\mathbf{z}}_i = n^{-1/2}\mathbf{U}_2\tilde{\mathbf{y}}_i,$$

and $\tilde{\mathbf{y}}_i = \sigma_n^{-1}\hat{\mathbf{y}}_i$. Moreover, let $\mathbf{W}_{i0} = (\mathbf{U}_1\tilde{\mathbf{y}}_1, ..., \mathbf{U}_1\tilde{\mathbf{y}}_{i-1}, \mathbf{w}_{i+1}, ..., \mathbf{w}_n)$ and

$$\begin{aligned}
&\mathbf{Z}_i = (\mathbf{U}_2\tilde{\mathbf{y}}_1, ..., \mathbf{U}_2\tilde{\mathbf{y}}_i, \mathbf{z}_{i+1}, ..., \mathbf{z}_n), && \mathbf{Z}_0 = (\mathbf{z}_1, ..., \mathbf{z}_n),\\
&\mathbf{W}_i = (\mathbf{U}_1\tilde{\mathbf{y}}_1, ..., \mathbf{U}_1\tilde{\mathbf{y}}_i, \mathbf{w}_{i+1}, ..., \mathbf{w}_n), && \mathbf{W}_0 = (\mathbf{w}_1, ..., \mathbf{w}_n),\\
&\mathbf{Z}_{i0} = (\mathbf{U}_2\tilde{\mathbf{y}}_1, ..., \mathbf{U}_2\tilde{\mathbf{y}}_{i-1}, \mathbf{z}_{i+1}, ..., \mathbf{z}_n), && A_{i0}^{-1} = (\mathbf{Z}_{i0}\mathbf{Z}_{i0}^T - \lambda\mathbf{I}_{p-K})^{-1},\\
&\tilde{\beta}_{i\ell} = (\mathbf{z}_i^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell)^2, \ \tilde{\zeta}_{i\ell} = \mathbf{w}_i^T\mathbf{e}_\ell\mathbf{e}_\ell^T\mathbf{w}_i, && \tilde{\gamma}_{i\ell} = \mathbf{z}_i^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell\mathbf{e}_\ell^T\mathbf{w}_i,\\
&\hat{\beta}_{i\ell} = (\hat{\mathbf{z}}_i^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell)^2, \ \hat{\zeta}_{i\ell} = \hat{\mathbf{w}}_i^T\mathbf{e}_\ell\mathbf{e}_\ell^T\hat{\mathbf{w}}_i, && \hat{\gamma}_{i\ell} = \hat{\mathbf{z}}_i^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell\mathbf{e}_\ell^T\hat{\mathbf{w}}_i,\\
&\beta_{i\ell} = n^{-1}\|\mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{Z}_{i0}^T\mathbf{A}_{i0}^{-1}\mathbf{U}_2\|^2, && \alpha_{i\ell} = n^{-1}\mathrm{tr}(\mathbf{A}_{i0}^{-1}\mathbf{U}_2\mathbf{U}_2^T),\\
&\tilde{\alpha}_{i\ell} = \mathbf{z}_i^T\mathbf{A}_{i0}^{-1}\mathbf{z}_i, \quad \hat{\alpha}_{i\ell} = \hat{\mathbf{z}}_i^T\mathbf{A}_{i0}^{-1}\hat{\mathbf{z}}_i, && \zeta_{i\ell} = n^{-1}\mathbf{e}_\ell^T\mathbf{U}_2\mathbf{U}_2^T\mathbf{e}_\ell.
\end{aligned}$$

Then

$$f(\ell, \ell, \tilde{\mathbf{Y}}) = \mathbf{e}_\ell^T\mathbf{W}_n(\mathbf{Z}_n^T\mathbf{Z}_n - \lambda\mathbf{I}_n)^{-1}\mathbf{W}_n^T\mathbf{e}_\ell, f(\ell, \ell, \mathbf{X}) = \mathbf{e}_\ell^T\mathbf{W}_0(\mathbf{Z}_0^T\mathbf{Z}_0 - \lambda\mathbf{I}_n)^{-1}\mathbf{W}_0^T\mathbf{e}_\ell.$$

The proof of Lemma S.7 has three steps:

- Step 1. Proving $\sum_{\ell=1}^K \mathrm{E}|\lambda f(\ell, \ell, \tilde{\mathbf{Y}}) - \lambda s_n(\lambda)|^2 = o(1)$;
- Step 2. Establishing $\sum_{\ell=1}^K \mathrm{E}\{[\lambda f(\ell, \ell, \mathbf{X}) - \lambda s_n(z)]^2\} - \sum_{\ell=1}^K \mathrm{E}\{[\lambda f(\ell, \ell, \tilde{\mathbf{Y}}) - \lambda s_n(z)]^2\} = o(1)$;
- Step 3. Proving $\sum_{1 \le \ell_1 \ne \ell_2 \le K}[\lambda f(\ell_1, \ell_2, \mathbf{X})]^2 = o_p(1)$.

Proof of Step 1. For large enough positive integer N, we have

$$1-\sigma_n^2 = \mathrm{E}Y^2 I_{\{|Y|\ge\sqrt{n}\eta_n\}} \le n^{-3N}\eta_n^{-6N}\mathrm{E}Y^{2(3N+1)} = o(n^{-N}). \tag{S.11}$$

Let $f(\ell,\mathbf{Y}) = n^{-1}\mathbf{e}_\ell^T\mathbf{U}_1\mathbf{Y}(n^{-1}\mathbf{Y}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{Y}-\lambda\mathbf{I}_n)^{-1}\mathbf{Y}^T\mathbf{U}_1^T\mathbf{e}_\ell$. Using

$$P(\mathbf{Y}\ne\hat{\mathbf{Y}}) \le \sum_{j=1}^{p+K}\sum_{i=1}^{n}\eta_n^{-8}n^{-4}\mathrm{E}(y_{ji}^8) = o(1),$$

$\hat{\mathbf{Y}} = \sigma_n\tilde{\mathbf{Y}}$ and (S.11), it is not difficult to prove

$$\begin{aligned} &\sum_{\ell=1}^{K}\mathrm{E}\{[f(\ell,\mathbf{Y})-f(\ell,\tilde{\mathbf{Y}})]^2 \\ \le\ & 2\sum_{\ell=1}^{K}\mathrm{E}\{[f(\ell,\mathbf{Y})-f(\ell,\hat{\mathbf{Y}})]^2 + 2\sum_{\ell=1}^{K}\mathrm{E}\{[f(\ell,\hat{\mathbf{Y}})-f(\ell,\tilde{\mathbf{Y}})]^2 = o(1). \end{aligned} \tag{S.12}$$

By the normality and orthogonality, we have $n^{-1/2}\mathbf{e}_\ell^T\mathbf{U}_1\mathbf{Y}\sim N(0,1)$, $n^{-1/2}\mathbf{U}_2\mathbf{Y}\sim N(\mathbf{0}_K,\mathbf{U}_2\mathbf{U}_2^T)$, $n^{-1/2}\mathbf{e}_\ell^T\mathbf{U}_1\mathbf{Y}$ and $n^{-1/2}\mathbf{U}_2\mathbf{Y}$ being independent. Thus,

$$\sum_{\ell=1}^{K}\mathrm{E}\{[f(\ell,\ell,\mathbf{Y}) - n^{-1}\mathrm{tr}(n^{-1}\mathbf{Y}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{Y}-\lambda\mathbf{I}_n)^{-1}]^2\} = O(Kn^{-1}) = o(n^{-5/6}). \tag{S.13}$$

By (6.2.34) of Bai and Silverstein (2010), we have

$$\mathrm{E}[n^{-1}\mathrm{tr}(n^{-1}\mathbf{Y}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{Y}-\lambda\mathbf{I}_n)^{-1} - n^{-1}\mathrm{E}\mathrm{tr}(n^{-1}\mathbf{Y}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{Y}-\lambda\mathbf{I}_n)^{-1}]^2 = O(n^{-1}).$$

This entails that

$$\sum_{\ell=1}^{K}\mathrm{E}\{[f(\ell,\ell,\mathbf{Y}) - n^{-1}\mathrm{E}\mathrm{tr}(n^{-1}\mathbf{Y}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{Y}-\lambda\mathbf{I}_n)^{-1}]^2\} = o(n^{-5/6}). \tag{S.14}$$

By (6.2.41) of Bai and Silverstein (2010), we have

$$|n^{-1}\mathrm{E}\mathrm{tr}(n^{-1}\mathbf{Y}^T\mathbf{U}_2^T\mathbf{U}_2\mathbf{Y}-\lambda\mathbf{I}_n)^{-1} - s_n(\lambda)| \to 0. \tag{S.15}$$

Combining (S.12), (S.14) and (S.15), we obtain

$$\sum_{\ell=1}^{K} \mathrm{E}|\lambda f(\ell,\ell,\tilde{\mathbf{Y}}) - \lambda s_n(\lambda)|^2 = o(1). \tag{S.16}$$

Proof of Step 2. The proof of this requires some tedious calculations. By simple linear algebra and matrix computation, we have

$$\begin{aligned}
&\mathbf{W}_{i-1}\mathbf{Z}_{i-1}^T = \mathbf{W}_{i0}\mathbf{Z}_{i0}^T + \mathbf{w}_i\mathbf{z}_i^T, \quad \mathbf{W}_{i-1}\mathbf{W}_{i-1}^T = \mathbf{W}_{i0}\mathbf{W}_{i0}^T + \mathbf{w}_i\mathbf{w}_i^T,\\
&\lambda(\mathbf{Z}_{i-1}^T\mathbf{Z}_{i-1} - \lambda\mathbf{I}_n)^{-1} = \mathbf{Z}_{i-1}^T(\mathbf{Z}_{i-1}\mathbf{Z}_{i-1}^T - \lambda\mathbf{I}_{p-K})^{-1}\mathbf{Z}_{i-1} - \mathbf{I}_n,\\
&(\mathbf{Z}_{i-1}\mathbf{Z}_{i-1}^T - \lambda\mathbf{I}_{p-K})^{-1} = \mathbf{A}_{i0}^{-1} - (1+\mathbf{z}_i^T\mathbf{A}_{i0}^{-1}\mathbf{z}_i)^{-1}\mathbf{A}_{i0}^{-1}\mathbf{z}_i\mathbf{z}_i^T\mathbf{A}_{i0}^{-1}, \qquad (S.17)\\
&\mathbf{W}_i\mathbf{Z}_i^T = \mathbf{W}_{i0}\mathbf{Z}_{i0}^T + \hat{\mathbf{w}}_i\hat{\mathbf{z}}_i^T, \quad \mathbf{W}_i\mathbf{W}_i^T = \mathbf{W}_{i0}\mathbf{W}_{i0}^T + \hat{\mathbf{w}}_i\hat{\mathbf{w}}_i^T,\\
&\lambda(\mathbf{Z}_i^T\mathbf{Z}_i - \lambda\mathbf{I}_n)^{-1} = \mathbf{Z}_i^T(\mathbf{Z}_i\mathbf{Z}_i^T - \lambda\mathbf{I}_{p-K})^{-1}\mathbf{Z}_i - \mathbf{I}_n,\\
&(\mathbf{Z}_i\mathbf{Z}_i^T - \lambda\mathbf{I}_{p-K})^{-1} = \mathbf{A}_{i0}^{-1} - (1+\hat{\mathbf{z}}_i^T\mathbf{A}_{i0}^{-1}\hat{\mathbf{z}}_i)^{-1}\mathbf{A}_{i0}^{-1}\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^T\mathbf{A}_{i0}^{-1}. \qquad (S.18)
\end{aligned}$$

From (S.17), we have

$$\begin{aligned}
&\lambda\mathbf{e}_\ell^T\mathbf{W}_{i-1}(\mathbf{Z}_{i-1}^T\mathbf{Z}_{i-1} - \lambda\mathbf{I}_n)^{-1}\mathbf{W}_{i-1}^T\mathbf{e}_\ell \qquad (S.19)\\
=\;& \mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{Z}_{i0}^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell - \mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell - \frac{\tilde{\beta}_{i\ell} + \tilde{\zeta}_{i\ell} - 2\tilde{\gamma}_{i\ell}}{1+\tilde{\alpha}_{i\ell}}\\
=\;& \mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{Z}_{i0}^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell - \mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell - \frac{\beta_{i\ell}+\zeta_{i\ell}}{1+\alpha_{i\ell}} - \frac{\tilde{\beta}_{i\ell}-\beta_{i\ell}}{1+\alpha_{i\ell}} + \frac{\tilde{\gamma}_{i\ell}}{1+\alpha_{i\ell}}\\
&+\frac{(\tilde{\beta}_{i\ell}-\beta_{i\ell})(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2} - \frac{(\tilde{\beta}_{i\ell}-\beta_{i\ell})(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\tilde{\alpha}_{i\ell})(1+\alpha_{i\ell})^2} - \frac{\beta_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\tilde{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}\\
&+\frac{\beta_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2} - \frac{\tilde{\zeta}_{i\ell}-\zeta_{i\ell}}{1+\alpha_{i\ell}} + \frac{(\tilde{\zeta}_{i\ell}-\zeta_{i\ell})(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2} - \frac{(\tilde{\zeta}_{i\ell}-\zeta_{i\ell})(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\tilde{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}\\
&+\frac{\zeta_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2} - \frac{\zeta_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\tilde{\alpha}_{i\ell})(1+\alpha_{i\ell})^2} - \frac{\tilde{\gamma}_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2} + \frac{\tilde{\gamma}_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\tilde{\alpha}_{i\ell})(1+\alpha_{i\ell})^2},
\end{aligned}$$

where the last equality is obtained by repeatedly using

$$(1-\tilde{\alpha}_{i\ell})^{-1} = (1-\alpha_{i\ell})^{-1} - (\tilde{\alpha}_{i\ell}-\alpha_{i\ell})[(1-\tilde{\alpha}_{i\ell})(1-\alpha_{i\ell})]^{-1}.$$

Similarly, we have

$$
\begin{aligned}
& \lambda\mathbf{e}_\ell^T\mathbf{W}_i(\mathbf{Z}_i^T\mathbf{Z}_i-\lambda\mathbf{I}_n)^{-1}\mathbf{W}_i^T\mathbf{e}_\ell \qquad\qquad (S.20)\\
= & \ \mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{Z}_{i0}^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell-\mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell-\frac{\beta_{i\ell}+\zeta_{i\ell}}{1+\alpha_{i\ell}}-\frac{\hat{\beta}_{i\ell}-\beta_{i\ell}}{1+\alpha_{i\ell}}+\frac{\hat{\gamma}_{i\ell}}{1+\alpha_{i\ell}}\\
& +\frac{(\hat{\beta}_{i\ell}-\beta_{i\ell})(\hat{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2}-\frac{(\hat{\beta}_{i\ell}-\beta_{i\ell})(\hat{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\hat{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}-\frac{\beta_{i\ell}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\hat{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}\\
& +\frac{\beta_{i\ell}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2}-\frac{\hat{\zeta}_{i\ell}-\zeta_{i\ell}}{1+\alpha_{i\ell}}+\frac{(\hat{\zeta}_{i\ell}-\zeta_{i\ell})(\hat{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2}-\frac{(\hat{\zeta}_{i\ell}-\zeta_{i\ell})(\hat{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\hat{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}\\
& +\frac{\zeta_{i\ell}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2}-\frac{\zeta_{i\ell}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\hat{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}-\frac{\hat{\gamma}_{i\ell}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})}{(1+\alpha_{i\ell})^2}+\frac{\hat{\gamma}_{i\ell}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})^2}{(1+\hat{\alpha}_{i\ell})(1+\alpha_{i\ell})^2}.
\end{aligned}
$$

Letting $b_{i0\ell}=\mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{Z}_{i0}^T\mathbf{A}_{i0}^{-1}\mathbf{Z}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell-\mathbf{e}_\ell^T\mathbf{W}_{i0}\mathbf{W}_{i0}^T\mathbf{e}_\ell-\frac{\beta_{i\ell}+\zeta_{i\ell}}{1+\alpha_{i\ell}}$, we have

$$
\begin{aligned}
& \sum_{\ell=1}^K \mathrm{E}\{[\lambda f(\ell,\ell,\tilde{\mathbf{Y}})-\lambda s_n(z)]^2\}-\sum_{\ell=1}^K \mathrm{E}\{[\lambda f(\ell,\ell,\mathbf{X})-\lambda s_n(z)]^2\}\\
= & \sum_{\ell=1}^K\sum_{i=1}^n \mathrm{E}\{[\lambda\mathbf{e}_\ell^T\mathbf{W}_i(\mathbf{Z}_i^T\mathbf{Z}_i-\lambda\mathbf{I}_n)^{-1}\mathbf{W}_i^T\mathbf{e}_\ell-\lambda s_n(z)]^2\\
& -\mathrm{E}[\lambda\mathbf{e}_\ell^T\mathbf{W}_{i-1}(\mathbf{Z}_{i-1}^T\mathbf{Z}_{i-1}-\lambda\mathbf{I}_n)^{-1}\mathbf{W}_{i-1}^T\mathbf{e}_\ell-\lambda s_n(z)]^2\}. \qquad (S.21)
\end{aligned}
$$

The main terms of $\mathrm{E}(\lambda\mathbf{e}_\ell^T\mathbf{W}_i(\mathbf{Z}_i^T\mathbf{Z}_i-\lambda\mathbf{I}_n)^{-1}\mathbf{W}_i^T\mathbf{e}_\ell)-\mathrm{E}(b_{i0\ell})$ or $\mathrm{E}[\lambda\mathbf{e}_\ell^T\mathbf{W}_i(\mathbf{Z}_i^T\mathbf{Z}_i-\lambda\mathbf{I}_n)^{-1}\mathbf{W}_i^T\mathbf{e}_\ell]^2-\mathrm{E}(b_{i0\ell}^2)$ are as follows:

$$
\begin{aligned}
& \mathrm{E}(\hat{\alpha}_{i\ell}-\alpha_{i\ell})=0,\ \mathrm{E}(\hat{\beta}_{i\ell}-\beta_{i\ell})=0,\ \mathrm{E}\hat{\gamma}_{i\ell}=0,\ \mathrm{E}\{\tilde{\gamma}_{i\ell}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})=O(n^{-2}),\\
& \mathrm{E}[\beta_{i\ell}(\tilde{\zeta}_{i\ell}-\zeta_{i\ell})^2]=O(n^{-2}),\quad \mathrm{E}[(\tilde{\zeta}_{i\ell}-\zeta_{i\ell})^2]=O(n^{-2}),\\
& \mathrm{E}[(\tilde{\beta}_{i\ell}-\beta_{i\ell})^2]=O(n^{-2}),\quad \mathrm{E}\{(\tilde{\zeta}_{i\ell}-\zeta_{i\ell})(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})=O(n^{-2}),\\
& \mathrm{E}\{(\tilde{\beta}_{i\ell}-\beta_{i\ell})(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})=O(n^{-2}),\quad \mathrm{E}[(\tilde{\gamma}_{i\ell})^2]=O(n^{-2}),\\
& |\mathrm{E}[(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^{m_0}]|\le C\eta_n^{2(m_0-2)}n^{-1}\mathrm{E}(\|\mathbf{A}_{i0}^{-1}\|^{m_0})=o(K^{-2}n^{-1}),
\end{aligned}
\qquad (S.22)
$$

$$
\begin{aligned}
&|\mathrm{E}\{(\tilde{\beta}_{i\ell}-\beta_{i\ell})^{m_0}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^{m_1}| \le C\eta_n^{2(m_0+m_1-2)}n^{-1} = o(K^{-2}n^{-1}),\\
&|\mathrm{E}\{(\tilde{\zeta}_{i\ell}-\zeta_{i\ell})^{m_0}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^{m_1}| \le C\eta_n^{2(m_0+m_1-2)}n^{-1} = o(K^{-2}n^{-1}),\\
&|\mathrm{E}\{(\tilde{\gamma}_{i\ell})^{m_0}(\tilde{\alpha}_{i\ell}-\alpha_{i\ell})^{m_1}| \le C\eta_n^{2(m_0+m_1-2)}n^{-1} = o(K^{-2}n^{-1}),
\end{aligned}
\tag{S.23}
$$

for non-negative integers m_0, m_1 satisfying $m_0 + m_1 \ge 3$ where $\eta_n \to 0$ and the last four inequalities are from Lemma S.4. From (S.19)-(S.20)-(S.21)-(S.23), we have

$$
\sum_{\ell=1}^{K} \mathrm{E}\{[\lambda f(\ell,\ell,\tilde{\mathbf{Y}}) - \lambda s_n(z)]^2\} - \sum_{\ell=1}^{K} \mathrm{E}\{[\lambda f(\ell,\ell,\mathbf{X}) - \lambda s_n(z)]^2\} = o(1). \tag{S.24}
$$

Proof of Step 3: From (S.16)-(S.24), we have

$$
\sum_{\ell=1}^{K} \mathrm{E}\{[\lambda f(\ell,\ell,\mathbf{X}) - \lambda s_n(z)]^2\} = o(1)
$$

leading to

$$
\sum_{\ell=1}^{K} [\lambda f(\ell,\ell,\mathbf{X}) - \lambda s_n(z)]^2 = o_p(1). \tag{S.25}
$$

By Lemma A.3 of Jiang and Bai (2019), we have

$$
\sum_{1\le \ell_1 \ne \ell_2 \le K} [\lambda f(\ell_1,\ell_2,\mathbf{X})]^2 = o_{a.s.}(1). \tag{S.26}
$$

From (S.25) and (S.26), we have

$$
\sum_{\ell=1}^{K} [\lambda f(\ell,\ell,\mathbf{X}) - \lambda s_n(\lambda)]^2 + \sum_{1\le \ell_1 \ne \ell_2 \le K} [\lambda f(\ell_1,\ell_2,\mathbf{X})]^2 = o_p(1).
$$

This completes the proof.

Lemma S.8 *For the high dimensional factor model (7) satisfying Conditions C1-C2-C3-C4-C5 and Assumptions (a)-(b)-(c)-(d)-(e) in the main text, we have*

$$
\max_{j\in[p]} |\hat{\sigma}_{jj} - 1| = o_{a.s.}(1),
$$

where $\hat{\sigma}_{jj} = n^{-1}\sum_{i=1}^n \mathbf{e}_j^T\mathbf{Q}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\mathbf{Q}^T\mathbf{e}_j$ *with* $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^n \mathbf{x}_i$, $\mathbf{Q}$ *being defined in (9) in the main text and* $\mathbf{e}_j$ *is the* j*th column of* $\mathbf{I}_p$.

Proof. Let $\mathbf{Q} = (\mathbf{q}_{(1)}, \cdots, \mathbf{q}_{(p)})^T$ where $\mathbf{q}_{(j)} = (q_{j1}, \cdots, q_{j,p+K})^T$, $j \in [p]$. From (S.2), we have

$$\max_{j\in[p+K]} |\mathrm{Var}(\hat{x}_{j1}) - 1| = 1 + o(1) \text{ and } \max_{j\in[p+K]} |\mathrm{E}\hat{x}_{j1}| = O(\eta_n^{-3}n^{-3/2}), \tag{S.27}$$

where $o(1)$ and $O(1)$ are uniformly for all $j \in [p+K]$. Recall $\tilde{x}_{ji} \le c\sqrt{n}\eta_n$, for all $j \in [p+K]$ and $i \in [n]$ where c is a constant. The proof has two steps.

Step 1. Define $\tilde{\sigma}_{jj} = n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\mathbf{q}_{(j)}$ for $j \in [p]$. We will establish

$$\max_{j\in[p]} |\hat{\sigma}_{jj} - 1| \le \max_{j\in[p]} |\tilde{\sigma}_{jj} - 1|(1 + o(1)) + o_{a.s.}(1). \tag{S.28}$$

To arrive at this target, we will prove

$$\begin{cases} \max\limits_{j\in[p]} |\hat{\sigma}_{jj} - 1| \le \max\limits_{j\in[p]} |n^{-1}\sum\limits_{i=1}^n \mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)} - 1| + o_{a.s.}(1), \\ \max\limits_{j\in[p]} |n^{-1}\sum\limits_{i=1}^n \mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)} - 1| \le \max\limits_{j\in[p]} |n^{-1}\sum\limits_{i=1}^n \mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)} - 1| + o_{a.s.}(1), \\ \max\limits_{j\in[p]} |n^{-1}\sum\limits_{i=1}^n \mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)} - 1| \\ \qquad\qquad \le \max\limits_{j\in[p]} |n^{-1}\sum\limits_{i=1}^n \mathbf{q}_{(j)}^T(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T\mathbf{q}_{(j)} - 1| + o_{a.s.}(1), \\ \max\limits_{j\in[p]} |n^{-1}\sum\limits_{i=1}^n \mathbf{q}_{(j)}^T(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i - \mathrm{E}\hat{\mathbf{x}}_i)^T\mathbf{q}_{(j)} - 1| \\ \qquad\qquad \le \max\limits_{j\in[p]} |\tilde{\sigma}_{jj} - 1|(1 + o(1)) + o(1). \end{cases}$$

Step 2 will prove $\max\limits_{j\in[p]} |\tilde{\sigma}_{jj} - 1| = o_{a.s.}(1)$.

We now furnish the details of the proofs of Steps 1 and 2.

Step 1.1. Using the elementary identity

$$\hat{\sigma}_{jj} = n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)} - (n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\mathbf{x}_i)^2,$$

we have

$$\max_{j\in[p]}|\hat{\sigma}_{jj}-1| \le \max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)}-1| + \max_{j\in[p]}(n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i)^2.$$

We now show that the second term is negligible. This is easily shown by appealing to the Markov inequality. For all $\epsilon>0$, we have

$$\begin{aligned} P(\max_{j\in[p]}(n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i)^2 \ge \epsilon) &\le \sum_{j=1}^{p}P((n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i)^2\ge\epsilon) \\ \le \quad \epsilon^{-3}\sum_{j=1}^{p}n^{-6}\sum_{i,h,k,\ell,u,v} &\mathrm{E}[\mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{q}_{(j)}^T\mathbf{x}_h\mathbf{q}_{(j)}^T\mathbf{x}_k\mathbf{q}_{(j)}^T\mathbf{x}_\ell\mathbf{q}_{(j)}^T\mathbf{x}_u\mathbf{q}_{(j)}^T\mathbf{x}_v]. \end{aligned}$$

By noting that all odd moments are zero and $p=O(n)$, we have

$$P(\max_{j\in[p]}(n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i)^2 = O(n^{-2}).$$

Consequently, $\sum_{n=1}^{\infty}P(\max_{j\in[p]}(n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i)^2\ge\epsilon)<\infty$, which leads to

$$\max_{j\in[p]}(n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i)^2 = o_{a.s.}(1). \tag{S.29}$$

Thus,

$$\max_{j\in[p]}|\hat{\sigma}_{jj}-1| \le \max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)}-1| + o_{a.s.}(1). \tag{S.30}$$

Step 1.2. Let $\mathbf{X}=(\mathbf{x}_1,\cdots,\mathbf{x}_n)$, $\hat{\mathbf{X}}=(\hat{\mathbf{x}}_1,\cdots,\hat{\mathbf{x}}_n)$ and $\breve{\mathbf{x}}_i=(\breve{x}_{1i},\cdots,\breve{x}_{p+K,i})^T$, $i\in[n]$ with $\breve{x}_{j1}=x_{j1}1_{\{|x_{j1}|\ge\eta_n\sqrt{n}\}}$. Then,

$$\begin{aligned} &\sup_{j\in[p]}\mathrm{E}(|\breve{x}_{j1}|) \le (\eta_n\sqrt{n})^{-(5+\delta_0)}\sup_{j\in[p]}\mathrm{E}(|\breve{x}_{j1}|^{(6+\delta_0)}) = O(\eta_n^{-(5+\delta_0)}n^{-(5+\delta_0)/2}), \\ &\sup_{j\in[p]}\mathrm{E}(|\breve{x}_{j1}|^2) \le (\eta_n\sqrt{n})^{-(4+\delta_0)}\sup_{j\in[p]}\mathrm{E}(|\breve{x}_{j1}|^{(6+\delta_0})) = O(\eta_n^{-(4+\delta_0)}n^{-(2+\delta_0/2)}), \\ &\max_{j\in[p]}|\mathbf{q}_{(j)}^T\mathrm{E}\breve{\mathbf{x}}_i| \le \sqrt{p+K}\max_{j\in[p]}\mathrm{E}(|\breve{x}_{j1}|) = o(1). \end{aligned} \tag{S.31}$$

Let $\bar{\breve{\mathbf{x}}} = n^{-1}\sum_{i=1}^n \breve{\mathbf{x}}_i$ and $\breve{\mathbf{\Lambda}} = \mathrm{diag}(\mathrm{Var}(\breve{x}_{11}), \ldots, \mathrm{Var}(\breve{x}_{p+K,1}))$. From Lemma S.6, we have

$$\lambda_1(n^{-1}\sum_{i=1}^n \breve{\mathbf{\Lambda}}^{-1/2}(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})^T\breve{\mathbf{\Lambda}}^{-1/2}) = O_{a.s.}(1). \tag{S.32}$$

From (S.29) and (S.31), we have

$$\max_{j\in[p]}(n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\breve{\mathbf{x}}_i)^2 \leq 2\max_{j\in[p]}[(n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T(\breve{\mathbf{x}}_i - \mathrm{E}\breve{\mathbf{x}}_i))^2 + (\mathbf{q}_{(j)}^T\mathrm{E}\breve{\mathbf{x}}_i)^2] = o_{a.s.}(1). \tag{S.33}$$

It follows from the triangular inequality that

$$\begin{aligned}
& \max_{j\in[p]} n^{-1}(\|\mathbf{q}_{(j)}^T\mathbf{X}\| - \|\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|)^2 \\
\leq\ & \max_{j\in[p]} n^{-1}\|\mathbf{q}_{(j)}^T(\mathbf{X} - \hat{\mathbf{X}})\|^2 \\
=\ & \max_{j\in[p]} n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i^T\mathbf{q}_{(j)} \\
\leq\ & \max_{j\in[p]} n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})^T\mathbf{q}_{(j)} + \max_{j\in[p]} \mathbf{q}_{(j)}^T\bar{\breve{\mathbf{x}}}\bar{\breve{\mathbf{x}}}^T\mathbf{q}_{(j)} \\
\leq\ & \max_{j\in[p]} n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\breve{\mathbf{\Lambda}}^{1/2}[\breve{\mathbf{\Lambda}}^{-1/2}(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})^T\breve{\mathbf{\Lambda}}^{-1/2}]\breve{\mathbf{\Lambda}}^{1/2}\mathbf{q}_{(j)} \\
& + \max_{j\in[p]}(n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\breve{\mathbf{x}}_i)^2 \\
\leq\ & \max_{j\in[p]} \mathrm{Var}(\breve{x}_{j1}) \cdot \max_{j\in[p]} \lambda_1(n^{-1}\sum_{i=1}^n \breve{\mathbf{\Lambda}}^{-1/2}(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})(\breve{\mathbf{x}}_i - \bar{\breve{\mathbf{x}}})^T\breve{\mathbf{\Lambda}}^{-1/2}) \\
& + \max_{j\in[p]}(n^{-1}\sum_{i=1}^n \mathbf{q}_{(j)}^T\breve{\mathbf{x}}_i)^2 = o_{a.s.}(1),
\end{aligned}$$

where the last equality is from the combination of (S.31), (S.32) and (S.33). Thus,

$$\begin{aligned}
&\max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)}-1| \\
\le\ & \max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)}-1|+\max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}(\mathbf{q}_{(j)}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{q}_{(j)}-\mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)})| \\
=\ & \max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)}-1| \\
& +\max_{j\in[p]}|(\|n^{-1/2}\mathbf{q}_{(j)}^T\mathbf{X}\|-n^{-1/2}\|\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|)(n^{-1/2}\|\mathbf{q}_{(j)}^T\mathbf{X}\|+n^{-1/2}\|\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|)| \\
=\ & \max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)}-1|+o_{a.s.}(1).
\end{aligned}\tag{S.34}$$

The last inequality follows from

$$n^{-1}\|\mathbf{q}_{(j)}^T\mathbf{X}\|^2\le\lambda_1(n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T)\le(1+\sqrt{\rho_n})^2+\epsilon_0, a.s. \tag{S.35}$$

by Lemma S.6 and

$$n^{-1}\|\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|^2\le\lambda_1(n^{-1}\sum_{i=1}^{n}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T)\le(1+\sqrt{\rho_n})^2+\epsilon_0, a.s. \tag{S.36}$$

Step 1.3. It follows easily from the triangular inequality that

$$\begin{aligned}
&\max_{j\in[p]}|(\|n^{-1/2}\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|-\|n^{-1/2}\mathbf{q}_{(j)}^T(\hat{\mathbf{X}}-\mathrm{E}\hat{\mathbf{X}})\|)| \\
\le\ & \max_{j\in[p]}\|n^{-1/2}\mathbf{q}_{(j)}^T\mathrm{E}\hat{\mathbf{X}}\|=|\mathbf{q}_{(j)}^T\mathrm{E}\hat{\mathbf{x}}_1|\le O(\eta_n^{-3}n^{-3/2}p^{1/2})=o_{a.s.}(1).
\end{aligned}$$

This together with (S.35) and (S.36) lead to

$$\begin{aligned}
&\max_{j\in[p]}|(\|n^{-1/2}\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|^2-\|n^{-1/2}\mathbf{q}_{(j)}^T(\hat{\mathbf{X}}-\mathrm{E}\hat{\mathbf{X}})\|^2)| \\
=\ & \max_{j\in[p]}n^{-1}|(\|\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|-\|\mathbf{q}_{(j)}^T(\hat{\mathbf{X}}-\mathrm{E}\hat{\mathbf{X}})\|)|(\|\mathbf{q}_{(j)}^T\hat{\mathbf{X}}\|+\|\mathbf{q}_{(j)}^T(\hat{\mathbf{X}}-\mathrm{E}\hat{\mathbf{X}}\|)=o_{a.s.}(1).
\end{aligned}$$

That is,

$$\max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\mathbf{q}_{(j)}-1| \le \max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T(\hat{\mathbf{x}}_i-\mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i-\mathrm{E}\hat{\mathbf{x}}_i)^T\mathbf{q}_{(j)}-1|+o_{a.s.}(1).$$

Step 1.4. It is easily seen

$$\begin{aligned}
&\max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T(\hat{\mathbf{x}}_i-\mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i-\mathrm{E}\hat{\mathbf{x}}_i)^T\mathbf{q}_{(j)}-1| \\
\le\ & \max_{j\in[p]}|\tilde{\sigma}_{jj}-1|\max_{j\in[p]}\mathrm{Var}(\hat{x}_{jj})+\max_{j\in[p]}|1-\mathrm{Var}(\hat{x}_{jj})| \\
\le\ & \max_{j\in[p]}|\tilde{\sigma}_{jj}-1|(1+o(1))+o(1),
\end{aligned}$$

where the last inequality is from (S.27). By (S.35) and (S.36), we have

$$\max_{j\in[p]}|n^{-1}\sum_{i=1}^{n}\mathbf{q}_{(j)}^T(\hat{\mathbf{x}}_i-\mathrm{E}\hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_i-\mathrm{E}\hat{\mathbf{x}}_i)^T\mathbf{q}_{(j)}-1| \le \max_{j\in[p]}|\tilde{\sigma}_{jj}-1|(1+o(1))+o(1). \quad \text{(S.37)}$$

By (S.30), (S.34) and (S.37), we have

$$\max_{j\in[p]}|\hat{\sigma}_{jj}-1| \le \max_{j\in[p]}|\tilde{\sigma}_{jj}-1|(1+o(1))+o_{a.s.}(1). \quad \text{(S.38)}$$

Step 2. Recalling $\sup_{j\in[p+K]}\mathrm{E}(|x_{ji}^{6+\delta_0}|)<\infty$ and $|\tilde{x}_{ji}|\le c\eta_n\sqrt{n}$, we have

$$\sup_{j\in[p+K]}|\mathrm{E}(\tilde{x}_{ji}^{\ell})| \le \sup_{j\in[p+K]}\mathrm{E}(|\tilde{x}_{ji}|^{6+\delta_0})n^{(\ell-6-\delta_0)/2}\eta_n^{\ell-6-\delta_0}=o(n^{(\ell-6-\delta_0)/2}), \quad \text{(S.39)}$$

for $\ell\ge 6+\delta_0$ with $\delta_0>0$. By (S.39) and direct computation, we have

$$\begin{aligned}
&\mathrm{E}[(\mathbf{q}_{(j)}^T\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\mathbf{q}_{(j)}-1)^{\ell}]=O(1), \quad \ell=1,2,3, \\
&|\mathrm{E}[(\mathbf{q}_{(j)}^T\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^T\mathbf{q}_{(j)}-1)^{\ell}]|=o(n^{\ell-3-0.5\delta_0}), \quad \ell\ge 4. \quad \text{(S.40)}
\end{aligned}$$

By using the union bound and the Markov inequality, we have

$$
\begin{aligned}
& P(\max_{j\in[p]} |\tilde{\sigma}_{jj} - 1| \geq \epsilon) \leq \epsilon^{-6} \sum_{j=1}^{p} \mathrm{E}[(\tilde{\sigma}_{jj} - 1)^6] \\
= \;& c_1 n^{-6}\epsilon^{-6} \sum_{j=1}^{p} \sum_{i\neq h\neq \ell} \{\mathrm{E}[(\mathbf{q}_{(j)}^T \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T \mathbf{q}_{(j)} - 1)^2]\}^3 \\
& + c_2 n^{-6}\epsilon^{-6} \sum_{j=1}^{p} \sum_{i\neq h} \mathrm{E}[(\mathbf{q}_{(j)}^T \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T \mathbf{q}_{(j)} - 1)^2]\mathrm{E}[(\mathbf{q}_{(j)}^T \tilde{\mathbf{x}}_h \tilde{\mathbf{x}}_h^T \mathbf{q}_{(j)} - 1)^4] \\
& + n^{-6}\epsilon^{-6} \sum_{j=1}^{p} \sum_{i=1}^{n} \mathrm{E}[(\mathbf{q}_{(j)}^T \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T \mathbf{q}_{(j)} - 1)^6] \qquad \text{(S.41)} \\
& + c_3 n^{-6}\epsilon^{-6} \sum_{j=1}^{p} \sum_{i\neq h} \mathrm{E}[(\mathbf{q}_{(j)}^T \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T \mathbf{q}_{(j)} - 1)^3]\mathrm{E}[(\mathbf{q}_{(j)}^T \tilde{\mathbf{x}}_h \tilde{\mathbf{x}}_h^T \mathbf{q}_{(j)} - 1)^3],
\end{aligned}
$$

where c_1, c_2, c_3 are positive constants. By (S.40) and (S.41), we have

$$
P(\max_{j\in[p]} |\tilde{\sigma}_{jj} - 1| \geq \epsilon) \leq o(n^{-1-0.5\delta_0}).
$$

Since this probability sequence is summable, we conclude that $\max_{j\in[p]} |\tilde{\sigma}_{jj} - 1| = o_{a.s.}(1)$. By (S.38), we have

$$
\max_{j\in[p]} |\hat{\sigma}_{jj} - 1| = o_{a.s.}(1).
$$

This completes the proof of Lemma S.8.

S.2 Proof of (5) in Introduction

Assume that $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are i.i.d. samples from the model (1) in the main text where f_1, $\cdots$, f_K, ϵ_1, $\cdots$, ϵ_p are independent and are satisfying $\mathrm{E}f_i = \mathrm{E}\epsilon_j = 0$, $\mathrm{Var}(f_i) = \mathrm{Var}(\epsilon_j) = 1$ for $i \in [K], j \in [p]$ with $\max\{\mathrm{E}|f_i|^{4+\delta_0}, \mathrm{E}|\epsilon_j|^{6+\delta_0}, i \in [K], j \in [p]\}$ being bounded. Assuming

$\nu_{K+1}^2 = \alpha > 1 + \sqrt{p/n} + \epsilon_0$ and $\mathbf{B}_1 = \mathbf{\Gamma}\mathrm{diag}(a_1, \cdots, a_K)\mathbf{\Gamma}^T$ where ϵ_0 is a very small positive constant and $a_j = \alpha + K - j$ for $j \in [K]$, then

$$\lambda_j(\mathbf{\Sigma}) = \alpha + K + 1 - j,\ j \in [K+1],$$

and $\lambda_\ell(\mathbf{\Sigma}) = 1,\ \ell = K+2, \cdots, p$. Then

$$\mathrm{tr}(\mathbf{\Sigma}) = \sum_{j=1}^{K+1} \lambda_j(\mathbf{\Sigma}) + \sum_{j=K+2}^{p} \lambda_j(\mathbf{\Sigma}) = \sum_{j=1}^{K+1} (\alpha + K + 1 - j) + (p - K - 1). \tag{S.42}$$

By (12) in the main text, we have

$$\hat{\mathbf{\Sigma}}_n = n^{-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T = \mathbf{Q} n^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{Q}^T,$$

where $\mathbf{Q} = (\mathbf{B}, \mathrm{diag}(\nu_1, ..., \nu_K))$. By Lemmas S.2-S.6, we have

$$0 \le \lambda_j(\hat{\mathbf{\Sigma}}_n) \le (1 + \sqrt{p/n})^2 + \epsilon_0,\ a.s. \tag{S.43}$$

for $j \ge K+2$. Thus, for $j \in [K+1]$, by (2.4) of Bai and Yao (2008), we have

$$\lambda_j(\hat{\mathbf{\Sigma}}_n) - \left[\alpha + K + 1 - j + (p/n)\frac{\alpha + K + 1 - j}{\alpha + K - j}\right] \to 0,\ a.s.$$

That is, for $j \in [K+1]$,

$$\begin{aligned} \lambda_j(\hat{\mathbf{\Sigma}}_n) &\ge \alpha + K + 1 - j + (p/n)\frac{\alpha + K + 1 - j}{\alpha + K - j} - \epsilon_0,\ a.s. && \text{(S.44)} \\ &\ge \alpha + K + 1 - j - \epsilon_0,\ a.s. && \text{(S.45)} \end{aligned}$$

$$\begin{aligned} \lambda_j(\hat{\mathbf{\Sigma}}_n) &\le \alpha + K + 1 - j + (p/n)\frac{\alpha + K + 1 - j}{\alpha + K - j} + \epsilon_0,\ a.s. && \text{(S.46)} \\ &\le \alpha + K + 1 - j + (p/n + \sqrt{p/n}) + \epsilon_0,\ a.s. && \text{(S.47)} \end{aligned}$$

From (S.44) and (S.46), for $j \in [K]$, we have

$$\lambda_j(\hat{\mathbf{\Sigma}}_n) - \lambda_{j+1}(\hat{\mathbf{\Sigma}}_n) \geq 1 - (p/n)[(\alpha + K - j)(\alpha + K - j - 1)]^{-1} - 2\epsilon_0, \ a.s. \tag{S.48}$$

$$\begin{aligned} \lambda_j(\hat{\mathbf{\Sigma}}_n) - \lambda_{j+1}(\hat{\mathbf{\Sigma}}_n) &\leq 1 - (p/n)[(\alpha + K - j)(\alpha + K - j - 1)]^{-1} + 2\epsilon_0, \ a.s. && \text{(S.49)} \\ &\leq 1 + 2\epsilon_0, \ a.s. && \text{(S.50)} \end{aligned}$$

From (S.43) and (S.44), we have

$$\begin{aligned} \lambda_{K+1}(\hat{\mathbf{\Sigma}}_n) - \lambda_{K+2}(\hat{\mathbf{\Sigma}}_n) &\geq \alpha + (p/n)\frac{\alpha}{\alpha - 1} - (1 + \sqrt{p/n})^2 - 2\epsilon_0, \ a.s. \\ &\geq \alpha - (1 + \sqrt{p/n})^2 - 2\epsilon_0, \ a.s. && \text{(S.51)} \\ \lambda_{K+2}(\hat{\mathbf{\Sigma}}_n) - \lambda_{K+3}(\hat{\mathbf{\Sigma}}_n) &\leq (1 + \sqrt{p/n})^2 + \epsilon_0, a.s. && \text{(S.52)} \end{aligned}$$

When α is large enough and p/n tends to a constant, for $j \in [K]$, we have

$$\alpha - (1 + \sqrt{p/n})^2 - 2\epsilon_0 > 1 + 2\epsilon_0, \tag{S.53}$$

which leads to

$$\lambda_j(\hat{\mathbf{\Sigma}}_n) - \lambda_{j+1}(\hat{\mathbf{\Sigma}}_n) < \lambda_{K+1}(\hat{\mathbf{\Sigma}}_n) - \lambda_{K+2}(\hat{\mathbf{\Sigma}}_n), a.s. \tag{S.54}$$

from (S.50) and (S.51), for $j \in [K]$. Thus, when α is large enough and p/n tends to a constant, from (S.54), we have

$$\hat{K}_{ED} = \max\{i \leq r_{\max} : \hat{\lambda}_i - \hat{\lambda}_{i+1} \geq t_0\} \geq K + 1, a.s.$$

where t_0 is a given threshold. From (S.48) and (S.50), for $j \in [K-1]$ we have

$$\begin{aligned} \frac{\lambda_j(\hat{\mathbf{\Sigma}}_n) - \lambda_{j+1}(\hat{\mathbf{\Sigma}}_n)}{\lambda_{j+1}(\hat{\mathbf{\Sigma}}_n) - \lambda_{j+2}(\hat{\mathbf{\Sigma}}_n)} &\leq \frac{1 + 2\epsilon_0}{1 - (p/n)[(\alpha + K - j - 1)(\alpha + K - j - 2)]^{-1} - 2\epsilon_0}, \ a.s. \\ &\leq \frac{1 + 2\epsilon_0}{1 - 2\epsilon_0}, \ a.s. && \text{(S.55)} \end{aligned}$$

where the second inequality holds when α is large enough. From (S.50) and (S.51), we have

$$\frac{\lambda_K(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{K+1}(\hat{\boldsymbol{\Sigma}}_n)}{\lambda_{K+1}(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{K+2}(\hat{\boldsymbol{\Sigma}}_n)} \leq \frac{1+2\epsilon_0}{\alpha - (1+\sqrt{p/n})^2 - 2\epsilon_0}, \ a.s. \tag{S.56}$$

From (S.51) and (S.52), we have

$$\frac{\lambda_{K+1}(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{K+2}(\hat{\boldsymbol{\Sigma}}_n)}{\lambda_{K+2}(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{K+3}(\hat{\boldsymbol{\Sigma}}_n)} \geq \frac{\alpha - (1+\sqrt{p/n})^2 - 2\epsilon_0}{(1+\sqrt{p/n})^2 + \epsilon_0}, \ a.s. \tag{S.57}$$

Thus, when α is large enough and p/n tends to a constant, from (S.55), (S.56) and (S.57), for $j \in [K-1]$, we have

$$\frac{\lambda_j(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{j+1}(\hat{\boldsymbol{\Sigma}}_n)}{\lambda_{j+1}(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{j+2}(\hat{\boldsymbol{\Sigma}}_n)} < \frac{\lambda_{K+1}(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{K+2}(\hat{\boldsymbol{\Sigma}}_n)}{\lambda_{K+2}(\hat{\boldsymbol{\Sigma}}_n) - \lambda_{K+3}(\hat{\boldsymbol{\Sigma}}_n)}, \ a.s..$$

That is,

$$\hat{K}_{ON} = \arg \max_{r_{\min} < i \leq r_{\max}} (\hat{\lambda}_i - \hat{\lambda}_{i+1})/(\hat{\lambda}_{i+1} - \hat{\lambda}_{i+2}) \geq K+1, \ a.s.$$

From (S.43), (S.45) and (S.47), for $j \in [K]$, we have

$$\begin{aligned} \frac{\hat{\lambda}_j(\hat{\boldsymbol{\Sigma}}_n)}{\hat{\lambda}_{j+1}(\hat{\boldsymbol{\Sigma}}_n)} &\leq \frac{\alpha + K + 1 - j + (p/n + \sqrt{p/n}) + \epsilon_0}{\alpha + K - j - \epsilon_0}, \ a.s. \\ \frac{\hat{\lambda}_{K+1}(\hat{\boldsymbol{\Sigma}}_n)}{\hat{\lambda}_{K+2}(\hat{\boldsymbol{\Sigma}}_n)} &\geq \frac{\alpha - \epsilon_0}{(1+\sqrt{p/n})^2 + \epsilon_0}, \ a.s. \end{aligned}$$

When α is large enough and p/n tends to a constant, we have

$$\frac{\alpha + K + 1 - j + (p/n + \sqrt{p/n}) + \epsilon_0}{\alpha + K - j - \epsilon_0} < \frac{\alpha - \epsilon_0}{(1+\sqrt{p/n})^2 + \epsilon_0},$$

which leads to

$$\frac{\hat{\lambda}_j(\hat{\boldsymbol{\Sigma}}_n)}{\hat{\lambda}_{j+1}(\hat{\boldsymbol{\Sigma}}_n)} < \frac{\hat{\lambda}_{K+1}(\hat{\boldsymbol{\Sigma}}_n)}{\hat{\lambda}_{K+2}(\hat{\boldsymbol{\Sigma}}_n)}, \ a.s.$$

Then we have

$$\hat{K}_{ER} = \arg \max_{1 \leq i \leq r_{\max}} \hat{\lambda}_i / \hat{\lambda}_{i+1} \geq K+1, \ a.s. \tag{S.58}$$

From (S.45) and (S.47), for $i \in [K]$, $\hat{\lambda}_i \le \alpha + K + (p/n + \sqrt{p/n}) + \epsilon_0$ and $\hat{\lambda}_i \ge \alpha - \epsilon_0$. Then for $i \in [K]$ and $V_i = \sum_{j=i+1}^p \hat{\lambda}_j$,

$$\log(V_{i-1}/V_i) = \log \frac{\sum_{j=i}^p \hat{\lambda}_j}{\sum_{j=i+1}^p \hat{\lambda}_j} = \log(1+\hat{\lambda}_i/\sum_{j=i+1}^p \hat{\lambda}_j) \le \log\left[1 + \frac{\alpha + K + (p/n + \sqrt{p/n}) + \epsilon_0}{(K+1-i)(\alpha - \epsilon_0)}\right],$$

which leads to

$$\log(V_{i-1}/V_i) \to \log 2, \tag{S.59}$$

in probability when α is large enough. Moreover, by (S.45) and (S.47), for $i \in [K-1]$, we have

$$\log(V_i/V_{i+1}) \ge \log\left[1 + \frac{\alpha + K - i - \epsilon_0}{(K-i)(\alpha + K + p/n + \sqrt{p/n} + \epsilon_0) + (p-K-1)[(1+\sqrt{p/n})^2 + \epsilon_0]}\right],$$

which leads to

$$\log(V_i/V_{i+1}) \to \log 2, \tag{S.60}$$

in probability when α is large enough. We have

$$\log(V_K/V_{K+1}) \ge \log\left[1 + \frac{\alpha - \epsilon_0}{(p-K-1)[(1+\sqrt{p/n})^2 + \epsilon_0]}\right]. \tag{S.61}$$

Because $p^{-1}\mathrm{tr}(\hat{\mathbf{\Sigma}}_n) - p^{-1}\mathrm{tr}(\mathbf{\Sigma}) = o_p(1)$, then we have

$$\begin{aligned}
& p^{-1}\sum_{j=K+3}^{p}\hat{\lambda}_j(\hat{\mathbf{\Sigma}}_n) \\
= \;& p^{-1}\mathrm{tr}(\hat{\mathbf{\Sigma}}_n) - p^{-1}\sum_{j=1}^{K+1}\hat{\lambda}_j(\hat{\mathbf{\Sigma}}_n) - p^{-1}\hat{\lambda}_{K+2}(\hat{\mathbf{\Sigma}}_n) \\
= \;& p^{-1}\mathrm{tr}(\mathbf{\Sigma}) - p^{-1}\sum_{j=1}^{K+1}\hat{\lambda}_j(\hat{\mathbf{\Sigma}}_n) - p^{-1}\hat{\lambda}_{K+2}(\hat{\mathbf{\Sigma}}_n) + o_p(1) \\
\geq \;& p^{-1}\mathrm{tr}(\mathbf{\Sigma}) - p^{-1}\sum_{j=1}^{K+1}(\alpha + K + 1 - j + (p/n + \sqrt{p/n}) + \epsilon_0) \\
& - p^{-1}(1+\sqrt{p/n})^2 - p^{-1}\epsilon_0 + o_p(1) \\
\geq \;& p^{-1}\sum_{j=1}^{K+1}(\alpha + K + 1 - j) + (p - K - 1)p^{-1} \\
& - p^{-1}\sum_{j=1}^{K+1}(\alpha + K + 1 - j + (p/n + \sqrt{p/n}) + \epsilon_0) - p^{-1}(1+\sqrt{p/n})^2 - p^{-1}\epsilon_0 + o_p(1) \\
= \;& (p - K - 1)p^{-1} - (K+1)p^{-1}((p/n + \sqrt{p/n}) + \epsilon_0) - p^{-1}(1+\sqrt{p/n})^2 - p^{-1}\epsilon_0 + o_p(1),
\end{aligned}$$

where the first inequality is from (S.43) and (S.47), and the second inequality is from (S.42). Then when p/n tends to a constant, we have

$$\begin{aligned}
& \frac{p^{-1}\hat{\lambda}_{K+2}(\hat{\mathbf{\Sigma}}_n)}{p^{-1}\sum_{j=K+3}^{p}\hat{\lambda}_j(\hat{\mathbf{\Sigma}}_n)} \\
\leq \;& \frac{p^{-1}(1+\sqrt{p/n})^2 + p^{-1}\epsilon_0}{(p - K - 1)p^{-1} - (K+1)p^{-1}((p/n + \sqrt{p/n}) + \epsilon_0) - p^{-1}(1+\sqrt{p/n})^2 - p^{-1}\epsilon_0 + o_p(1)} \\
\to \;& 0,
\end{aligned}$$

in probability. Thus, when p/n tends to a constant, we have

$$\log(V_{K+1}/V_{K+2}) = \log\left[1 + \frac{\hat{\lambda}_{K+2}(\hat{\mathbf{\Sigma}}_n)}{\sum_{j=K+3}^{p}\hat{\lambda}_j(\hat{\mathbf{\Sigma}}_n)}\right] \to 1. \tag{S.62}$$

From (S.59), (S.60), (S.61) and (S.62), we have

$$\log(V_{i-1}/V_i)/\log(V_i/V_{i+1}) \to 1,$$

for $j \in [K]$ in probability and $\log(V_K/V_{K+1})/\log(V_{K+1}/V_{K+2})$ is large enough when α is large enough. That is, when α is large enough,

$$\log(V_{i-1}/V_i)/\log(V_i/V_{i+1}) < \log(V_K/V_{K+1})/\log(V_{K+1}/V_{K+2}),$$

in probability. That is,

$$\hat{K}_{GR} = \arg\max_{1 \le i \le r_{\max}} \log(V_{i-1}/V_i)/\log(V_i/V_{i+1}) \ge K+1,$$

in probability. Then when α is large enough, we conclude that

$$P(\hat{K}_{ON} \ge K+1) \to 1, \qquad P(\hat{K}_{ED} \ge K+1) \to 1,$$
$$P(\hat{K}_{ER} \ge K+1) \to 1, \qquad P(\hat{K}_{GR} \ge K+1) \to 1.$$

S.3 Proof of Theorem 1

Proof. By Lemma S.1 and for $i, j, k \le p$, we have

$$\lambda_i(\mathbf{Q}_1\mathbf{Q}_1^T + \mathbf{Q}_2\mathbf{Q}_2^T) \le \lambda_j(\mathbf{Q}_1\mathbf{Q}_1^T) + \lambda_k(\mathbf{Q}_2\mathbf{Q}_2^T),\ i \ge j+k-1,$$
$$\lambda_j(\mathbf{Q}_1\mathbf{Q}_1^T) + \lambda_k(\mathbf{Q}_2\mathbf{Q}_2^T) \le \lambda_{j+k-p}(\mathbf{Q}_1\mathbf{Q}_1^T + \mathbf{Q}_2\mathbf{Q}_2^T),\ j+k \ge p.$$

Notice that $\lambda_{K+1}(\mathbf{Q}_1\mathbf{Q}_1^T) = \cdots = \lambda_p(\mathbf{Q}_1\mathbf{Q}_1^T) = 0$ because of $\mathrm{rank}(\mathbf{Q}_1\mathbf{Q}_1^T) = K < p$. Following Lemma S.1, we have for $i \ge K+1$,

$$\lambda_i(\mathbf{R}) \le \lambda_{K+1}(\mathbf{Q}_1\mathbf{Q}_1^T) + \lambda_1(\mathbf{Q}_2\mathbf{Q}_2^T) = \lambda_1(\mathbf{Q}_2\mathbf{Q}_2^T) \le 1,$$

if $\lambda_1(\mathbf{Q}_2\mathbf{Q}_2^T) = \|\mathbf{Q}_2\mathbf{Q}_2^T\| = \|[\text{diag}(\mathbf{\Sigma})]^{-1}\mathbf{\Psi}\|^2 \leq 1$.

Next, letting ν_j^2 be the jth diagonal element of $\mathbf{\Psi}$, then by using $\text{tr}(\mathbf{R}) = p$, we have $p = \text{tr}\mathbf{Q}_1^T\mathbf{Q}_1 + \sum_{j=1}^p \nu_j^2/\sigma_{jj}$, and

$$\lambda_1(\mathbf{Q}_1\mathbf{Q}_1^T) + \cdots + \lambda_K(\mathbf{Q}_1\mathbf{Q}_1^T) = p - \sum_{j=1}^p \nu_j^2/\sigma_{jj} = \|[\text{diag}(\mathbf{\Sigma})]^{-1/2}\mathbf{B}\|_F^2.$$

By the assumption, we have

$$\lambda_1(\mathbf{Q}_1^T\mathbf{Q}_1)/\lambda_K(\mathbf{Q}_1^T\mathbf{Q}_1) = \|\mathbf{B}^T[\text{diag}(\mathbf{\Sigma})]^{-1}\mathbf{B}\| \cdot \|\{\mathbf{B}^T[\text{diag}(\mathbf{\Sigma})]^{-1}\mathbf{B}\}^{-1}\| = O(p^{\delta_2}).$$

Hence,

$$O(p^{\delta_2})K\lambda_K(\mathbf{Q}_1^T\mathbf{Q}_1) \geq \|[\text{diag}(\mathbf{\Sigma})]^{-1/2}\mathbf{B}\|_F^2 = O(p^{\delta_1}).$$

This entails that $\lambda_K(\mathbf{Q}_1^T\mathbf{Q}_1) \geq O(p^{\delta_1-\delta_2-\delta_3})$. Consequently, we have $\lambda_K(\mathbf{R}) \geq \lambda_K(\mathbf{Q}_1^T\mathbf{Q}_1) > 1$ when p is large enough with $\delta_1 - \delta_2 - \delta_3 > 0$. This completes the proof of Theorem 1.

S.4 Proof of Theorem 2

By (12) in the main text and $\mathbf{y}_i = (\mathbf{B}, \mathbf{\Psi}^{1/2})\mathbf{x}_i$, $i \in [n]$, we have

$$\hat{\mathbf{\Sigma}}_n = n^{-1}\sum_{i=1}^n (\mathbf{B}, \mathbf{\Psi}^{1/2})(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T(\mathbf{B}, \mathbf{\Psi}^{1/2})^T.$$

Note that

$$\hat{\mathbf{R}} = [\text{diag}(\hat{\mathbf{\Sigma}}_n)]^{-1/2}\hat{\mathbf{\Sigma}}_n[\text{diag}(\hat{\mathbf{\Sigma}}_n)]^{-1/2} = [\text{diag}(\mathbf{S}_n)]^{-1/2}\mathbf{S}_n[\text{diag}(\mathbf{S}_n)]^{-1/2},$$

where

$$\begin{aligned}\mathbf{S}_n &= n^{-1}\sum_{i=1}^n [\text{diag}(\mathbf{\Sigma})]^{-1/2}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T[\text{diag}(\mathbf{\Sigma})]^{-1/2} \\ &= n^{-1}\sum_{i=1}^n \mathbf{Q}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\mathbf{Q}^T = (\hat{\sigma}_{\ell_1\ell_2})_{\ell_1,\ell_2\in[p]}.\end{aligned}$$

The proof of Theorem 2 consists of two steps.

- Step 1 will prove that $\max_{k=K+1,\cdots,p} |\lambda_k(\hat{\mathbf{R}}) - \lambda_k(\mathbf{S}_n)| \to 0,\ a.s.$
- Step 2 will prove that the Stieltjes transform $m(z)$ of the limiting spectral distribution of $\mathbf{S}_n$ satisfies

$$z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1 + t\underline{m}(z)} = -\underline{m}^{-1}(z)\psi(-\underline{m}^{-1}(z)),$$

  with $\underline{m}(z) = -(1-\rho)z^{-1} + \rho m(z)$.

We now establish the results in Steps 1 and 2. Let $0/0 = 1$.

Step 1. By Lemma S.2, we have

$$\min_{j\in[p]} \hat{\sigma}_{jj} = \min_{j\in[p]} \lambda_j(\mathrm{diag}(\mathbf{S}_n)) \le \frac{\lambda_k(\mathbf{S}_n)}{\lambda_k(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1})},$$
$$\frac{\lambda_k(\mathbf{S}_n)}{\lambda_k(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1})} \le \max_{j\in[p]} \lambda_j(\mathrm{diag}(\mathbf{S}_n)) = \max_{j\in[p]} \hat{\sigma}_{jj},$$

for $k \in [p]$. Thus,

$$\begin{aligned} & \max_{k=1,\cdots,p} \left| \frac{\lambda_k(\mathbf{S}_n)}{\lambda_k(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1})} - 1 \right| \\ \le\ & \max\{|\min_{j\in[p]} \hat{\sigma}_{jj} - 1|, |\max_{j\in[p]} \hat{\sigma}_{jj} - 1|\} = \max_{j\in[p]} |\hat{\sigma}_{jj} - 1|, \end{aligned} \tag{S.63}$$

which converges to 0, a.s., by Lemma S.8. Consequently, we have

$$\max_{k\in[p]} \left| \frac{\lambda_k(\mathbf{S}_n)}{\lambda_k(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1})} - 1 \right| = o_{a.s.}(1). \tag{S.64}$$

Note that

$$\begin{aligned} & \max_{k=K+1,\cdots,p} \left| \frac{\lambda_k(\mathbf{S}_n)}{\lambda_k(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1})} - 1 \right| \\ \ge\ & \frac{\max_{k=K+1,\cdots,p} |\lambda_k(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1}) - \lambda_k(\mathbf{S}_n)|}{\lambda_{K+1}(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1})} = o_{a.s.}(1), \end{aligned} \tag{S.65}$$

by using (S.64). By using Lemma S.2 again, we have

$$\lambda_{K+1}(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1}) \leq \frac{\lambda_{K+1}(\mathbf{S}_n)}{\min_{j\in[p]} \hat{\sigma}_{jj}}$$

$$\leq \frac{\lambda_{K+1}(\mathbf{Q}\mathbf{Q}^T)\lambda_1(n^{-1}\sum_{i=1}^n(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T)}{\min_{j\in[p]} \hat{\sigma}_{jj}}.$$

By Lemma S.6 and Lemma 1 and Theorem 1 in the main text, the above quantity is further bounded by

$$\leq \frac{(1+\sqrt{\rho})^2+\epsilon_0+o_{a.s.}(1)}{1+o_{a.s.}(1)}.$$

That is,

$$\lambda_{K+1}(\mathbf{S}_n[\mathrm{diag}(\mathbf{S}_n)]^{-1}) \leq (1+\sqrt{\rho})^2+\epsilon_0+o_{a.s.}(1). \tag{S.66}$$

By (S.65) and (S.66), we have

$$\max_{k=K+1,\cdots,p} |\lambda_k(\hat{\mathbf{R}})-\lambda_k(\mathbf{S}_n)| = o_{a.s.}(1). \tag{S.67}$$

Step 2. Recall the definition of the Stieltjes transform $m_n(z)$ by (15) in the main text. Define similarly the Stieltjes transform of the empirical spectral distribution $F^{\mathbf{S}_n}(t) = p^{-1}\sum_{j=1}^p 1(\lambda_j(\mathbf{S}_n)\leq t)$ as

$$m_n^{\mathbf{S}_n}(z) = \int (t-z)^{-1} dF^{\mathbf{S}_n}(t).$$

By using $K = o(p^{1/6})$, it is obvious that

$$p^{-1}|\sum_{j=1}^K(\lambda_j(\mathbf{S}_n)-z)^{-1}| \leq Kp^{-1}v^{-1} \to 0,$$

for $z = u + \mathbf{i}v$ with $v > 0$. Therefore,

$$m_n^{\mathbf{S}_n}(z) = (p-K)^{-1}\sum_{j=K+1}^p(\lambda_j(\mathbf{S}_n)-z)^{-1} + o(1), \quad z \in \mathcal{C}^+,$$

Step 2.1. For all $z = u + \mathbf{i}v \in \mathcal{C}^+$, we have

$$\begin{aligned} m_n(z) - m_n^{\mathbf{S}_n}(z) &= (p-K)^{-1} \sum_{j=K+1}^{p} [(\lambda_j(\hat{\mathbf{R}}) - z)^{-1} - (\lambda_j(\mathbf{S}_n) - z)^{-1}] + o(1) \\ &= (p-K)^{-1} \sum_{j=K+1}^{p} \frac{\lambda_j(\mathbf{S}_n) - \lambda_j(\hat{\mathbf{R}})}{(\lambda_j(\hat{\mathbf{R}}) - z)(\lambda_j(\mathbf{S}_n) - z)} + o(1). \end{aligned}$$

Thus, we have

$$\begin{aligned} & |m_n(z) - m_n^{\mathbf{S}_n}(z)| \\ \le\; & \frac{\max_{j=K+1,\cdots,p} |\lambda_j(\mathbf{S}_n) - \lambda_j(\hat{\mathbf{R}})|}{\min_{j=K+1,\cdots,p} \sqrt{(\lambda_j(\hat{\mathbf{R}}) - u)^2 + v^2}\sqrt{(\lambda_j(\mathbf{S}_n) - u)^2 + v^2}} + o(1) \\ \le\; & \frac{\max_{j=K+1,\cdots,p} |\lambda_j(\mathbf{S}_n) - \lambda_j(\hat{\mathbf{R}})|}{v^2} + o(1). \end{aligned}$$

By (S.67), for $z = u + \mathbf{i}v \in \mathcal{C}^+$, we have

$$|m_n(z) - m_n^{\mathbf{S}_n}(z)| = o_{a.s.}(1). \tag{S.68}$$

Step 2.2. Letting $\bar{\boldsymbol{\epsilon}} = n^{-1}\sum_{i=1}^{n} \boldsymbol{\epsilon}_i$, $\bar{\mathbf{f}} = n^{-1}\sum_{i=1}^{n} \mathbf{f}_i$, we have

$$\mathbf{S}_n = (\mathbf{Q}_1, \mathbf{Q}_2) n^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T (\mathbf{Q}_1, \mathbf{Q}_2)^T = \mathbf{S}_{11} + \mathbf{S}_{12} + \mathbf{S}_{21} + \mathbf{S}_{22},$$

where $\mathbf{S}_{11} = \mathbf{Q}_1 n^{-1} \sum_{i=1}^{n} (\mathbf{f}_i - \bar{\mathbf{f}})(\mathbf{f}_i - \bar{\mathbf{f}})^T \mathbf{Q}_1^T$ and

$$\begin{aligned} \mathbf{S}_{12} &= \mathbf{Q}_1 n^{-1} \sum_{i=1}^{n} (\mathbf{f}_i - \bar{\mathbf{f}})[\Psi^{-1/2}(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})]^T \mathbf{Q}_2^T, \quad \mathbf{S}_{21} = \mathbf{S}_{12}^T, \\ \mathbf{S}_{22} &= \mathbf{Q}_2 n^{-1} \sum_{i=1}^{n} [\Psi^{-1/2}(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})][\Psi^{-1/2}(\boldsymbol{\epsilon}_i - \bar{\boldsymbol{\epsilon}})]^T \mathbf{Q}_2^T. \end{aligned}$$

Let $m_n^{\mathbf{S}_{22}}(z)$ be the Stieltjes transform of $\mathbf{S}_{22}$ as follows:

$$m_n^{\mathbf{S}_{22}}(z) = \int (t-z)^{-1} dF^{\mathbf{S}_{22}}(t) = p^{-1} \sum_{j=1}^{p} (\lambda_j(\mathbf{S}_{22}) - z)^{-1},$$

for $z \in \mathcal{C}^+$, where $F^{\mathbf{S}_{22}}(t) = p^{-1} \sum_{j=1}^{p} 1(\lambda_j(\mathbf{S}_{22}) \leq t)$. Then, using $|(t-z)^{-1}| \leq v^{-1}$ with $z = u + \mathbf{i}v, v > 0$,

$$|m_n^{\mathbf{S}_n}(z) - m_n^{\mathbf{S}_{22}}(z)| = \left| \int (t-z)^{-1} d(F^{\mathbf{S}_n}(t) - F^{\mathbf{S}_{22}}(t)) \right| \leq \frac{2\|F^{\mathbf{S}_n} - F^{\mathbf{S}_{22}}\|}{v}, \tag{S.69}$$

where $\|F^{\mathbf{S}_n} - F^{\mathbf{S}_{22}}\| = \sup_t |F^{\mathbf{S}_n}(t) - F^{\mathbf{S}_{22}}(t)|$. By Lemma S.3, we have

$$\|F^{\mathbf{S}_n} - F^{\mathbf{S}_{22}}\| \leq \frac{\text{rank}(\mathbf{S}_{11} + \mathbf{S}_{12} + \mathbf{S}_{21})}{p} \leq \frac{3K}{p} = o(1), \tag{S.70}$$

by Assumption (d) in the main text. Combination of (S.69) and (S.70), we conclude that for $z = u + \mathbf{i}v$ with $v > 0$

$$|m_n^{\mathbf{S}_n}(z) - m_n^{\mathbf{S}_{22}}(z)| = o(1). \tag{S.71}$$

Step 2.3. Note that $\mathbf{R} = \mathbf{Q}_2\mathbf{Q}_2^T + \mathbf{Q}_1\mathbf{Q}_2^T + \mathbf{Q}_2\mathbf{Q}_1^T + \mathbf{Q}_1\mathbf{Q}_1^T$ with $\text{rank}(\mathbf{Q}_1\mathbf{Q}_2^T) \leq K$, $\text{rank}(\mathbf{Q}_2\mathbf{Q}_1^T) \leq K$ and $\text{rank}(\mathbf{Q}_1\mathbf{Q}_1^T) \leq K$. By Lemma S.3, we have

$$\sup_t |F^{\mathbf{Q}_2\mathbf{Q}_2^T}(t) - F^{\mathbf{R}}(t)| \leq 3Kp^{-1} = o(1).$$

Moreover, $\sup_t |F^{\mathbf{R}}(t) - (p-K)^{-1} \sum_{j=K+1}^{p} 1(\lambda_j(\mathbf{R}) \leq t)| = o(1)$. By Assumption (e), $H(t)$ is the limit of $(p-K)^{-1} \sum_{j=K+1}^{p} 1(\lambda_j(\mathbf{R}) \leq t)$. Thus, $H(t)$ is also the limit of the empirical spectral distribution $F^{\mathbf{Q}_2\mathbf{Q}_2^T}(t)$.

Step 2.4. Under Assumption (a)-(b)-(c)-(d)-(e) and by Silverstein and Choi (1995), we have $|m_n^{\mathbf{S}_{22}} - m(z)| = o_{a.s.}(1)$, where

$$z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{t dH^{\mathbf{Q}_2\mathbf{Q}_2^T}(t)}{1 + t\underline{m}(z)} = -\frac{1}{\underline{m}(z)} + \rho \int \frac{t dH(t)}{1 + t\underline{m}(z)}, \; z \in \mathcal{C}^+,$$

where $H^{\mathbf{Q}_2\mathbf{Q}_2^T}(t)$ is the limit of the empirical spectral distribution $F^{\mathbf{Q}_2\mathbf{Q}_2^T}$. By (S.68)-(S.71), we have $|m_n(z) - m_n^{\mathbf{S}_{22}}(z)| = o_{a.s.}(1)$. Thus, we have $|m_n(z) - m(z)| = o_{a.s.}(1)$. That is, $m(z)$ is the limit of $m_n(z)$.

Step 2.5. Letting $\psi(x) = 1 + \rho \int \frac{t}{x-t} dH(t)$, we have

$$z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1 + t\underline{m}(z)} = -\underline{m}^{-1}(z)\psi(-\underline{m}^{-1}(z)).$$

Then the Stieltjes transform of the limiting spectral distribution from the eigenvalues $\lambda_{K+1}(\hat{\mathbf{R}}), \cdots, \lambda_p(\hat{\mathbf{R}})$ of $\hat{\mathbf{R}}$ also satisfies

$$z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1 + t\underline{m}(z)} = -\underline{m}^{-1}(z)\psi(-\underline{m}^{-1}(z)). \tag{S.72}$$

This finishes the proof of Theorem 2.

S.5 Lemma S.9 and its Proof

The proof of Theorem 3 in the main text requires the follow lemma.

Lemma S.9 *For the high dimensional factor model (7) satisfying Conditions C1-C2-C3-C4-C5 and Assumptions (a)-(b)-(c)-(d)-(e) in the main text, we have*

$$\lambda_j(\hat{\mathbf{R}}) > \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})) + \epsilon_2, a.s., \quad j \in [K],$$

for a very small positive constant ϵ_2 if $\lambda_K(\mathbf{R}) > \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho}) + \epsilon_0$ for a very small positive constant ϵ_0.

Proof. From (16), for $z \in \mathcal{C}^+$, we have

$$z \;=\; -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1 + t\underline{m}(z)} = [-\underline{m}(z)]^{-1}\psi([-\underline{m}(z)]^{-1}), \tag{S.73}$$

where $H(t)$ is the limiting spectral distribution of the empirical spectral distribution $H_{p-K}(t) = (p-K)^{-1}\sum_{j=K+1}^{p} 1(\lambda_j(\mathbf{R}) \le t)$ and $\psi(x) = 1 + \rho\int t(x-t)^{-1}dH(t)$. Let $z = x + \mathbf{i}\nu$ and $\underline{m}(z) = m_1(z) + \mathbf{i}m_2(z)$. The proof of Lemma S.9 consists of the following three steps.

- Step 1 is to prove $\underline{m}'(x) > 0$ for x outside the support set of $\underline{F}(t)$;

- Step 2 is to prove for x outside the support set of $\underline{F}$,

$$[\underline{m}(x)]^{-2} - \rho\int t^2(1+t\underline{m}(x))^{-2}dH(t) > 0$$

  and for x_0 being the right edge of the support set of $\underline{F}$,

$$[\underline{m}(x_0)]^{-2} - \rho\int t^2(1+t\underline{m}(x_0))^{-2}dH(t) = 0;$$

- Step 3 is to prove that as $\lambda_K(\mathbf{R}) \ge \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho}) + \epsilon_0$, for $j \in [K]$,

$$\lambda_j(\hat{\mathbf{R}}) > \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})) + \epsilon_2, a.s.$$

Proof of Step 1: Note that

$$\underline{m}(z) = \int \frac{1}{t-z}d\underline{F}(t) = \int \frac{t-x}{(t-x)^2+\nu^2}d\underline{F}(t) + \mathbf{i}\int \frac{\nu}{(t-x)^2+\nu^2}d\underline{F}(t).$$

For x outside the support set of $\underline{F}(t)$, we have

$$\begin{aligned} m_2(z) > 0, &\qquad m_2(x+\mathbf{i}0) = 0, \\ \lim_{\nu\to 0}\frac{m_1(x+\mathbf{i}\nu) - m_1(x+\mathbf{i}0)}{\nu} = 0, &\qquad \lim_{\nu\to 0}\frac{m_2(x+\mathbf{i}\nu)}{\nu} > 0. \end{aligned} \tag{S.74}$$

By (S.74), we have

$$\begin{aligned} \underline{m}'(x) = \underline{m}'(x+\mathbf{i}0) &= \lim_{\nu\to 0}\frac{\underline{m}(x+\mathbf{i}\nu) - \underline{m}(x+\mathbf{i}0)}{\mathbf{i}\nu} \\ &= \lim_{\nu\to 0}\frac{\mathbf{i}[m_2(x+\mathbf{i}\nu) - m_2(x+\mathbf{i}0)]}{\mathbf{i}\nu} \\ &= \lim_{\nu\to 0}\frac{m_2(x+\mathbf{i}\nu)}{\nu} > 0. \end{aligned} \tag{S.75}$$

Proof of Step 2: By (S.73), we have

$$z = \frac{-m_1(z) + \mathbf{i}m_2(z)}{[m_1(z)]^2 + [m_2(z)]^2} + \rho \int \frac{t(1 + tm_1(z) - \mathbf{i}tm_2(z))dH(t)}{(1 + tm_1(z))^2 + t^2 m_2(z)^2}.$$

Letting $z = x + \mathbf{i}\nu$, we have

$$\nu = \frac{m_2(z)}{[m_1(z)]^2 + [m_2(z)]^2} - \rho \int \frac{t^2 m_2(z)dH(t)}{(1 + tm_1(z))^2 + t^2[m_2(z)]^2},$$

$$\frac{\nu}{m_2(z)} = \frac{1}{[m_1(z)]^2 + [m_2(z)]^2} - \rho \int \frac{t^2 dH(t)}{(1 + tm_1(z))^2 + t^2[m_2(z)]^2}.$$

Thus, we have

$$\begin{aligned} & \underline{m}'(x) = \lim_{\nu \to 0} \frac{m_2(z)}{\nu} \\ = & \lim_{\nu \to 0} \left\{ \frac{1}{[m_1(z)]^2 + [m_2(z)]^2} - \rho \int \frac{t^2 dH(t)}{(1 + tm_1(z))^2 + t^2[m_2(z)]^2} \right\}^{-1} \\ = & \lim_{\nu \to 0} \frac{1}{[m_1(x)]^{-2} - \rho \int t^2[1 + tm_1(x)]^{-2} dH(t)} \\ = & \lim_{\nu \to 0} \frac{1}{[\underline{m}(x)]^{-2} - \rho \int t^2[1 + t\underline{m}(x)]^{-2} dH(t)}. \end{aligned} \tag{S.76}$$

From (S.75) and (S.76), for x outside the support set of $\underline{F}$, we have

$$[\underline{m}(x)]^{-2} - \rho \int t^2(1 + t\underline{m}(x))^{-2} dH(t) > 0. \tag{S.77}$$

Similarly, we obtain that for x_0 being the right edge of the support set of $\underline{F}$,

$$[\underline{m}(x_0)]^{-2} - \rho \int t^2(1 + t\underline{m}(x_0))^{-2} dH(t) = 0, \tag{S.78}$$

which leads to

$$-[\underline{m}(x_0)]^{-1} \leq \lambda_{K+1}(\mathbf{R})(1 + \sqrt{\rho}). \tag{S.79}$$

Proof of Step 3: By (S.77) and (S.78), it can easily be proved that

$$
\begin{aligned}
&\partial[t\psi(t)]/\partial t > 0, \qquad \text{if } t > -[\underline{m}(x_0)]^{-1}, \\
&\partial[t\psi(t)]/\partial t = 0, \qquad \text{if } t = -[\underline{m}(x_0)]^{-1}.
\end{aligned} \tag{S.80}
$$

Let $[-\underline{m}(x_1)]^{-1} = \lambda_{K+1}(1+\sqrt{\rho}) + \epsilon_0$. By Theorem 1.2 of Bai and Silverstein (1999), if $\lambda_K(\mathbf{R}) \geq -[\underline{m}(x_1)]^{-1} > -[\underline{m}(x_0)]^{-1}$, then for $j \in [K]$,

$$
\hat{\lambda}_j(\mathbf{R}) \geq x_1, \quad a.s. \tag{S.81}
$$

By (S.81) and recalling $z = [-\underline{m}(z)]^{-1}\psi([-\underline{m}(z)]^{-1})$, for $j \in [K]$, we have

$$
\begin{aligned}
\hat{\lambda}_j(\mathbf{R}) &\geq -[\underline{m}(x_1)]^{-1}\psi(-[\underline{m}(x_1)]^{-1}), a.s. \\
&> \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})) + \epsilon_2 \\
&\geq -[\underline{m}(x_0)]^{-1}\psi(-[\underline{m}(x_0)]^{-1}) + \epsilon_2 \\
&= x_0 + \epsilon_2,
\end{aligned}
$$

with a very small positive constant ϵ_2 where the third inequality is from (S.80) and the last equality is (S.73).

S.6 Proof of Theorem 3

By (S.64), we have $\max_{k\in[p]} \left|\frac{\lambda_k(\mathbf{S}_n)}{\lambda_k(\hat{\mathbf{R}})} - 1\right| = o_{a.s.}(1)$. Therefore, we only consider the convergence of $\lambda_k(\mathbf{S}_n)$ for $j \in [K]$. By the singular value decomposition, the $p \times (p+K)$ dimensional matrix $\mathbf{Q}$, defined in (9) in the main text, can be decomposed as

$$
\mathbf{Q} = \mathbf{CDV}
$$

where $\mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2)$ is an orthogonal matrix with $\mathbf{C}_1$ being of $p \times K$ dimension and $\mathbf{C}_2$ being of $p \times (p - K)$ dimension, $\mathbf{D} = \text{diag}(\mathbf{D}_1, \mathbf{D}_2)$ with $\mathbf{D}_1$ being $K \times K$ dimension and $\mathbf{D}_2$ being $(p-K) \times p$ dimension, and $\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{pmatrix}$ is an orthogonal matrix with $\mathbf{V}_1$ being $K \times (p + K)$ dimension and $\mathbf{V}_2$ being $p \times (p + K)$ dimension. With the above notation, we have

$$\mathbf{C}^T \mathbf{R} \mathbf{C} = \mathbf{D}\mathbf{D}^T = \text{diag}(\lambda_1(\mathbf{R}), \cdots, \lambda_p(\mathbf{R})),$$

where

$$\mathbf{D}_1^2 = \text{diag}(\lambda_1(\mathbf{R}), \cdots, \lambda_K(\mathbf{R})), \ \mathbf{D}_2\mathbf{D}_2^T = \text{diag}(\lambda_{K+1}(\mathbf{R}), \cdots, \lambda_p(\mathbf{R})). \tag{S.82}$$

By Lemma S.9, as $\lambda_K(\mathbf{R}) \geq \lambda_{K+1}(1 + \sqrt{\rho}) + \epsilon_0$, we have

$$P(\hat{\lambda}_j > \lambda_{K+1}(1 + \sqrt{\rho})\psi(\lambda_{K+1}(1 + \sqrt{\rho})) + \epsilon_2) \to 1, \tag{S.83}$$

with very small positive constants ϵ_0, ϵ_2, for $j \leq K$. In order to prove the convergence of $\hat{\lambda}_j, j \in [K]$, we seperate them into the following four steps.

- Step 1: To prove that $|\mathbf{S}_n - \hat{\lambda}_j \mathbf{I}_p| = 0$ leads to $|\mathbf{K}_n(\hat{\lambda}_j) + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0$ in probability 1 with $\mathbf{1}_K$ being a vector with K elements 1 where

$$\mathbf{K}_n(\hat{\lambda}_j) = \mathbf{D}_1^{-1}[\mathbf{C}_1^T\mathbf{B}_n\mathbf{C}_1 + \mathbf{C}_1^T\mathbf{B}_n\mathbf{C}_2(\hat{\lambda}_j\mathbf{I}_p - \mathbf{C}_2^T\mathbf{B}_n\mathbf{C}_2)^{-1}\mathbf{C}_2^T\mathbf{B}_n\mathbf{C}_1 - \hat{\lambda}_j\mathbf{I}_K]\mathbf{D}_1^{-1};$$

- Step 2: To prove that $|\mathbf{K}_n(\hat{\lambda}_j) + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0$ leads to

$$|\hat{\lambda}_j\mathbf{D}_1^{-2} + \hat{\lambda}_j n^{-1}\mathbf{V}_1\mathbf{X}(\boldsymbol{\eta}^T\boldsymbol{\eta} - \hat{\lambda}_j\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{V}_1^T + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0$$

  where $\boldsymbol{\eta} = n^{-1/2}\mathbf{D}_2\mathbf{V}_2(\mathbf{x}_1, \ldots, \mathbf{x}_n)$;

- Step 3. To prove that $\hat{\lambda}_j\lambda_j^{-1} + n^{-1}\text{tr}[(\theta_j^{-1}\boldsymbol{\eta}^T\boldsymbol{\eta} - \mathbf{I}_n)^{-1}] = o_p(1)$;

- Step 4. To prove $\frac{\hat{\lambda}_j}{\lambda_j} = \psi(\lambda_j) + o_p(1)$.

Proof of Step 1. We have

$$\mathbf{C}^T\mathbf{S}_n\mathbf{C} = \begin{pmatrix} \mathbf{C}_1^T\mathbf{S}_n\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{S}_n\mathbf{C}_2 \\ \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_1 & \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2 \end{pmatrix}.$$

Using $\mathbf{C}_2^T\mathbf{Q} = \mathbf{C}_2^T\mathbf{C}\mathbf{D}\mathbf{V} = (\mathbf{0}, \mathbf{D}_2)\mathbf{V}$, we have

$$\mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2 = (\mathbf{0}, \mathbf{D}_2)\mathbf{V}n^{-1}\sum_{j=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T\mathbf{V}^T(\mathbf{0}, \mathbf{D}_2)^T.$$

By using this special structure, we have

$$\begin{aligned} \lambda_1(\mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2) &\leq \lambda_1((\mathbf{0}, \mathbf{D}_2\mathbf{D}_2^T)n^{-1}\sum_{j=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T) \\ &\leq \lambda_{K+1}(\mathbf{R})(1+\sqrt{\rho})\psi(\lambda_{K+1}(1+\sqrt{\rho})) + o_p(1), \end{aligned}$$

for any small positive constant $\epsilon_0 > 0$ where the last inequality follows from Lemma S.6. Therefore, by (S.83) $|\hat{\lambda}_j\mathbf{I}_{p-K} - \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2| \neq 0$ for $j \in [K]$ in probability 1. Let

$$\tilde{\mathbf{K}}_n(\hat{\lambda}_j) = \mathbf{C}_1^T\mathbf{S}_n\mathbf{C}_1 + \mathbf{C}_1^T\mathbf{S}_n\mathbf{C}_2(\hat{\lambda}_j\mathbf{I}_p - \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2)^{-1}\mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_1 - \hat{\lambda}_j\mathbf{I}_K.$$

By matrix factorization, we obtain from $|\mathbf{C}^T\mathbf{S}_n\mathbf{C} - \hat{\lambda}_j\mathbf{I}_p| = 0$ that $|\tilde{\mathbf{K}}_n(\hat{\lambda}_j)| = 0$ in probability 1. Since $K = o(n^{1/6})$, it can be shown that

$$\begin{aligned} &\mathrm{E}\|\mathbf{D}_1^{-1}\mathbf{C}_1^T\mathbf{Q}\bar{\mathbf{x}}\bar{\mathbf{x}}^T\mathbf{Q}^T\mathbf{C}_1\mathbf{D}_1^{-1}\| \leq K^2n^{-1} = o(K^{-2}), \\ &\|\mathbf{D}_1^{-1}\mathbf{C}_1^T\mathbf{Q}\bar{\mathbf{x}}\bar{\mathbf{x}}^T\mathbf{Q}^T\mathbf{C}_2(\hat{\lambda}_j\mathbf{I}_p - \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2)^{-1}\mathbf{C}_2^T\bar{\mathbf{x}}\bar{\mathbf{x}}^T\mathbf{C}_1\mathbf{D}_1^{-1}\| = O_p(K^2n^{-1}) = o_p(K^{-2}), \\ &\|\mathbf{D}_1^{-1}\mathbf{C}_1^T\mathbf{Q}\bar{\mathbf{x}}\bar{\mathbf{x}}^T\mathbf{Q}^T\mathbf{C}_2(\hat{\lambda}_j\mathbf{I}_p - \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2)^{-1}\mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_1\mathbf{D}_1^{-1}\| = O_p(Kn^{-1/2}) = o_p(K^{-2}), \\ &\|\mathbf{D}_1^{-1}\mathbf{C}_1^T\mathbf{B}_n\mathbf{C}_2[(\hat{\lambda}_j\mathbf{I}_p - \mathbf{C}_2^T\mathbf{B}_n\mathbf{C}_2)^{-1} - (\hat{\lambda}_j\mathbf{I}_p - \mathbf{C}_2^T\mathbf{S}_n\mathbf{C}_2)^{-1}]\mathbf{C}_2^T\mathbf{B}_n\mathbf{C}_1\mathbf{D}_1^{-1}\| = o_p(K^{-2}). \end{aligned}$$

Then $|\mathbf{K}_n(\hat{\lambda}_j) + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0$.

Proof of Step 2. Let us now reexpress the matrix $\mathbf{K}_n(\hat{\lambda}_j)$. Note that from the definition, we have $\mathbf{D}_1^{-1}\mathbf{C}_1^T\mathbf{Q} = \mathbf{V}_1$ and $\mathbf{C}_2^T\mathbf{Q} = \mathbf{D}_2\mathbf{V}_2$. Let the $K \times n$ dimensional matrix be $\boldsymbol{\xi} = n^{-1/2}\mathbf{V}_1\mathbf{X}$ with $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Recall $\boldsymbol{\eta} = n^{-1/2}\mathbf{D}_2\mathbf{V}_2\mathbf{X}$. Then, we have

$$
\begin{aligned}
\mathbf{D}_1^{-1}\mathbf{C}_1^T\mathbf{B}_n\mathbf{C}_1\mathbf{D}_1^{-1} &= n^{-1}\mathbf{V}_1\mathbf{X}\mathbf{X}^T\mathbf{V}_1^T = \boldsymbol{\xi}\boldsymbol{\xi}^T, \\
\mathbf{C}_2^T\mathbf{B}_n\mathbf{C}_1\mathbf{D}_1^{-1} &= n^{-1}\mathbf{D}_2\mathbf{V}_2\mathbf{X}\mathbf{X}^T\mathbf{V}_1^T = \boldsymbol{\eta}\boldsymbol{\xi}^T, \\
\mathbf{C}_2^T\mathbf{B}_n\mathbf{C}_2 &= n^{-1}\mathbf{D}_2\mathbf{V}_2\mathbf{X}\mathbf{X}^T\mathbf{V}_2^T\mathbf{D}_2^T = \boldsymbol{\eta}\boldsymbol{\eta}^T.
\end{aligned}
$$

Letting $\mathbf{A}_n = \boldsymbol{\eta}^T(\hat{\lambda}_j\mathbf{I}_{p-K} - \boldsymbol{\eta}\boldsymbol{\eta}^T)^{-1}\boldsymbol{\eta} = \hat{\lambda}_j(\hat{\lambda}_j\mathbf{I}_n - \boldsymbol{\eta}^T\boldsymbol{\eta})^{-1} - \mathbf{I}_n$, then we have

$$
\mathbf{K}_n(\hat{\lambda}_j) = \boldsymbol{\xi}\boldsymbol{\xi}^T + \boldsymbol{\xi}\mathbf{A}_n\boldsymbol{\xi}^T - \hat{\lambda}_j\mathbf{D}_1^{-2} = \hat{\lambda}_j\boldsymbol{\xi}(\hat{\lambda}_j\mathbf{I}_n - \boldsymbol{\eta}^T\boldsymbol{\eta})^{-1}\boldsymbol{\xi}^T - \hat{\lambda}_j\mathbf{D}_1^{-2}.
$$

Thus

$$
\begin{aligned}
& |\mathbf{K}_n(\hat{\lambda}_j) + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0 \\
\iff & |\hat{\lambda}_j\mathbf{D}_1^{-2} + \hat{\lambda}_j\boldsymbol{\xi}(\boldsymbol{\eta}^T\boldsymbol{\eta} - \hat{\lambda}_j\mathbf{I}_n)^{-1}\boldsymbol{\xi}^T + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0 \\
\iff & |\hat{\lambda}_j\mathbf{D}_1^{-2} + \hat{\lambda}_j n^{-1}\mathbf{V}_1\mathbf{X}(\boldsymbol{\eta}^T\boldsymbol{\eta} - \hat{\lambda}_j\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{V}_1^T + o_p(K^{-2})\mathbf{1}_K\mathbf{1}_K^T| = 0.
\end{aligned}
$$

Proof of Step 3. Let us simplify the second term in the last step. For each given λ, it is easy to prove that

$$
\begin{aligned}
& \sum_{j=1}^{K}\{\lambda n^{-1}\mathbf{e}_j^T\mathbf{V}_1\mathbf{X}(\boldsymbol{\eta}^T\boldsymbol{\eta} - \lambda\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{V}_1^T\mathbf{e}_j - \lambda n^{-1}\mathrm{tr}[(\boldsymbol{\eta}^T\boldsymbol{\eta} - \lambda\mathbf{I}_n)^{-1}]\}^2 \\
& + \sum_{1 \le i \ne j \le K}[\lambda n^{-1}\mathbf{e}_i^T\mathbf{V}_1\mathbf{X}(\boldsymbol{\eta}^T\boldsymbol{\eta} - \lambda\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{V}_1^T\mathbf{e}_j]^2 = o_p(1),
\end{aligned}
$$

using Lemma S.7 for $\lambda > \lambda_{K+1}(\mathbf{R})(1 + \sqrt{\rho})\psi(\lambda_{K+1}(\mathbf{R})(1 + \sqrt{\rho})) + \epsilon_2$ with a very small positive constant ϵ_2. Thus,

$$
\hat{\lambda}_j n^{-1}\mathbf{V}_1\mathbf{X}(\boldsymbol{\eta}^T\boldsymbol{\eta} - \hat{\lambda}_j\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{V}_1^T - \hat{\lambda}_j n^{-1}\mathrm{tr}[(\boldsymbol{\eta}^T\boldsymbol{\eta} - \hat{\lambda}_j\mathbf{I}_n)^{-1}]\mathbf{I}_K = O_p(n^{-1/2})\mathbf{1}_K\mathbf{1}_K^T,
$$

and

$$
\begin{aligned}
& \hat{\lambda}_j \mathbf{D}_1^{-2} + \hat{\lambda}_j n^{-1} \mathbf{V}_1 \mathbf{X} (\boldsymbol{\eta}^T \boldsymbol{\eta} - \hat{\lambda}_j \mathbf{I}_n)^{-1} \mathbf{X}^T \mathbf{V}_1^T \\
= \;& \hat{\lambda}_j \mathrm{diag}(\lambda_1^{-1}, ..., \lambda_p^{-1}) + \hat{\lambda}_j n^{-1} \mathrm{tr}[(\boldsymbol{\eta}^T \boldsymbol{\eta} - \hat{\lambda}_j \mathbf{I}_n)^{-1}] \mathbf{I}_K + O_p(n^{-1/2}) \mathbf{1}_K \mathbf{1}_K^T.
\end{aligned}
$$

Obviously,

$$
\begin{aligned}
& |\hat{\lambda}_j \mathbf{D}_1^{-2} + \hat{\lambda}_j n^{-1} \mathbf{V}_1 \mathbf{X} (\boldsymbol{\eta}^T \boldsymbol{\eta} - \hat{\lambda}_j \mathbf{I}_n)^{-1} \mathbf{X}^T \mathbf{V}_1^T + o_p(1) \mathbf{1}_K \mathbf{1}_K^T| = 0 \\
\Longleftrightarrow \;& |\hat{\lambda}_j \mathrm{diag}(\lambda_1^{-1}, ..., \lambda_p^{-1}) + \hat{\lambda}_j n^{-1} \mathrm{tr}[(\boldsymbol{\eta}^T \boldsymbol{\eta} - \hat{\lambda}_j \mathbf{I}_n)^{-1}] \mathbf{I}_K + O_p(n^{-1/2}) \mathbf{1}_K \mathbf{1}_K^T| = 0.
\end{aligned}
$$

The determinant $|O_p(n^{-1/2}) \mathbf{1}_K \mathbf{1}_K^T| = o_P(1)$. Therefore, there must exist λ_j (an eigenvalue of $\mathbf{R}$) satisfying

$$
\hat{\lambda}_j \lambda_j^{-1} + n^{-1} \mathrm{tr}[(\hat{\lambda}_j^{-1} \boldsymbol{\eta}^T \boldsymbol{\eta} - \mathbf{I}_n)^{-1}] + O_p(n^{-1/2}) = 0. \tag{S.84}
$$

Proof of Step 4. By Silverstein and Choi (1995), we have $n^{-1}\mathrm{tr}[(\boldsymbol{\eta}^T \boldsymbol{\eta} - z\mathbf{I}_n)^{-1}] \to \underline{m}(z)$, *a.s.* for $z \in \mathcal{C}^+$ where $z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{t dH(t)}{1 + t\underline{m}(z)}$. In the following, we will prove $\hat{\lambda}_j \lambda_j^{-1} = \psi(\lambda_j) + o_p(1)$. Let θ_j be the limit of $\hat{\lambda}_j$ in the sense that $\hat{\lambda}_j / \theta_j = 1 + o_p(1)$, which exists by the result of Step 3. Then, we have

$$
\begin{aligned}
& n^{-1} \mathrm{tr}[(\hat{\lambda}_j^{-1} \boldsymbol{\eta}^T \boldsymbol{\eta} - \mathbf{I}_n)^{-1}] - n^{-1} \mathrm{tr}[(\theta_j^{-1} \boldsymbol{\eta}^T \boldsymbol{\eta} - \mathbf{I}_n)^{-1}] \\
= \;& -(\hat{\lambda}_j^{-1} - \theta_j^{-1}) \boldsymbol{\eta}^T \boldsymbol{\eta} n^{-1} \mathrm{tr}[(\hat{\lambda}_j^{-1} \boldsymbol{\eta}^T \boldsymbol{\eta} - \mathbf{I}_n)^{-1} (\theta_j^{-1} \boldsymbol{\eta}^T \boldsymbol{\eta} - \mathbf{I}_n)^{-1}] = o_p(1).
\end{aligned}
$$

Combining this with (S.84), we have

$$
\hat{\lambda}_j \lambda_j^{-1} + n^{-1} \mathrm{tr}[(\theta_j^{-1} \boldsymbol{\eta}^T \boldsymbol{\eta} - \mathbf{I}_n)^{-1}] = o_p(1).
$$

This leads to $\hat{\lambda}_j \lambda_j^{-1} + \theta_j \underline{m}(\theta_j) = o_p(1)$. That is, $\lambda_j \hat{\lambda}_j^{-1} = -[\theta_j \underline{m}(\theta_j)]^{-1} + o_p(1)$. Theorem 2 in the main text have proved $\underline{m}_n(z) - \underline{m}(z) = o_{a.s.}(1)$. It is easily proved that $\underline{m}_{n,j}(z) - \underline{m}_n(z) = o_{a.s.}(1)$. Then we have

$$
\underline{m}_{n,j}(z) - \underline{m}(z) = o_{a.s.}(1).
$$

That is, $\underline{m}(z)$ can be estimated by $\underline{m}_{n,j}(z)$. That is,

$$\lambda_j \hat{\lambda}_j^{-1} = -[\hat{\lambda}_j \underline{m}_{n,j}(\hat{\lambda}_j)]^{-1} + o_p(1) = \hat{\lambda}_j^{-1}\lambda_j^C + o_p(1),$$

or

$$\frac{\lambda_j^C}{\lambda_j} = 1 + o_p(1).$$

In fact, Bai and Ding (2012) first proved the result. But they required the block diagonal structure of the covariance matrix $\mathbf{\Sigma}$ and λ_j being bounded. However, our paper doesn't require the condition. By $z = -\frac{1}{\underline{m}(z)} + \rho \int \frac{tdH(t)}{1+t\underline{m}(z)}$, we have

$$\hat{\lambda}_j = -\frac{1}{\underline{m}(\hat{\lambda}_j)} + \rho \int \frac{-[\underline{m}(\hat{\lambda}_j)]^{-1} tdH(t)}{-[\underline{m}(\hat{\lambda}_j)]^{-1} - t}.$$

Combining the last three results, we have

$$\begin{aligned}
\frac{\hat{\lambda}_j}{\lambda_j} &= -\frac{\hat{\lambda}_j}{\lambda_j} \frac{1}{\hat{\lambda}_j \underline{m}(\hat{\lambda}_j)} + \rho \int \frac{-(\hat{\lambda}_j/\lambda_j)[\hat{\lambda}_j \underline{m}(\hat{\lambda}_j)]^{-1} tdH(t)}{-\lambda_j(\hat{\lambda}_j/\lambda_j)[\hat{\lambda}_j \underline{m}(\hat{\lambda}_j)]^{-1} - t} \\
&= -\frac{\hat{\lambda}_j}{\lambda_j} \frac{1}{\hat{\lambda}_j \underline{m}_{n,j}(\hat{\lambda}_j)} + \rho \int \frac{-(\hat{\lambda}_j/\lambda_j)[\hat{\lambda}_j \underline{m}_{n,j}(\hat{\lambda}_j)]^{-1} tdH(t)}{-\lambda_j(\hat{\lambda}_j/\lambda_j)[\hat{\lambda}_j \underline{m}_{n,j}(\hat{\lambda}_j)]^{-1} - t} + o_p(1) \\
&= 1 + \rho \int \frac{tdH(t)}{\lambda_j - t} + o_p(1) \\
&= \psi(\lambda_j) + o_p(1).
\end{aligned}$$

The proof of Theorem 3 is completed.

S.7 Some additional simulation results

The simulation setup is the same as that in the main paper except Case 6 which is

Case 6: Let $b_{\ell j}$ be iid from $N(0,1)$ and $\nu_1^2, \cdots, \nu_p^2$ be iid from Unif$(0, 180)$.

In Tables S.1-S.3, simulation results will be presented for Cases 1-3 for uniform population in the main paper. In Table S.4, simulation results will be presented for Cases 6 for Gaussian population and uniform population.

Table S.1: Percentages of the estimated number of common factors for Case 1 with $n = 300$ in 1000 simulations: "$\hat{K} = K$", "$\hat{K} > K$" and "$\hat{K} < K$"

truly estimates, overestimates and underestimates the number of common factors, respectively. "ave($\hat{K}$)" is the average of the estimated number of common factors.

p		PC_3	ON_2	ER	GR	ACT
		Case 1 and Uniform population				
100	$\hat{K} = K$	99.9	99.9	44.7	81.2	100
	$\hat{K} > K$	0	0.1	0	0	0
	$\hat{K} < K$	0.1	0	55.3	18.8	0
	ave($\hat{K}$)	5	5	2.97	4.52	5
300	$\hat{K} = K$	93.3	100	4.2	9.0	100
	$\hat{K} > K$	0	0	0	0	0
	$\hat{K} < K$	6.7	0	95.8	91.0	0
	ave($\hat{K}$)	4.93	5	2.07	2.55	5
500	$\hat{K} = K$	0	99.9	0.1	0.3	99.2
	$\hat{K} > K$	0	0.1	0	0	0.8
	$\hat{K} < K$	100	0	99.9	99.7	0
	ave($\hat{K}$)	3.92	5	1.75	1.97	5.01
1000	$\hat{K} = K$	0	83.3	0	0	89.8
	$\hat{K} > K$	0	0.1	0	0	1.5
	$\hat{K} < K$	100	16.6	100	100	8.7
	ave($\hat{K}$)	1.82	4.84	1.32	1.37	4.93

References

Bai, Z. D. and X. Ding (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices: Theory and Applications 1*(2), 1150011–1–1150011–21.

Table S.2: Percentages of the estimated number of common factors for Case 2 with $n = 300$ in 1000 simulations: "$\hat{K} = K$", "$\hat{K} > K$" and "$\hat{K} < K$" truly estimates, overestimates and underestimates the number of common factors, respectively. "ave($\hat{K}$)" is the average of the estimated number of common factors.

p		PC_3	ON_2	ER	GR	ACT
		Uniform population				
100	$\hat{K} = K$	0	0	5.0	5.1	0
	$\hat{K} > K$	0	0	6.8	7.0	0
	$\hat{K} < K$	100	100	88.2	87.9	100
	ave($\hat{K}$)	1	1.27	2.33	2.35	1.08
300	$\hat{K} = K$	0	0.5	5.2	5.5	4.6
	$\hat{K} > K$	0	0	1.2	1.3	0.1
	$\hat{K} < K$	100	99.5	93.6	93.2	95.3
	ave($\hat{K}$)	1	2.87	2.25	2.28	2.92
500	$\hat{K} = K$	0	37.3	31.5	32.6	76.1
	$\hat{K} > K$	0	0.1	0.2	0.2	1.1
	$\hat{K} < K$	100	62.6	68.3	67.2	22.8
	ave($\hat{K}$)	1	4.26	3.08	3.13	4.76
1000	$\hat{K} = K$	0	99.8	94.5	94.7	96.8
	$\hat{K} > K$	0	0.1	0	0	3.2
	$\hat{K} < K$	100	0.1	5.5	5.3	0
	ave($\hat{K}$)	1	5	4.88	4.88	5.03

Table S.3: Percentages of the estimated number of common factors for Case 3 with $n = 300$ in 1000 simulations: "$\hat{K} = K$", "$\hat{K} > K$" and "$\hat{K} < K$" truly estimates, overestimates and underestimates the number of common factors, respectively. "ave($\hat{K}$)" is the average of the estimated number of common factors.

p		PC_3	ON_2	ER	GR	ACT
		Uniform population				
100	$\hat{K} = K$	0.3	1.3	4.7	5.0	96.0
	$\hat{K} > K$	0	0	2.4	2.8	0.5
	$\hat{K} < K$	99.7	98.7	92.9	92.2	3.5
	ave($\hat{K}$)	2.32	2.87	2.21	2.28	4.97
300	$\hat{K} = K$	99.6	98.8	87.7	88.7	99.6
	$\hat{K} > K$	0	0.1	0	0	0.4
	$\hat{K} < K$	0.4	1.1	12.3	11.3	0
	ave($\hat{K}$)	5	4.99	4.73	4.76	5
500	$\hat{K} = K$	67.1	99.8	99.8	99.8	99.7
	$\hat{K} > K$	0	0.2	0	0	0.3
	$\hat{K} < K$	32.9	0	0.2	0.2	0
	ave($\hat{K}$)	4.66	5	5	5	5
1000	$\hat{K} = K$	6.4	99.9	100	100	99.3
	$\hat{K} > K$	0	0.1	0	0	0.7
	$\hat{K} < K$	93.6	0	0	0	0
	ave($\hat{K}$)	3.71	5	5	5	5.01

Table S.4: Percentages of the estimated number of common factors for Case 6 with $n = 300$ in 1000 simulations: "$\hat{K} = K$", "$\hat{K} > K$" and "$\hat{K} < K$" truly estimates, overestimates and underestimates the number of common factors, respectively. "ave($\hat{K}$)" is the average of the estimated number of common factors.

p		PC_3	ON_2	ER	GR	ACT
		Gaussian population				
100	$\hat{K} = K$	0	0.1	4.2	4.4	64.3
	$\hat{K} > K$	0	0	6.6	7.3	0.10
	$\hat{K} < K$	100	99.9	89.2	88.3	35.6
	ave($\hat{K}$)	1.18	1.53	2.29	2.37	4.58
300	$\hat{K} = K$	47.0	31.2	27.0	28.2	98.9
	$\hat{K} > K$	0	0.1	0.4	0.4	1.1
	$\hat{K} < K$	53.0	68.7	72.6	71.4	0
	ave($\hat{K}$)	4.42	4.17	3.01	3.07	5.01
500	$\hat{K} = K$	0	98.8	88.9	89.7	98.9
	$\hat{K} > K$	0	0	0	0	1.1
	$\hat{K} < K$	100	1.2	11.1	10.3	0
	ave($\hat{K}$)	2.44	4.99	4.76	4.78	5.01
1000	$\hat{K} = K$	0	99.9	99.9	99.9	99.1
	$\hat{K} > K$	0	0.1	0	0	0.9
	$\hat{K} < K$	100	0	0.1	0.1	0
	ave($\hat{K}$)	1.17	5	5	5	5.01
		Uniform population				
100	$\hat{K} = K$	0	0.1	5.0	5.4	60.7
	$\hat{K} > K$	0	0.1	8.4	9.0	0.4
	$\hat{K} < K$	100	99.8	86.6	85.6	38.9
	ave($\hat{K}$)	1.17	1.57	2.38	2.45	4.54
300	$\hat{K} = K$	48.4	37.8	31.7	33.7	99.4
	$\hat{K} > K$	0	0	0.3	0.4	0.6
	$\hat{K} < K$	51.6	62.2	68.0	65.9	0
	ave($\hat{K}$)	4.45	4.27	3.16	3.25	5.01
500	$\hat{K} = K$	0	99.4	91.0	91.6	99.1
	$\hat{K} > K$	0	0.1	0	0	0.9
	$\hat{K} < K$	100	0.5	9.0	8.4	0
	ave($\hat{K}$)	2.44	5	4.81	4.83	5.01
1000	$\hat{K} = K$	0	99.9	100	100	99.0
	$\hat{K} > K$	0	0.1	0	0	1.0
	$\hat{K} < K$	100	0	0	0	0
	ave($\hat{K}$)	1.12	5	5	5	5.01

Bai, Z. D. and J. W. Silverstein (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *The Annals of Probability 27*(3), 1536–1555.

Bai, Z. D. and J. W. Silverstein (2010). *Spectral Analysis of Large Dimensional Random Matrices*. 1st Edition. Science Press.

Bai, Z. D. and J. F. Yao (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44*(3), 447–474.

Bao, Z. G., G. M. Pan, and W. Zhou (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics 43*(1), 382–421.

Jiang, D. and Z. D. Bai (2019). Generalized four moment theorem and an application to clt for spiked eigenvalues of large-dimensional covariance matrices. *arXiv: 1808.05362V2*.

Silverstein, W. J. and S. I. Choi (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* (54), 295–309.

Yin, Y. Q., Z. D. Bai, and P. R. Krishnaish (2013). Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Processes and their Applications 123*, 2899–2920.