# Supplement to

# Test Assets and Weak Factors


Stefano Giglio*
Yale School of Management
NBER and CEPR

Dacheng Xiu†
Booth School of Business
University of Chicago

Dake Zhang‡
Booth School of Business
University of Chicago


This Version: June 25, 2021


**Abstract**

This supplementary appendix provides model assumptions and mathematical proofs.



*Address: 165 Whitney Avenue, New Haven, CT 06520, USA. E-mail address: stefano.giglio@yale.edu.
†Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.
‡Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. Email: dkzhang@chicagobooth.edu.

# A Model Assumptions

To derive the asymptotic properties of the SPCA and alternative estimators, we need the following high-level assumptions, which can be easily verified by standard and more primitive assumptions. We start with assumptions that characterize the DGP of returns and factor proxies.

**Assumption A.1.** *The factor innovation $V$ satisfies:*

$$\|\bar{v}\| \lesssim_p T^{-1/2}, \quad \left\|T^{-1}VV^\intercal - \Sigma_v\right\| \lesssim_p T^{-1/2}, \quad \|V\|_{\text{MAX}} \lesssim_p \sqrt{\log T},$$

*where $\Sigma_v \in \mathbb{R}^{p\times p}$ is a positive-definite matrix with $\lambda_p(\Sigma_v) \gtrsim 1$ and $\lambda_1(\Sigma_v) \lesssim 1$.*

**Assumption A.2.** *The residual innovation $Z$ satisfies:*

$$\|\bar{z}\| \lesssim_p T^{-1/2}, \quad \left\|T^{-1}ZZ^\intercal - \Sigma_z\right\| \lesssim_p T^{-1/2}, \quad \|Z\|_{\text{MAX}} \lesssim_p \sqrt{\log T}.$$

*where $\Sigma_z \in \mathbb{R}^{d\times d}$ is a positive-definite matrix with $\lambda_d(\Sigma_z) \gtrsim 1$ and $\lambda_1(\Sigma_z) \lesssim 1$. In addition,*

$$\|ZV^\intercal\| \lesssim_p T^{1/2}.$$

Assumptions A.1 and A.2 impose rather weak conditions on the time series behavior of the factors and measurement error. Since $v_t$ and $z_t$ have a finite cross-sectional dimension, both assumptions hold if these processes are stationary, strong mixing, and satisfy some moment conditions.

**Assumption A.3.** *The factor loading matrix $\beta$ satisfies*

$$\|\beta\|_{\text{MAX}} \lesssim 1, \qquad \lambda_p(\beta_{[I_0]}^\intercal \beta_{[I_0]}) \gtrsim N_0,$$

*for some index set $I_0$, where $N_0 = |I_0|$.*

Assumption A.3 implies that there exists a subset of test assets, within which all latent factors are strong. Because the number of factors is finite, requiring *all* factors to be strong within a *common* index set $I_0$ is equivalent to requiring each factor to be strong in its own index set. One direction of the equivalence is trivial. To prove the other direction, suppose that for factor $i$, there exists an index set, $I_i$, in which this factor is strong, that is, $\lambda_1(\beta_{[I_i]}^\intercal \beta_{[I_i]}) \gtrsim |I_i|$. Then we can find $k^\star := \min_k |I_k|$, and build up $I_0$ from $I_{k^*}$ (so that $|I_0| \geq |I_{k^\star}|$) by adding randomly selected $|I_{k^\star}|$ number of assets from each $I_j, j = 1, 2, \ldots, p, j \neq k^\star$. The resulting index set $I_0$ contains at most $p \times |I_{k^\star}|$ number of test assets, barring from repeated counts. We thereby construct a common index set such that all factors are strong within this set.

Next, we need the following moment conditions.

**Assumption A.4.** *The idiosyncratic component $U$ satisfies:*

$$\|U\|_{\text{MAX}} \lesssim_p (\log T)^{1/2} + (\log N)^{1/2}, \quad \|\bar{u}\|_{\text{MAX}} \lesssim_p T^{-1/2}(\log N)^{1/2}.$$

*In addition, for any non-random subset $I \subset [N]$,*

$$\left\|U_{[I]}\right\| \lesssim_p |I|^{1/2} + T^{1/2}, \quad \left\|\bar{u}_{[I]}\right\| \lesssim_p |I|^{1/2} T^{-1/2}.$$

Assumption A.4 imposes restrictions on the time-series dependence and heteroskedasticity of $u_t$. The first two inequalities are results of some large deviation theorem, see, e.g., Fan et al. (2011). The last inequality can be shown by random matrix theory, see Bai and Silverstein (2009), if $u_t$ is i.i.d. both in time and in the cross-section.

**Assumption A.5.** *For any non-random subset $I \subset [N]$, the factor loading $\beta_{[I]}$ and the idiosyncratic error $U_{[I]}$ satisfy the following conditions:*

$$(i) \quad \left\|(\beta_{[I]}^\intercal \beta_{[I]})^{-1/2} \beta_{[I]}^\intercal U_{[I]}\right\| \lesssim_p T^{1/2}.$$
$$(ii) \quad \left\|(\beta_{[I]}^\intercal \beta_{[I]})^{-1/2} \beta_{[I]}^\intercal U_{[I]} \iota_T\right\| \lesssim_p T^{1/2}.$$

*If $\beta_{[I]}^\intercal \beta_{[I]}$ is singular, we need replace the matrix inverse above by the Moore-Penrose inverse.*

**Assumption A.6.** *The following conditions hold for $U$, $V$, $\beta$, and any non-random subset $I \subset [N]$:*

$$(i) \quad \left\|U_{[I]} V^\intercal\right\| \lesssim_p |I|^{1/2} T^{1/2}, \quad \left\|U_{[I]} V^\intercal\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2} T^{1/2}.$$
$$(ii) \quad \left\|(\beta_{[I]}^\intercal \beta_{[I]})^{-1/2} \beta_{[I]}^\intercal U_{[I]} V^\intercal\right\| \lesssim_p T^{1/2}.$$

**Assumption A.7.** *The following conditions hold for $U$, $Z$, $\beta$, and any non-random subset $I \subset [N]$:*

$$(i) \quad \left\|U_{[I]} Z^\intercal\right\| \lesssim_p |I|^{1/2} T^{1/2}, \quad \left\|U_{[I]} Z^\intercal\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2} T^{1/2}.$$
$$(ii) \quad \left\|(\beta_{[I]}^\intercal \beta_{[I]})^{-1/2} \beta_{[I]}^\intercal U_{[I]} Z^\intercal\right\| \lesssim_p T^{1/2}.$$

Assumptions A.5 - A.7 resemble Assumptions A.7, A.9, and A.10 of Giglio and Xiu (2021), except that here we impose their stronger versions which hold for any non-random subset $I \subset [N]$. Of course, these two sets of assumptions are equivalent if $u_t$ is identically distributed along the cross-sectional dimension.

In the main text, we denote the selected subsets in the SPCA procedure as $\widehat{I}_k$, $k = 1, 2, \ldots$. We now define their population counterparts. For simplicity, we consider the case $\Sigma_v = \mathbb{I}_p$ here. In general case, replace $\beta$ and $\eta$ by $\beta' = \beta \Sigma_v^{1/2}$ and $\eta' = \eta \Sigma_v^{1/2}$ in the following definiiton. In detail, we start with $a_i^{(1)} := \left\|\beta_{[i]} \eta^\intercal\right\|_{\text{MAX}}$ and define $I_1 := \{a_i^{(1)} \geq c_{qN}^{(1)}\}$, where $c_{qN}^{(1)}$ is the $(qN)$th largest value in $\left\{a_i^{(1)}\right\}_{i=1,\ldots,N}$. Then, we denote the largest right singular vector of $\beta_{(1)} := \beta_{[I_1]}$ by $b_1$. For $k > 1$, we obtain $a_i^{(k)} := \left\|\beta_{[i]} \prod_{j<k} \mathbb{M}_{b_j} \eta^\intercal\right\|_{\text{MAX}}$, $I_k := \{a_i^{(k)} \geq c_{qN}^{(k)}\}$ and $b_k$ is the largest right singular vector of $\beta_{(k)} := \beta_{[I_k]} \prod_{j<k} \mathbb{M}_{b_j}$. This procedure is stopped at step $\tilde{p}$ (for some $\tilde{p}$ not necessarily equal to $p$) if $c_{qN}^{(\tilde{p}+1)} < c$. In a nutshell, $I_k$'s are what we will select if we do SPCA directly on $\beta \in \mathbb{R}^{N \times p}$ and

$\eta \in \mathbb{R}^{d\times p}$, while $\widehat{I}_k$'s are obtained by SPCA on $\bar{R} \in \mathbb{R}^{N\times T}$ and $\bar{G} \in \mathbb{R}^{d\times T}$. We need the following assumption to guarantee the selection consistency, that is, $\mathrm{P}(\widehat{I}_k = I_k) \to 1$ for any $1 \le k \le \tilde{p}$.

**Assumption A.8.** *We assume that $\beta_{(k)}$, $a_i^{(k)}$ and $c$ in the above procedure satisfy:*

*(i) $\sigma_1(\beta_{(k)})$ and $\sigma_2(\beta_{(k)})$ are distinct in the sense that there exists a constant $\delta > 0$ such that*

$$\sigma_2(\beta_{(k)}) \le (1+\delta)^{-1}\sigma_1(\beta_{(k)}).$$

*(ii) $c_{qN}^{(k)}$ and $c_{qN+1}^{(k)}$ are distinct in the sense that there exists a constant $\delta > 0$ such that*

$$c_{qN+1}^{(k)} \le (1+\delta)^{-1} c_{qN}^{(k)},$$

*where $c_{qN}^{(k)}$ and $c_{qN+1}^{(k)}$ are the $(qN)$th and $(qN+1)$th largest value in $\left\{a_i^{(k)}\right\}_{i=1,\dots,N}$, respectively.*

*(iii) $c_{qN}^{(\tilde{p}+1)}$ and $c$ are distinct in the sense that there exists a constant $\delta > 0$ such that*

$$c_{qN}^{(\tilde{p}+1)} \le (1+\delta)^{-1} c.$$

Assumption A.8 requires that these singular values are distinguishable, so that their (relative) differences will not vanish asymptotically. This assumption is rather mild, despite not being very explicit.

**Assumption A.9.** *As $T \to \infty$, the following joint central limit theorem holds:*

$$T^{1/2}\begin{pmatrix} T^{-1}\mathrm{vec}(ZV^\intercal) \\ \bar{v} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{12}^\intercal & \Pi_{22} \end{pmatrix}\right),$$

*where $\Pi_{11}$, $\Pi_{12}$, $\Pi_{22}$ are $dp \times dp$, $dp \times p$, and $p \times p$ matrices, respectively, defined as:*

$$\begin{aligned}
\Pi_{11} &= \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(\mathrm{vec}(ZV^\intercal)\mathrm{vec}(ZV^\intercal)^\intercal\right), \\
\Pi_{12} &= \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(\mathrm{vec}(ZV^\intercal)\iota_T^\intercal V^\intercal\right), \\
\Pi_{22} &= \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(V\iota_T\iota_T^\intercal V^\intercal\right).
\end{aligned}$$

Assumption A.9 characterizes the joint asymptotic distribution of $ZV^\intercal$ and $V\iota_T$. Since the dimensions of these random processes are finite, this CLT is a fairly standard result of a central limit theory for mixing processes.

In the same vein, we make an assumption on the central limit result between $ZV^\intercal$ and $Z\iota_T$, which we use for inference on $\alpha_g$.

**Assumption A.10.** *As $T \to \infty$, the following joint central limit theorem holds:*

$$T^{1/2}\begin{pmatrix} T^{-1}\mathrm{vec}(ZV^\intercal) \\ \bar{z} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Pi_{11} & \Pi_{13} \\ \Pi_{13}^\intercal & \Pi_{33} \end{pmatrix}\right),$$

*where $\Pi_{13}$ and $\Pi_{33}$ are $dp \times d$, and $d \times d$ matrices, respectively, defined as:*

$$\Pi_{13} = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(\mathrm{vec}(ZV^\intercal)\iota_T^\intercal Z^\intercal\right),$$
$$\Pi_{33} = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left(Z\iota_T\iota_T^\intercal Z^\intercal\right).$$

Blow we introduce assumptions needed for the SDF estimation. Assumption A.11 ensures that the SDF concept is well defined. Assumption A.12 again can be shown by some large deviation result and certain central limit theorem.

**Assumption A.11.** *Suppose that $v_t$ and $u_t$ are stationary time series independent of $\beta$, and that $\Sigma_v = \mathrm{Cov}(v_t)$ and $\Sigma_u = \mathrm{Cov}(u_t)$ satisfy $\lambda_{\min}(\Sigma_v) \gtrsim 1$ and $\lambda_{\max}(\Sigma_u) \lesssim 1$. Consequently, $\Sigma = \mathrm{Cov}(r_t) = \beta\Sigma_v\beta^\intercal + \Sigma_u$.*

**Assumption A.12.** *The time series $r_t$ and the SDF defined by $m_t = 1 - b^\intercal(r_t - \mathrm{E}(r))$ with $b = \Sigma^{-1}\mathrm{E}(r_t)$ satisfy:*

$$(1) \quad \left\| T^{-1}\sum_{t=1}^{T}(r_t - \bar{r}_t)(m_t - \bar{m}_t) - \mathrm{Cov}(r_t, m_t)\right\|_{\mathrm{MAX}} \lesssim_p (\log N)^{1/2}T^{-1/2}.$$
$$(2) \quad \left\| T^{-1}\sum_{t=1}^{T}(r_t - \bar{r}_t)(r_t - \bar{r}_t)^\intercal - \mathrm{Cov}(r_t)\right\|_{\mathrm{MAX}} \lesssim_p (\log N)^{1/2}T^{-1/2}.$$
$$(3) \quad \left| T^{-1}\sum_{t=1}^{T} m_t - \mathrm{E}(m_t)\right| \lesssim_p T^{-1/2}.$$
$$(4) \quad \left\| T^{-1}\sum_{t=1}^{T} r_t - \mathrm{E}(r_t)\right\|_{\mathrm{MAX}} \lesssim_p (\log N)^{1/2}T^{-1/2}.$$

Finally, we need the following assumption for establishing the convergence of the ridge-based SDF estimator. It ensures that all eigenvalues of $\beta\Sigma_v\beta^\intercal$ are well separated. This assumption shares the spirit with Assumption A.8. A similar assumption has been adopted by, e.g., Wang and Fan (2017).

**Assumption A.13.** *The eigenvalues of $\beta\Sigma_v\beta^\intercal$ are separated in the sense that*

$$(\lambda_j - \lambda_{j+1})/\lambda_j \geq \delta$$

*for some constant $\delta > 0$, where $\lambda_j := \lambda_j(\beta\Sigma_v\beta^\intercal)$ is the $j$th eigenvalue of $\beta\Sigma_v\beta^\intercal$.*

# B Mathematical Proofs

## B.1 Proof of Proposition 1

*Proof.* Note that for any orthogonal matrix $\Gamma \in \mathbb{R}^{N\times N}$, the estimators based on PCA, PLS and Ridge on $R' = \Gamma R$ are the same as those based on $R$. Thus, without loss of generality, we can assume $\beta = (\lambda^{1/2}, 0, \cdots, 0)^\intercal$, where $\lambda = \|\beta\|^2$. The same simplifying assumption is adopted in the proofs of Propositions 1, 2, and 3. Also, since $z_t = 0$, $\bar{G} = \eta\bar{V}$.

We start with $\widehat{\gamma}_g^{PCA}$. We write $\bar{R}$ in the following form:

$$\bar{R} = \beta\bar{V} + \bar{U} = \begin{pmatrix} \sqrt{\lambda}\bar{V} + \bar{U}_1 \\ \bar{U}_2 \end{pmatrix}, \tag{B.1}$$

where $\bar{U}_1$ is the first row of $\bar{U}$ and $\bar{U}_2$ contains the remaining rows. Correspondingly, we write the largest left singular vector of $\bar{R}$ as $\varsigma = (\varsigma_1, \varsigma_2^\intercal)^\intercal$, where $\varsigma_1$ is the first element of $\varsigma$ and $\varsigma_2$ is a vector of the remaining $N-1$ entries of $\varsigma$. Recall that in Algorithm 1, we denote $\xi$ and $\varsigma$ as the largest right and left singular vectors of $\bar{R}$ with the singular value $\sqrt{T\widehat{\lambda}}$, so that by simple algebra we have

$$\varsigma_1 = \frac{(\sqrt{\lambda}\bar{V} + \bar{U}_1)\xi}{\sqrt{T\widehat{\lambda}}}, \quad \varsigma_2 = \frac{\bar{U}_2\xi}{\sqrt{T\widehat{\lambda}}}. \tag{B.2}$$

Since the entries of $U$ and $V$ are i.i.d $\mathcal{N}(0,1)$, we have

$$|T^{-1}\bar{V}\bar{V}^\intercal - 1| = |T^{-1}V(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)V^\intercal - 1| \le |T^{-1}VV^\intercal - 1| + |\bar{v}|^2 \lesssim_p T^{-1/2},$$

where we use large deviation results $|T^{-1}VV^\intercal - 1| \lesssim_p T^{-1/2}$ and $|\bar{v}| \lesssim_P T^{-1/2}$ in the last equation. This equation also implies that $\left\|\bar{V}\right\| - \sqrt{T} \lesssim_p 1$.

Similarly, we can get $|T^{-1}\bar{U}_1\bar{U}_1^\intercal - 1| \lesssim_p T^{-1/2}$ and $\left\|\bar{U}_1\right\| - \sqrt{T} \lesssim_p 1$.

In addition, by Lemma A.1 in Wang and Fan (2017), we have $\left\|N^{-1}U^\intercal U - \mathbb{I}_T\right\| \lesssim_p \sqrt{T/N}$, which leads to

$$\left\|N^{-1}\bar{U}^\intercal\bar{U} - (\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)\right\| = \left\|(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)(N^{-1}U^\intercal U - \mathbb{I}_T)(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)\right\| \lesssim_p \sqrt{T/N}.$$

Next, by direct calculation using the above inequalities we obtain

$$\left\|\frac{\bar{V}^\intercal\bar{U}_1 + \bar{U}_1^\intercal\bar{V}}{T\sqrt{\lambda}} + \frac{\bar{U}^\intercal\bar{U} - N(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)}{T\lambda}\right\| \lesssim_p \frac{1}{\sqrt{\lambda}} + \frac{\sqrt{NT}}{T\lambda} \lesssim_p \frac{1}{\sqrt{\lambda}}.$$

Together with (B.1), we have

$$\left\|\frac{\bar{R}^\intercal\bar{R}}{T\lambda} - \frac{\bar{V}^\intercal\bar{V}}{T} - \frac{N(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)}{T\lambda}\right\| \lesssim_p \frac{1}{\sqrt{\lambda}}. \tag{B.3}$$

Because of this result, to study the eigenstructure of $\bar{R}^\intercal\bar{R}/(T\lambda)$, we need analyze the eigenstructure of

$$M := \frac{\bar{V}^\intercal\bar{V}}{T} + \frac{N(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)}{T\lambda} = \frac{\bar{V}^\intercal\bar{V}}{T} + \tilde{B}(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal),$$

where $\tilde{B} = N/(T\lambda)$ and the assumption of the proposition implies that $\tilde{B} \to B$ for a constant $B$.

Note that $\bar{V}\iota_T = 0$, the eigenvalues of $M$ can be explicitly given by:

$$\lambda_i = \begin{cases} T^{-1}\bar{V}\bar{V}^\intercal + \tilde{B} & i = 1; \\ \tilde{B} & 2 \le i \le T-1; , \\ 0 & i = T. \end{cases} \tag{B.4}$$

and the first eigenvector is $\bar{V}^\intercal/\left\|\bar{V}^\intercal\right\|$. Since the largest eigenvalue of $\bar{R}^\intercal\bar{R}/(T\lambda)$ is $\widehat{\lambda}/\lambda$ with its corresponding eigenvector $\xi$, Weyl's theorem yields that

$$\frac{\widehat{\lambda}}{\lambda} = \frac{\bar{V}\bar{V}^\intercal}{T} + \tilde{B} + O_p\left(\frac{1}{\sqrt{\lambda}}\right) = 1 + \tilde{B} + O_p\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{T}}\right), \tag{B.5}$$

and the sin-theta theorem in Davis and Kahan (1970) implies that

$$\|\mathbb{P}_{\bar{V}^\intercal} - \mathbb{P}_\xi\| = \left\|\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V} - \xi\xi^\intercal\right\| \lesssim_p \frac{1}{\sqrt{\lambda}}, \tag{B.6}$$

which implies that $(\bar{V}\bar{V})^{-1}(\bar{V}\xi)^2 = \xi^\intercal\bar{V}^\intercal(\bar{V}\bar{V})^{-1}\bar{V}\xi = 1 + O_p(\lambda^{-1/2} + T^{-1/2})$. Together with $|T^{-1}\bar{V}\bar{V}^\intercal - 1| \lesssim T^{-1/2}$, we have

$$\frac{|\bar{V}\xi|}{\sqrt{T}} = 1 + O_p\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{T}}\right). \tag{B.7}$$

It is easy to observe that the sign of $\xi$ plays no role in the estimator $\widehat{\gamma}_g^{PCA}$, we can choose $\xi$ such that

$$\frac{\bar{V}\xi}{\sqrt{T}} = 1 + O_p\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{T}}\right). \tag{B.8}$$

Recall that the risk premium estimator is $\widehat{\gamma}_g^{PCA} = \widehat{\eta}\widehat{\gamma}$, where

$$\widehat{\eta} = \frac{\bar{G}\xi}{\sqrt{T}} \quad \text{and } \widehat{\gamma} = \frac{\varsigma^\intercal\bar{r}}{\sqrt{\widehat{\lambda}}}. \tag{B.9}$$

Using $\bar{G} = \eta\bar{V}$ and (B.8), we have

$$\widehat{\eta} = \eta + O_p\left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{T}}\right). \tag{B.10}$$

Write

$$\widehat{\gamma} = \frac{\varsigma^\intercal \bar{r}}{\sqrt{\widehat{\lambda}}} = \frac{\varsigma^\intercal \beta(\gamma+\bar{v})}{\sqrt{\widehat{\lambda}}} + \frac{\varsigma^\intercal \bar{u}}{\sqrt{\widehat{\lambda}}} = \frac{\sqrt{\lambda}\varsigma_1}{\sqrt{\widehat{\lambda}}}(\gamma+\bar{v}) + \frac{\varsigma^\intercal \bar{u}}{\sqrt{\widehat{\lambda}}}, \tag{B.11}$$

where we use $\beta = (\sqrt{\lambda}, 0, \ldots, 0)^\intercal$ in the last step. Now we analyze the two terms on the right hand side of (B.11) one by one. For the first term, using (B.2), we have

$$\frac{\sqrt{\lambda}\varsigma_1}{\sqrt{\widehat{\lambda}}} = \frac{\lambda}{\widehat{\lambda}} \frac{(\bar{V} + \lambda^{-1/2}\bar{U}_1)\xi}{\sqrt{T}} = \frac{\lambda}{\widehat{\lambda}} \left( \frac{\bar{V}\xi}{\sqrt{T}} + \frac{\bar{U}_1 \xi}{\sqrt{T\lambda}} \right).$$

Using (B.5) and (B.8) and $\left\|\bar{U}_1\right\| \lesssim_p \sqrt{T}$, it follows that

$$\frac{\sqrt{\lambda}\varsigma_1}{\sqrt{\widehat{\lambda}}} = \frac{1}{1+\tilde{B}} + O_p\left( \frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{T}} \right). \tag{B.12}$$

For the second term in (B.11), using (B.2) again, we can write

$$\frac{\varsigma^\intercal \bar{u}}{\sqrt{\widehat{\lambda}}} = \frac{\varsigma_1 U_1 \iota_T}{T\sqrt{\widehat{\lambda}}} + \frac{\varsigma_2^\intercal U_2 \iota_T}{T\sqrt{\widehat{\lambda}}} = \frac{\varsigma_1 U_1 \iota_T}{T\sqrt{\widehat{\lambda}}} + \frac{\xi^\intercal (\mathbb{I}_T - T^{-1}\iota_T \iota_T^\intercal) U_2^\intercal U_2 \iota_T}{T^{3/2}\widehat{\lambda}}. \tag{B.13}$$

The condition that entries of $U$ are independent $\mathcal{N}(0,1)$ implies that $\|U_1 \iota_T\| \lesssim_p \sqrt{T}$, with $\widehat{\lambda}/\lambda \xrightarrow{p} 1+B$ as shown in (B.5), the first term in (B.13) is of order $O_p(T^{-1/2}\lambda^{-1/2})$. For the second term in (B.13), using $\left\|(N-1)^{-1}U_2^\intercal U_2 - \mathbb{I}_T\right\| \lesssim_p \sqrt{T/N}$, we have

$$\begin{aligned} \left|\frac{\xi^\intercal (\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal) U_2^\intercal U_2 \iota_T}{T^{3/2}\widehat{\lambda}}\right| \leq & \left|\frac{(N-1)\xi^\intercal (\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)\iota_T}{T^{3/2}\widehat{\lambda}}\right| + \frac{N-1}{T\widehat{\lambda}} \left\|(N-1)^{-1}U_2^\intercal U_2 - \mathbb{I}_T\right\| \\ = & \frac{N-1}{T\widehat{\lambda}} \left\|(N-1)^{-1}U_2^\intercal U_2 - \mathbb{I}_T\right\| \lesssim_p \frac{1}{\sqrt{\lambda}}, \end{aligned}$$

which leads to $|\widehat{\lambda}^{-1/2}\varsigma^\intercal \bar{u}| \lesssim_p \lambda^{-1/2}$. Plugging this and (B.12) into (B.11), we obtain

$$\widehat{\gamma} = \frac{\varsigma^\intercal \bar{r}}{\sqrt{\widehat{\lambda}}} = \frac{\gamma}{1+\tilde{B}} + O_p\left( \frac{1}{\sqrt{\lambda}} + \frac{1}{\sqrt{T}} \right), \tag{B.14}$$

and thus $\widehat{\gamma}_g^{PCA} \xrightarrow{p} (1+B)^{-1}\eta\gamma$ by (B.10), (B.14) and $\tilde{B} \to B$. □

## B.2 Proof of Proposition 2

*Proof.* Recall that in Section 2.3.2, we have

$$\widehat{\gamma}_g^{PLS} = \left\|\bar{G}\bar{R}^\intercal \bar{R}\right\|^{-2} \bar{G}\bar{R}^\intercal \bar{R}\bar{G}^\intercal \bar{G}\bar{R}^\intercal \bar{r}. \tag{B.15}$$

We analyze $\|\bar{G}\bar{R}^\intercal\bar{R}\|$, $\bar{G}\bar{R}^\intercal\bar{R}\bar{G}^\intercal$ and $\bar{G}\bar{R}^\intercal\bar{r}$ separately. Recall that from (B.3), we have

$$\left\|\frac{\bar{R}^\intercal\bar{R}}{T\lambda}-\frac{\bar{V}^\intercal\bar{V}}{T}-\tilde{B}(\mathbb{I}_T-T^{-1}\iota_T\iota_T^\intercal)\right\|\lesssim_p\frac{1}{\sqrt{\lambda}},$$

where $\tilde{B}=N/(T\lambda)$ satisfies $\tilde{B}\to B$. Together with $\bar{G}=\eta\bar{V}$ and $\|\bar{G}\|\lesssim_p\sqrt{T}$, we have

$$\begin{aligned}\frac{1}{T\lambda\sqrt{T}}\|\bar{G}\bar{R}^\intercal\bar{R}\|&=\frac{1}{\sqrt{T}}\left\|\bar{G}\left(\frac{\bar{V}^\intercal\bar{V}}{T}+\tilde{B}(\mathbb{I}_T-T^{-1}\iota_T\iota_T^\intercal)\right)\right\|+O_p\left(\frac{1}{\sqrt{\lambda}}\right)\\&=\frac{\eta}{\sqrt{T}}\left\|\frac{\bar{V}\bar{V}^\intercal\bar{V}}{T}+\tilde{B}\bar{V}\right\|+O_p\left(\frac{1}{\sqrt{\lambda}}\right)\xrightarrow{p}\eta(1+B),\end{aligned}\tag{B.16}$$

where we use $|T^{-1}\bar{V}\bar{V}^\intercal-1|\lesssim_p T^{-1/2}$ and $\|\bar{V}\|-\sqrt{T}\lesssim_p 1$ in the last equation. For the same reason, by direct calculation we have

$$\begin{aligned}\frac{1}{T^2\lambda}\bar{G}\bar{R}^\intercal\bar{R}\bar{G}^\intercal&=\frac{1}{T}\bar{G}\left(\frac{\bar{V}^\intercal\bar{V}}{T}+\tilde{B}(\mathbb{I}_T-T^{-1}\iota_T\iota_T^\intercal)\right)\bar{G}^\intercal+O_p\left(\frac{1}{\sqrt{\lambda}}\right)\\&=\eta^2\frac{\bar{V}\bar{V}^\intercal\bar{V}\bar{V}^\intercal}{T^2}+\eta^2\tilde{B}\frac{\bar{V}\bar{V}^\intercal}{T}+O_p\left(\frac{1}{\sqrt{\lambda}}\right)\xrightarrow{p}\eta^2(1+B).\end{aligned}\tag{B.17}$$

Next, we write

$$\frac{1}{T\lambda}\bar{G}\bar{R}^\intercal\bar{r}=\frac{1}{T\lambda}\bar{G}\bar{R}^\intercal\beta(\gamma+\bar{v})+\frac{1}{T\lambda}\bar{G}\bar{R}^\intercal\bar{u}.\tag{B.18}$$

We analyze these two terms in (B.18) separately. For the first term, we can write $\bar{R}$ in the form of (B.1) as in the proof of Proposition 1. Then, using $\|\bar{U}_1\|\lesssim_p\sqrt{T}$ we have

$$\frac{1}{T\lambda}\bar{G}\bar{R}^\intercal\beta=\eta\frac{\bar{V}\bar{V}^\intercal}{T}+\eta\frac{\bar{V}\bar{U}_1^\intercal}{T\sqrt{\lambda}}=\eta\frac{\bar{V}\bar{V}^\intercal}{T}+O_p\left(\frac{1}{\sqrt{\lambda}}\right).\tag{B.19}$$

For the second term in (B.18), we have

$$\begin{aligned}\frac{1}{T\lambda}\bar{G}\bar{R}^\intercal\bar{u}&=\eta\frac{1}{T^2\sqrt{\lambda}}\bar{V}\bar{V}^\intercal\bar{U}_1\iota_T+\eta\frac{1}{T^2\lambda}\bar{V}\bar{U}^\intercal U\iota_T=\eta\frac{1}{\sqrt{\lambda}}\frac{\bar{V}\bar{V}^\intercal}{T}\frac{\bar{U}_1\iota_T}{T}+\eta\frac{1}{T^2\lambda}\bar{V}U^\intercal U\iota_T\\&=O_p\left(\frac{1}{\sqrt{T\lambda}}\right)+\eta\frac{N}{T^2\lambda}\bar{V}\left(N^{-1}U^\intercal U-\mathbb{I}_T\right)\iota_T+\eta\frac{N}{T^2\lambda}\bar{V}\iota_T=O_p\left(\frac{1}{\sqrt{T\lambda}}\right)+O_p\left(\frac{1}{\sqrt{\lambda}}\right),\end{aligned}\tag{B.20}$$

where we use $\|N^{-1}U^\intercal U-\mathbb{I}_T\|\lesssim_p\sqrt{T/N}$ and $\bar{V}\iota_T=0$ in the last equation. Plugging (B.19) and (B.20) into (B.18), we have

$$\frac{1}{T\lambda}\bar{G}\bar{R}^\intercal\bar{r}=\eta\frac{\bar{V}\bar{V}^\intercal}{T}(\gamma+\bar{v})+O_p\left(\frac{1}{\sqrt{\lambda}}\right)\xrightarrow{p}\eta\gamma.\tag{B.21}$$

Plug (B.16), (B.17), (B.21) into (B.15), we have

$$\widehat{\gamma}_g^{PLS} \xrightarrow{p} \frac{1}{\eta^2(1+B)^2}\eta^2(1+B)\eta\gamma = \frac{1}{1+B}\eta\gamma.$$

□

### B.3 Proof of Proposition 3

*Proof.* Since $\text{Rank}(\bar{R}) \leq \min\{N, T-1\}$, and the assumptions of the proposition imply that $N/T \to \infty$, we thereby have a condensed SVD of $\bar{R}$ as

$$\bar{R} = \sqrt{T}(\varsigma, \varsigma_*)\widehat{\Lambda}^{1/2}(\xi, \xi_*)^\intercal = \sqrt{T}\varsigma\widehat{\lambda}^{1/2}\xi^\intercal + \sqrt{T}\varsigma_*\widehat{\Lambda}_*^{1/2}\xi_*^\intercal,$$

where $\widehat{\Lambda}^{1/2}$ is the diagonal matrix of $T-1$ singular values, $\varsigma$, $\xi$ are the left and right singular vectors corresponding to the largest singular value of $T^{-1/2}\bar{R}$, which is denoted by $\widehat{\lambda}^{1/2}$. In addition, $\varsigma_* \in \mathbb{R}^{N\times(T-2)}$ and $\xi_* \in \mathbb{R}^{T\times(T-2)}$ are the singular vectors corresponding to the rest $T-2$ nonzero singular values, $\widehat{\Lambda}_*^{1/2} \in \mathbb{R}^{(T-2)\times(T-2)}$. By direct calculation, we have

$$\sqrt{T}\bar{R}^\intercal\left(\bar{R}\bar{R}^\intercal + \mu I\right)^{-1} = (\xi, \xi_*)\widehat{\Lambda}^{1/2}(\widehat{\Lambda} + T^{-1}\mu I)^{-1}(\varsigma, \varsigma_*)^\intercal = \frac{\widehat{\lambda}^{1/2}}{\widehat{\lambda} + T^{-1}\mu}\xi\varsigma^\intercal + \xi_*\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_* + T^{-1}\mu I\right)^{-1}\varsigma_*^\intercal,$$

and thus, with $\bar{G} = \eta\bar{V}$, the Ridge estimator can be written as

$$\begin{aligned}\widehat{\gamma}_g^{Ridge} = \bar{G}\bar{R}^\intercal\left(\bar{R}\bar{R}^\intercal + \mu I\right)^{-1}\bar{r} &= \frac{\widehat{\lambda}}{\widehat{\lambda} + T^{-1}\mu}\frac{\eta\bar{V}\xi}{\sqrt{T}}\frac{\varsigma^\intercal\bar{r}}{\sqrt{\widehat{\lambda}}} + \frac{\eta\bar{V}\xi_*}{\sqrt{T}}\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_* + T^{-1}\mu\right)^{-1}\varsigma_*^\intercal\bar{r} \\ &= \frac{\widehat{\lambda}}{\widehat{\lambda} + T^{-1}\mu}\widehat{\gamma}_g^{PCA} + \frac{\eta\bar{V}\xi_*}{\sqrt{T}}\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_* + T^{-1}\mu\right)^{-1}\varsigma_*^\intercal\bar{r}. \qquad \text{(B.22)}\end{aligned}$$

Using (B.5) and the fact that $T^{-1}\lambda^{-1}\mu \to D$ and Proposition 1, we can show that the first term in (B.22) converges to $(1+B+D)^{-1}\eta\gamma$. With respect to the second term, as shown in (B.3), we have

$$\left\|\frac{\bar{R}^\intercal\bar{R}}{T\lambda} - \frac{\bar{V}^\intercal\bar{V}}{T} - \frac{N(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)}{T\lambda}\right\| \lesssim_p \frac{1}{\sqrt{\lambda}},$$

and the eigenvalues of

$$M = \frac{\bar{V}^\intercal\bar{V}}{T} + \frac{N(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal)}{T\lambda}$$

are given by (B.4), it then follows from Weyl's theorem that $\lambda_i(T^{-1}\lambda^{-1}\bar{R}^\intercal\bar{R}) = \tilde{B} + O_p(\lambda^{-1/2})$ for $2 \leq i \leq T-1$. Note that $\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_* + T^{-1}\mu\right)^{-1}$ is a $(T-2)\times(T-2)$ diagonal matrix and the $i$th

element on the diagonal is

$$\frac{\lambda_{i+1}(T^{-1}\bar{R}^\intercal\bar{R})^{1/2}}{\lambda_{i+1}(T^{-1}\bar{R}^\intercal\bar{R})+T^{-1}\mu}=\frac{1}{\sqrt{\lambda}}\frac{\lambda_{i+1}(T^{-1}\lambda^{-1}\bar{R}^\intercal\bar{R})^{1/2}}{\lambda_{i+1}(T^{-1}\lambda^{-1}\bar{R}^\intercal\bar{R})+T^{-1}\lambda^{-1}\mu}.$$

Together with $T^{-1}\lambda^{-1}\mu\to D$, we have

$$\left\|\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_*+T^{-1}\mu\right)^{-1}\right\|=\max_{1\le i\le T-2}\frac{\lambda_{i+1}(T^{-1}\bar{R}^\intercal\bar{R})^{1/2}}{\lambda_{i+1}(T^{-1}\bar{R}^\intercal\bar{R})+T^{-1}\mu}\lesssim_p\frac{1}{\sqrt{\lambda}}. \tag{B.23}$$

Also, with $\|\bar{u}\|\lesssim_p\sqrt{N/T}$, we have

$$\|\varsigma_*^\intercal\bar{r}\|\le\|\varsigma_*^\intercal\beta(\gamma+\bar{v})\|+\|\varsigma_*^\intercal\bar{u}\|\le\|\beta(\gamma+\bar{v})\|+\|\bar{u}\|\lesssim_p\sqrt{\lambda}+\sqrt{N/T}\lesssim_p\sqrt{\lambda} \tag{B.24}$$

and

$$\left\|\frac{\bar{V}\xi_*}{\sqrt{T}}\right\|^2=\left\|\frac{\bar{V}(\xi,\xi_*)}{\sqrt{T}}\right\|^2-\left\|\frac{\bar{V}\xi}{\sqrt{T}}\right\|^2\le\left\|\frac{\bar{V}}{\sqrt{T}}\right\|^2-\left\|\frac{\bar{V}\xi}{\sqrt{T}}\right\|^2=1+O_p\left(\frac{1}{\sqrt{T}}\right)-\left\|\frac{\bar{V}\xi}{\sqrt{T}}\right\|^2\lesssim_p\frac{1}{\sqrt{\lambda}}+\frac{1}{\sqrt{T}}, \tag{B.25}$$

where we use (B.8) in the last inequality. Consequently, using (B.23), (B.24) and (B.25), we have

$$\left|\frac{\eta\bar{V}\xi_*}{\sqrt{T}}\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_*+T^{-1}\mu\right)^{-1}\varsigma_*^\intercal\bar{r}\right|\le\left\|\frac{\eta\bar{V}\xi_*}{\sqrt{T}}\right\|\left\|\widehat{\Lambda}_*^{1/2}\left(\widehat{\Lambda}_*+T^{-1}\mu\right)^{-1}\right\|\|\varsigma_*^\intercal\bar{r}\|\lesssim T^{-1/4}+\lambda^{-1/4}.$$

By comparing this with the limit of the first term in (B.22), we obtain

$$\widehat{\gamma}_g^{Ridge}\xrightarrow{p}\frac{1}{1+B+D}\eta\gamma.$$

□

## B.4 Proof of Proposition 4

*Proof.* By direct calculation, we can write

$$RR^\intercal+T\mu\bar{r}\bar{r}^\intercal=R\left(\mathbb{I}_T+\frac{\mu}{T}\iota_T\iota_T^\intercal\right)R^\intercal=R\left(\mathbb{I}_T+\frac{\tilde{\mu}}{T}\iota_T\iota_T^\intercal\right)^2R^\intercal, \tag{B.26}$$

where $\tilde{\mu}=\sqrt{\mu+1}-1$. Hence, the eigenvectors of $RR^\intercal+T\mu\bar{r}\bar{r}^\intercal$ are equivalent to the left singular vectors of $R\left(\mathbb{I}_T+T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal\right)$. Let $\varsigma$ and $\xi$ denote the largest left and right singular vector of $R\left(\mathbb{I}_T+T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal\right)$. Note that $\xi$ can be viewed as the largest eigenvector of

$$(\mathbb{I}_T+T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)R^\intercal R(\mathbb{I}_T+T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal),$$

we analyze the eigenspace of this matrix first. Similar to (B.3) in the PCA case, we have the following approximation of $R^\intercal R$

$$\left\| \frac{R^\intercal R}{T\lambda} - \frac{\bar{V}^\intercal \bar{V}}{T} - \gamma \frac{\iota_T \bar{V} + \bar{V}^\intercal \iota_T^\intercal}{T} - \gamma^2 \frac{\iota_T \iota_T^\intercal}{T} - \frac{N}{T\lambda} \mathbb{I}_T \right\| \lesssim_p \frac{1}{\sqrt{T}} + \frac{1}{\sqrt{\lambda}}, \tag{B.27}$$

by $|T^{-1}\bar{V}\bar{V}^\intercal - 1| \lesssim_p T^{-1/2}$, $\|\bar{U}_1\| \lesssim_p T^{1/2}$ and $\|N^{-1}\bar{U}^\intercal \bar{U} - (\mathbb{I}_T - T^{-1}\iota_T \iota_T^\intercal)\| \lesssim_p \sqrt{T/N}$.

Then, with (B.27) and $N/(T\lambda) \to B$, we have

$$\|T^{-1}\lambda^{-1}(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T \iota_T^\intercal) R^\intercal R (\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T \iota_T^\intercal) - M^*\| = o_p(1) \tag{B.28}$$

where the matrix $M^*$ here is defined by

$$M^* := B\mathbb{I}_T + T^{-1}\bar{V}^\intercal \bar{V} + T^{-1}(1+\tilde{\mu})\gamma(\iota_T \bar{V} + \bar{V}^\intercal \iota_T^\intercal) + T^{-1}\Big((1+\tilde{\mu})^2\gamma^2 + \tilde{\mu}^2 B + 2\tilde{\mu}B\Big)\iota_T \iota_T^\intercal.$$

Recall that $\xi$ is the eigenvector of $T^{-1}\lambda^{-1}(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T \iota_T^\intercal) R^\intercal R (\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T \iota_T^\intercal)$, we can analyze the eigenspace of $M^*$ first and then use sin-theta theorem to characterize $\xi$.

Firstly, the rank of $M^* - B\mathbb{I}_T$ is at most 2. Using the fact that $\bar{V}\iota_T = 0$, by direct calculation, we have the two nozero eigenvalues of $M^* - B\mathbb{I}_T$ are the solutions of the equation

$$(x - a_1)(x - a_3) - a_2^2 = 0, \tag{B.29}$$

where $a_1 = T^{-1}\|\bar{V}\|^2$, $a_2 = T^{-1/2}(1+\tilde{\mu})\gamma\|\bar{V}\|$ and $a_3 = (1+\tilde{\mu})^2\gamma^2 + \tilde{\mu}^2 B + 2\tilde{\mu}B$. Since the larger solution of (B.29) is

$$\frac{a_1 + a_3 + \sqrt{(a_1 - a_3)^2 + 4a_2^2}}{2} \geq a_1 > 0 \tag{B.30}$$

with probability 1, it is also the largest eigenvalue of $M^* - B\mathbb{I}_T$. In addition, the second largest eigenvalue of $M^* - B\mathbb{I}_T$ should be distinct with $\lambda_1(M^* - B\mathbb{I}_T)$. To see this, if the second eigenvalue is the other solution of (B.29), we have $\lambda_1(M^* - B\mathbb{I}_T) - \lambda_2(M^* - B\mathbb{I}_T) = \sqrt{(a_1 - a_3)^2 + 4a_2^2} \geq \max\{2a_2, |a_1 - a_3|\} > 0$. If the second eigenvalue is 0 (in which case the second solution of the above equation must be negative), we have shown in (B.30) that $\lambda_1(M^* - B\mathbb{I}_T) - \lambda_2(M^* - B\mathbb{I}_T) = \lambda_1(M^* - B\mathbb{I}_T) \geq a_1 > 0$. In both cases, we have $\lambda_1(M^* - B\mathbb{I}_T) - \lambda_2(M^* - B\mathbb{I}_T) \geq \delta > 0$ for some constant $\delta > 0$. Consequently,

$$\lambda_1(M^*) - \lambda_2(M^*) = \lambda_1(M^* - B\mathbb{I}_T) - \lambda_2(M^* - B\mathbb{I}_T) \geq \delta, \tag{B.31}$$

for some constant $\delta > 0$. Now we calculate the first eigenvector of $M^*$, which should also be the first eigenvector of $M^* - B\mathbb{I}_T$. We use $\tilde{\xi}$ to denote this eigenvector. Since we already know that the largest eigenvalue of $\lambda_1(M^* - B\mathbb{I}_T)$ is a solution of (B.29), which means that $\tilde{\xi}$ should be in the space

spanned by $\bar{V}^\intercal$ and $\iota_T$. Writing $\tilde{\xi} = K_1 \left\|\bar{V}\right\|^{-1} \bar{V}^\intercal + K_2 T^{-1/2} \iota_T$ and plugging the largest eigenvalue of $\lambda_1(M^* - B\mathbb{I}_T)$ of (B.30) into $\lambda_1(M - B\mathbb{I}_T)\tilde{\xi} = (M - B\mathbb{I}_T)\tilde{\xi}$, we directly get

$$\frac{K_2}{K_1} = \frac{\sqrt{(a_1 - a_3)^2 + 4a_2^2} + a_3 - a_1}{2a_2}, \tag{B.32}$$

which will pin down $K_1$ and $K_2$ because we also have $\left\|\tilde{\xi}\right\| = 1$.

Using $\left\|T^{-1}\lambda^{-1}(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)R^\intercal R(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal) - M\right\| = o_p(1)$, (B.31) and sin-theta theorem, we have

$$\left\|\mathbb{P}_\xi - \mathbb{P}_{\tilde{\xi}}\right\| \leq \frac{o_p(1)}{\delta - o_p(1)} = o_p(1),$$

which implies that $|\tilde{\xi}^\intercal \xi| \xrightarrow{p} 1$ and consequently,

$$\left\|\xi - K_1 \left\|\bar{V}\right\|^{-1} \bar{V}^\intercal - K_2 T^{-1/2} \iota_T\right\| = o_p(1) \quad \text{or} \quad \left\|\xi + K_1 \left\|\bar{V}\right\|^{-1} \bar{V}^\intercal + K_2 T^{-1/2} \iota_T\right\| = o_p(1).$$

Since the sign of $\xi$ plays no role in the estimator $\widehat{\gamma}_g^{rpPCA}$, we can simply assume the former one.

Also, the relationship between singular vectors implies that

$$\widehat{F} = \varsigma^\intercal R = \left\|R(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)\right\|^{-1} \xi^\intercal(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)R^\intercal R. \tag{B.33}$$

With the approximation of $R^\intercal R$ in (B.27), $\bar{V}\iota_T = 0$, $T^{-1}\bar{V}\bar{V}^\intercal = 1 + O_p(T^{-1/2})$ and $N/(T\lambda) \to B$, by direct calculation, we have

$$\left\|\left\|\bar{V}\right\|^{-1} \bar{V}(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)R^\intercal R - \lambda T^{1/2}\Big((1+B)\bar{V} + \gamma\iota_T^\intercal\Big)\right\| = o_p(\lambda T), \tag{B.34}$$

and

$$\left\|T^{-1/2}\iota_T^\intercal(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)R^\intercal R - \lambda T^{1/2}(1+\tilde{\mu})\Big(\gamma\bar{V} + (\gamma^2 + B)\iota_T^\intercal\Big)\right\| = o_p(\lambda T). \tag{B.35}$$

Plugging (B.34), (B.35) and $\left\|\xi - K_1 \left\|\bar{V}\right\|^{-1} \bar{V}^\intercal + K_2 T^{-1/2} \iota_T\right\| = o_p(1)$ into (B.33) we have

$$\left\|\left\|R(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)\right\| \widehat{F} - \lambda T^{1/2}(L_1\bar{V} + L_2\iota_T^\intercal)\right\| = o_p(\lambda T), \tag{B.36}$$

where

$$L_1 = K_1(1+B) + K_2(1+\tilde{\mu})\gamma, \qquad L_2 = K_1\gamma + K_2(1+\tilde{\mu})(\gamma^2 + B). \tag{B.37}$$

It is easy to observe that scalar plays no role in the estimator $\widehat{\gamma}_g^{rpPCA}$, we can redefine

$$\widehat{F}^* = \lambda^{-1}T^{-1/2}L_1^{-1}\left\|R(\mathbb{I}_T + T^{-1}\tilde{\mu}\iota_T\iota_T^\intercal)\right\|\widehat{F}$$

and use $\widehat{F}^*$ to create $\widehat{\gamma}_g^{rpPCA}$. Then, (B.36) becomes $\left\|\widehat{F}^* - \bar{V} - L_1^{-1}L_2\iota_T^\intercal\right\| = o_p\left(T^{1/2}\right)$. Consequently, $\left\|\widehat{V} - \bar{V}\right\| = \left\|\widehat{F}^*(\mathbb{I}_T - T^{-1}\iota_T\iota_T^\intercal) - \bar{V}\right\| = o_p\left(T^{1/2}\right)$, $\widehat{\gamma} = T^{-1}\widehat{F}^*\iota_T = L_1^{-1}L_2 + o_p(1)$, and

$$\widehat{\eta} = \bar{G}\widehat{V}^\intercal(\widehat{V}\widehat{V}^\intercal)^{-1} = \eta\bar{V}\widehat{V}^\intercal(\widehat{V}\widehat{V}^\intercal)^{-1} = \eta\Big(\bar{V}\bar{V}^\intercal + o_p(T)\Big)\Big(\bar{V}\bar{V}^\intercal + o_p(T)\Big)^{-1} = \eta + o_p(1),$$

and the estimator $\widehat{\gamma}_g^{rpPCA} = \widehat{\eta}\widehat{\gamma} \xrightarrow{p} \eta L_1^{-1}L_2$, where $L_1$ and $L_2$ are defined in (B.37).

In light of that $a_1 \xrightarrow{p} 1$, $a_2 \xrightarrow{p} (1+\tilde{\mu})\gamma$, $\tilde{\mu} = \sqrt{1+\mu} - 1$, $\widehat{\gamma}_g^{rpPCA} \xrightarrow{p} \eta L_2/L_1$, (B.32) and the definitions of $L_1$ and $L_2$ in (B.37), we have

$$\widehat{\gamma}_g^{rpPCA} \xrightarrow{p} w(1+B)^{-1}\eta\gamma + (1-w)\eta(\gamma + \gamma^{-1}B),$$

where

$$w = \frac{2+2B}{1+2B+\sqrt{(1-a)^2 + 4(1+\mu)\gamma} + a}, \qquad a = (1+\mu)(\gamma^2 + B) - B.$$

□

## B.5 Proof of Proposition 5

*Proof.* Consider the set $I = \{|\beta_{[i]}| \geq \beta_{\{qN\}}\}$, where $|\beta|_{\{qN\}}$ is the $(qN)$th largest value in $\left\{|\beta_{[i]}|\right\}_{i\in[N]}$. Since

$$T^{-1}\bar{R}\bar{G}^\intercal - \beta\eta^\intercal = \beta\left(T^{-1}\bar{V}\bar{V}^\intercal - 1\right)\eta^\intercal + T^{-1}\bar{U}\bar{V}^\intercal\eta^\intercal + T^{-1}\beta\bar{V}\bar{Z}^\intercal + T^{-1}\bar{U}\bar{Z}^\intercal,$$

we have

$$\begin{aligned}\left\|T^{-1}\bar{R}\bar{G}^\intercal - \beta\eta^\intercal\right\|_{\text{MAX}} \lesssim{}& \|\beta\|_{\text{MAX}}\left|T^{-1}\bar{V}\bar{V}^\intercal - 1\right|\|\eta\| + T^{-1}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}}\|\eta\| \\ &+ T^{-1}\|\beta\|_{\text{MAX}}\left\|\bar{V}\bar{Z}^\intercal\right\| + T^{-1}\left\|\bar{U}\bar{Z}^\intercal\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2}T^{-1/2}.\end{aligned}$$

In other words, the difference between $T^{-1}\bar{R}\bar{G}^\intercal$ and $\beta\eta^\intercal$ for all test assets is bounded by $O_p\left((\log N)^{1/2}T^{-1/2}\right)$, which is $o(1)$ under our assumption.

On the other hand, with the assumption that $\|\beta\|_{\text{MAX}} \lesssim 1$ and the definition of $|\beta|_{\{qN\}}$, we have $\left\|\beta_{[I_0]}\right\|^2 \lesssim qN + (N_0 - qN)|\beta|_{\{qN\}}^2$. Together with the assumption that $qN/N_0 \to 0$ and $\left\|\beta_{[I_0]}\right\| \asymp \sqrt{N_0}$, it leads to $|\beta|_{\{qN\}}^2 \gtrsim \|\beta_{I_0}\|^2/N_0 \asymp 1$. Then, with the assumption that $|\beta|_{\{qN+1\}} \leq (1+\delta)^{-1}|\beta|_{\{qN\}}$, we have that the difference between $|\beta|_{\{qN+1\}}$ and $|\beta|_{\{qN\}}$ should be at the same rate as $|\beta|_{\{qN\}} \gtrsim 1$, which is larger than the difference between $T^{-1}\bar{R}\bar{G}^\intercal$ and $\beta\eta^\intercal$. Therefore, with

probability approaching one, we have $\widehat{I} = I$. In what follows, we only need consider the case of $\widehat{I} = I$.

Since $qN/N_0 \to 0$, by the definition of $I$, we have $\left\|\beta_{[I]}\right\| / \sqrt{|I|} \geq \left\|\beta_{[I_0]}\right\| / \sqrt{|I_0|}$. Together with the assumption that $\left\|\beta_{[I_0]}\right\| \asymp \sqrt{N_0}$, $\left\|\beta_{[I_0]}\right\| \to \infty$ and $|I| = qN \to \infty$, we have $|I|/(T\left\|\beta_{[I]}\right\|^2) \to 0$ and $\left\|\beta_{[I]}\right\| \to \infty$. Now compared to the case with PCA, the expansion on $\widehat{\gamma}_g^{SPCA}$ resembles that of (B.11), except for an extra term that depends on $\bar{Z}$ and the restriction of $\bar{r}$ on $I$:

$$\widehat{\gamma}_g^{SPCA} = \frac{\eta \bar{V}\xi}{\sqrt{T}} \frac{\varsigma^\intercal \bar{r}_{[I]}}{\sqrt{\widehat{\lambda}}} + \frac{\bar{Z}\xi}{\sqrt{T}} \frac{\varsigma^\intercal \bar{r}_{[I]}}{\sqrt{\widehat{\lambda}}}. \tag{B.38}$$

In restriction to the smaller set $I$, the first term matches exactly the case of $|I|/(T\left\|\beta_{[I]}\right\|^2) \to 0 = B$ in Proposition 1, which yields

$$\frac{\eta \bar{V}\xi}{\sqrt{T}} \frac{\varsigma^\intercal \bar{r}_{[I]}}{\sqrt{\widehat{\lambda}}} = \eta\gamma + o_p(1).$$

We now analyze the second term in (B.38). As shown in (B.14), we have

$$\left\| \frac{\varsigma^\intercal \bar{r}_{[I]}}{\sqrt{\widehat{\lambda}}} \right\| \lesssim_p 1,$$

so to prove that SPCA is consistent in this case, it is sufficient to show that $T^{-1/2}\left\|\bar{Z}\xi\right\| \xrightarrow{p} 0$, where $\xi$ is the largest right singular vector of $\bar{R}_{[I]}$. Similar to the proof of (B.6) in Proposition 1, we can show that the difference between projection matrices, $\mathbb{P}_\xi$ and $\mathbb{P}_{\bar{V}^\intercal}$ is small by sin-theta theorem. That is to say, we have $\left\|\xi\xi^\intercal - \bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\right\| \xrightarrow{p} 0$. Then, with the fact that

$$\left\|\bar{Z}\bar{V}^\intercal(\bar{V}\bar{V}^\intercal)^{-1}\bar{V}\right\| \leq \left\|\bar{Z}\bar{V}^\intercal\right\| \left\|(\bar{V}\bar{V}^\intercal)^{-1}\right\| \left\|\bar{V}\right\| \lesssim_p T^{1/2} \times T^{-1} \times T^{1/2} \lesssim_p 1,$$

we have $T^{-1/2}\left\|\bar{Z}\xi\xi^\intercal\right\| \xrightarrow{p} 0$. Consequently,

$$T^{-1/2}\left\|\bar{Z}\xi\right\| = T^{-1/2}\left\|\bar{Z}\xi\xi^\intercal\xi\right\| \leq T^{-1/2}\left\|\bar{Z}\xi\xi^\intercal\right\| \|\xi\| \xrightarrow{p} 0.$$

Hence, $z_t$ does not affect the consistency of the SPCA estimator. This completes the proof. □

## B.6 Proof of Theorem 1

*Proof.* It is sufficient to consider the case $\Sigma_v = \mathbb{I}_p$. Otherwise, we can do transformation $V' = \Sigma_v^{-\frac{1}{2}} V$, $\beta'_{[I]} = \beta_{[I]}\Sigma_v^{\frac{1}{2}}$, $\eta' = \eta\Sigma_v^{\frac{1}{2}}$ and $\gamma' = \Sigma_v^{-\frac{1}{2}}\gamma$. All the Assumptions A.1-A.8 still hold for the new $V'$, $\beta'_{[I]}$. Therefore, we only need analyze the case of $\Sigma_v = \mathbb{I}_p$.

For notation simplicity, throughout the proofs of Theorems 1-4, we use $\widetilde{R}_{(k)} := \left(\bar{R}_{(k)}\right)_{[\widehat{I}_k]}$ to denote the matrix on which we perform SVD in each step of Algorithm 5. Similarly, we use $\widetilde{r}_{(k)} := \left(\bar{r}_{(k)}\right)_{[\widehat{I}_k]}$. The first left and right singular vectors of $\widetilde{R}_{(k)}$ are denoted by $\varsigma_{(k)}$ and $\xi_{(k)}$, while the largest singular

value of $\widetilde{R}_{(k)}$ is denoted by $\sqrt{T\widehat{\lambda}_{(k)}}$. As a result, $\widehat{\lambda}_{(k)} = T^{-1}\left\|\widetilde{R}_{(k)}\right\|^2$.

Using the above notation, our estimated factor at $k$-th step is $\widehat{V}_{(k)} = \sqrt{T}\xi_{(k)}^{\intercal} \in \mathbb{R}^{1\times T}$, the risk premium of this factor is given by $\widehat{\gamma}_{(k)} = \widehat{\lambda}_{(k)}^{-1/2}\varsigma_{(k)}^{\intercal}\widetilde{r}_{(k)}$, the loading matrix of $R$ on this factor is $\widehat{\beta}_{(k)} = T^{-1/2}\bar{R}\xi_{(k)}$, and the loading of $G$ on this factor is $\widehat{\eta}_{(k)} = T^{-1/2}\bar{G}\xi_{(k)}$. By footnote 14, we can use $\bar{G}$ instead of $\bar{G}_{(k)}$ in Algorithm 5 and throughout the proof. We denote $\widehat{\eta} = (\widehat{\eta}_{(1)}, \ldots, \widehat{\eta}_{(\tilde{p})})$ and $\widehat{\gamma} = (\widehat{\gamma}_{(1)}, \ldots, \widehat{\gamma}_{(\tilde{p})})^{\intercal}$, so the risk premium estimator is $\widehat{\gamma}_g^{SPCA} = \widehat{\eta}\widehat{\gamma}$.

By Lemma 2, we have $\xi_{(i)}^{\intercal}\xi_{(j)} = 0$ for $i \neq j \leq \tilde{p}$. This suggests that $\widehat{V}_{(k)}$ at each step $k$ are pairwise orthogonal. Using this property and the definition of $\widetilde{R}_{(k)}$, we have

$$\widetilde{R}_{(k)} := \left(\bar{R}_{(k)}\right)_{[\widehat{I}_k]} = \bar{R}_{[\widehat{I}_k]}\prod_{i=1}^{k-1}\mathbb{M}_{\widehat{V}_{(i)}^{\intercal}} = \bar{R}_{[\widehat{I}_k]}\left(\mathbb{I}_T - \sum_{i=1}^{k-1}\xi_{(i)}\xi_{(i)}^{\intercal}\right), \tag{B.39}$$

for $k > 1$ and when $k = 1$,

$$\widetilde{R}_{(1)} = \bar{R}_{[\widehat{I}_1]} = \beta_{[\widehat{I}_1]}\bar{V} + \bar{U}_{[\widehat{I}_1]}.$$

If we define $\widetilde{\beta}_{(1)} = \beta_{[\widehat{I}_1]}$ and $\widetilde{U}_{(1)} = \bar{U}_{[\widehat{I}_1]}$, then $\widetilde{R}_{(1)}$ can be written in the form $\widetilde{R}_{(1)} = \widetilde{\beta}_{(1)}\bar{V} + \widetilde{U}_{(1)}$. We can iteratively define

$$\widetilde{U}_{(k)} := \bar{U}_{[\widehat{I}_k]} - \sum_{i=1}^{k-1}\frac{\bar{R}_{[\widehat{I}_k]}\xi_{(i)}}{\sqrt{T}}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}}{\sqrt{\widetilde{\lambda}_{(i)}}} \quad \text{and} \quad \widetilde{\beta}_{(k)} := \beta_{[\widehat{I}_k]} - \sum_{i=1}^{k-1}\frac{\bar{R}_{[\widehat{I}_k]}\xi_{(i)}}{\sqrt{T}}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{\widetilde{\lambda}_{(i)}}}. \tag{B.40}$$

Recall that $\xi_{(k)}$ and $\varsigma_{(k)}$ are singular vectors of $\widetilde{R}_{(k)}$, we have

$$\varsigma_{(k)} = \frac{\widetilde{R}_{(k)}\xi_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}, \quad \xi_{(k)} = \frac{\widetilde{R}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}. \tag{B.41}$$

Using (B.41), if $\widetilde{R}_{(i)} = \widetilde{\beta}_{(i)}\bar{V} + \widetilde{U}_{(i)}$ for $i < k$, we can write (B.39) as

$$\begin{aligned}
\widetilde{R}_{(k)} = \bar{R}_{[\widehat{I}_k]}\left(\mathbb{I}_T - \sum_{i=1}^{k-1}\xi_{(i)}\xi_{(i)}^{\intercal}\right) =& \bar{R}_{[\widehat{I}_k]} - \sum_{i=1}^{k-1}\bar{R}_{[\widehat{I}_k]}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{R}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \\
=& \widetilde{\beta}_{[\widehat{I}_k]}\bar{V} + \widetilde{U}_{[\widehat{I}_k]} - \sum_{i=1}^{k-1}\bar{R}_{[\widehat{I}_k]}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}\bar{V}}{\sqrt{T\widehat{\lambda}_{(i)}}} - \sum_{i=1}^{k-1}\bar{R}_{[\widehat{I}_k]}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \\
=& \widetilde{\beta}_{(k)}\bar{V} + \widetilde{U}_{(k)}.
\end{aligned}$$

Consequently, by induction, $\widetilde{R}_{(k)} = \widetilde{\beta}_{(k)}\bar{V} + \widetilde{U}_{(k)}$ for $k \leq \tilde{p} + 1$. Similarly, we can write

$$\widetilde{r}_{(k)} = \widetilde{\beta}_{(k)}(\gamma + \bar{v}) + \widetilde{u}_{(k)}, \tag{B.42}$$

where $\widetilde{u}_{(k)}$ is defined by

$$\widetilde{u}_{(k)} := \bar{u}_{[\widehat{I}_k]} - \sum_{i=1}^{k-1} \frac{\bar{R}_{[\widehat{I}_k]}\xi_{(i)}}{\sqrt{T}} \frac{\varsigma_{(i)}^{\intercal}\widetilde{u}_{(i)}}{\sqrt{\widehat{\lambda}_{(i)}}}, \tag{B.43}$$

and $\widetilde{u}_{(1)} = \bar{u}_{[\widehat{I}_1]}$.

Similar representations can be created for $\widetilde{G}_{(k)} := \bar{G}\prod_{i=1}^{k-1}\mathbb{M}_{\widehat{V}_{(i)}^{\intercal}}$. Specifically, we have

$$\begin{aligned}\widetilde{G}_{(k)} := \bar{G}\left(\mathbb{I}_T - \sum_{i=1}^{k-1}\xi_{(i)}\xi_{(i)}^{\intercal}\right) =& \bar{G} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{R}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} = \eta\bar{V} + \bar{Z} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}\bar{V}}{\sqrt{T\widehat{\lambda}_{(i)}}} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \\ =& \left(\eta - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right)\bar{V} + \left(\bar{Z} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right).\end{aligned}$$

Using the following notation

$$\widetilde{\eta}_{(k)} := \eta - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}, \quad \text{and} \quad \widetilde{Z}_{(k)} := \bar{Z} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}, \tag{B.44}$$

$\widetilde{G}_{(k)}$ can be written as $\widetilde{G}_{(k)} = \widetilde{\eta}_{(k)}\bar{V} + \widetilde{Z}_{(k)}$.

To sum up, we have defined $\widetilde{R}_{(k)}, \widetilde{r}_{(k)}, \widetilde{\beta}_{(k)}, \widetilde{U}_{(k)}, \widetilde{u}_{(k)}, \widetilde{\eta}_{(k)}$ and $\widetilde{Z}_{(k)}$ at the $k$th step of the algorithm. Note that $\widetilde{\beta}_{(k)} \in \mathbb{R}^{|I_k|\times p}$ and $\widetilde{\eta}_{(k)} \in \mathbb{R}^{d\times p}$ can be viewed as the loading of $\widetilde{R}_{(k)}$ and $\widetilde{G}_{(k)}$ on $\bar{V}$, but they are not the same as the estimators defined in Algorithm 5, $\widehat{\beta}_{(k)} \in \mathbb{R}^{N\times 1}$ and $\widehat{\eta}_{(k)} \in \mathbb{R}^{d\times 1}$, which are the estimated loadings of $R$ and $G$ on the $k$th factor.

By Lemma 4, we have $\mathrm{P}(\widehat{I}_k = I_k) \to 1$ for $k \le \tilde{p}$ and $\mathrm{P}(\widehat{p} = \tilde{p}) \to 1$. Thus, we can impose that $\widehat{I}_k = I_k$ for any $k$ and $\widehat{p} = \tilde{p}$ in what follows. In addition, Lemma 3(ii) and Lemma 4(iii) imply that $\widehat{\lambda}_{(k)} \asymp qN$ and that $|I_k| = qN$. Therefore, the assumptions of Lemmas 6-9 hold.

Since our algorithm stops at $\tilde{p}$, it implies that at most $qN-1$ test assets satisfy $T^{-1}\left\|\left(\bar{R}_{(\tilde{p}+1)}\right)_{[i]}\bar{G}^{\intercal}\right\|_{\mathrm{MAX}} \ge c$. Consider the test assets in $I_0$, we have

$$T^{-1}\left\|\widetilde{G}_{(\tilde{p}+1)}\bar{R}_{[I_0]}^{\intercal}\right\| = T^{-1}\left\|\left(\bar{R}_{(\tilde{p}+1)}\right)_{[I_0]}\bar{G}^{\intercal}\right\| \lesssim_p q^{1/2}N^{1/2} + cN_0^{1/2} = o\left(N_0^{1/2}\right), \tag{B.45}$$

where we use the the assumptions $c \to 0$ and $qN/N_0 \to 0$ in the last equation.

Write the left hand side of (B.45) as

$$\widetilde{G}_{(\tilde{p}+1)}\bar{R}_{[I_0]}^{\intercal} = \widetilde{\eta}_{(\tilde{p}+1)}\bar{V}\bar{V}^{\intercal}\beta_{[I_0]} + \widetilde{\eta}_{(\tilde{p}+1)}\bar{V}\bar{U}_{[I_0]}^{\intercal} + \bar{Z}_{(\tilde{p}+1)}\bar{V}^{\intercal}\beta_{[I_0]} + \bar{Z}_{(\tilde{p}+1)}\bar{U}_{[I_0]}^{\intercal}. \tag{B.46}$$

Using (B.45), (B.46) and Lemma 8(i)(ii), we have

$$\left\|\widetilde{\eta}_{(\tilde{p}+1)}\left(\bar{V}\bar{V}^{\intercal}\beta_{[I_0]}+\bar{V}\bar{U}_{[I_0]}^{\intercal}\right)\right\|=o_p\left(N_0^{1/2}T\right). \tag{B.47}$$

Also, using Assumption A.6, Lemma 1(i) and Weyl's theorem, we have

$$|\sigma_p(\bar{V}\bar{V}^{\intercal}\beta_{[I_0]}+\bar{V}\bar{U}_{[I_0]}^{\intercal})-\sigma_p(T\beta_{[I_0]})|\leq\left\|\bar{V}\bar{U}_{[I_0]}^{\intercal}\right\|+\left\|T^{-1}\bar{V}\bar{V}^{\intercal}-\mathbb{I}_p\right\|\left\|T\beta_{[I_0]}\right\|\lesssim_p N_0^{1/2}T^{1/2}. \tag{B.48}$$

Since Assumption A.3 implies that $\sigma_p(\beta_{[I_0]})\asymp N_0^{1/2}$, we have $\sigma_p(\bar{V}\bar{V}^{\intercal}\beta_{[I_0]}+\bar{V}\bar{U}_{[I_0]}^{\intercal})\asymp_p N_0^{1/2}T$. Using this result, (B.47) and the inequality $\left\|\widetilde{\eta}_{(\tilde{p}+1)}\left(\bar{V}\bar{V}^{\intercal}\beta_{[I_0]}+\bar{V}\bar{U}_{[I_0]}^{\intercal}\right)\right\|\geq\sigma_p(\bar{V}\bar{V}^{\intercal}\beta_{[I_0]}+\bar{V}\bar{U}_{[I_0]}^{\intercal})\left\|\widetilde{\eta}_{(\tilde{p}+1)}\right\|$, we have $\left\|\widetilde{\eta}_{(\tilde{p}+1)}\right\|\xrightarrow{p}0$. That is, by definition of $\tilde{\eta}_{(\tilde{p}+1)}$ in (B.44),

$$\left\|\eta-\sum_{i=1}^{\tilde{p}}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\|=o_p(1). \tag{B.49}$$

Multiplying (B.49) by $\gamma$ from the right-hand side, we have

$$\left\|\eta\gamma-\sum_{i=1}^{\tilde{p}}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\gamma\right\|=o_p(1). \tag{B.50}$$

Recall that our final estimator of $\gamma_g$ is

$$\widehat{\gamma}_g^{SPCA}=\widehat{\eta}\widehat{\gamma}=\sum_{i=1}^{\tilde{p}}\widehat{\eta}_{(i)}\widehat{\gamma}_{(i)}=\sum_{i=1}^{\tilde{p}}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{r}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}=\sum_{i=1}^{\tilde{p}}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}(\gamma+\bar{v})+\sum_{i=1}^{\tilde{p}}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{u}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}. \tag{B.51}$$

Combining (B.50) and (B.51), we have

$$\|\eta\gamma-\widehat{\eta}\widehat{\gamma}\|\leq\sum_{i=1}^{\tilde{p}}\left\|\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\bar{v}\right\|+\sum_{i=1}^{\tilde{p}}\left\|\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{u}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\|+o_p(1). \tag{B.52}$$

Using $\left\|\bar{G}\right\|\lesssim_p T^{1/2}$, Lemma 7(ii), Lemma 9(i) and the assumptions that $qN\to\infty$, we have

$$\left\|\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\bar{v}\right\|\leq\left\|\bar{G}\xi_{(i)}\right\|\left\|\frac{\varsigma_{(i)}^{\intercal}\widetilde{\beta}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\|\|\bar{v}\|=o_p(1),$$

and

$$\left\| \bar{G}\xi_{(i)} \frac{\varsigma_{(i)}^{\intercal} \widetilde{u}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \right\| \leq \left\| \bar{G}\xi_{(i)} \right\| \left\| \frac{\varsigma_{(i)}^{\intercal} \widetilde{u}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \right\| = o_p(1).$$

Plugging them into (B.52) completes the proof.

## B.7 Proof of Theorem 2

To derive the asymptotic distribution, we need a more intricate analysis. As in the proof of Theorem 1, we impose that $\widehat{p} = \tilde{p}$ and $\widehat{I}_k = I_k$, since Lemma 4 shows that both events occur with probability approaching 1.

Recall that in Algorithm 5 the SPCA estimator is written as $\widehat{\gamma}_g^{SPCA} = \widehat{\eta}\widehat{\gamma} = \sum_{k=1}^{\widehat{p}} \widehat{\eta}_{(k)}\widehat{\gamma}_{(k)}$, where $\widehat{p}$ is the number of factors selected and, with the notation defined in the proof of Theorem 1,

$$\widehat{\eta}_{(k)} = \frac{\bar{G}\xi_{(k)}}{\sqrt{T}} = \frac{\eta\bar{V}\xi_{(k)}}{\sqrt{T}} + \frac{\bar{Z}\xi_{(k)}}{\sqrt{T}}, \qquad \widehat{\gamma}_{(k)} = \frac{\varsigma_{(k)}^{\intercal}\widetilde{r}_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}} = \frac{\varsigma_{(k)}^{\intercal}\widetilde{\beta}_{(k)}(\gamma + \bar{v})}{\sqrt{\widehat{\lambda}_{(k)}}} + \frac{\varsigma_{(k)}^{\intercal}\widetilde{u}_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}. \tag{B.53}$$

Denote $H_1 = (h_{11}, \ldots, h_{\widehat{p}1})$, $H_2 = (h_{12}, \ldots, h_{\widehat{p}2})$, where

$$h_{k1} = T^{-1/2}\bar{V}\xi_{(k)}, \quad h_{k2} = \widehat{\lambda}_{(k)}^{-1/2}\widetilde{\beta}_{(k)}^{\intercal}\varsigma_{(k)}. \tag{B.54}$$

Therefore, we can write (B.53) as

$$\widehat{\eta}_{(k)} - \eta h_{k1} = \frac{\bar{Z}\xi_{(k)}}{\sqrt{T}}, \quad \widehat{\gamma}_{(k)} - h_{k2}^{\intercal}(\gamma + \bar{v}) = \frac{\varsigma_{(k)}^{\intercal}\widetilde{u}_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}. \tag{B.55}$$

Since $\xi_{(k)}$ and $\varsigma_{(k)}$ are the largest singular vectors of $\widetilde{R}_{(k)}$ with the singular value $\sqrt{T\widehat{\lambda}_{(k)}}$, we have

$$\varsigma_{(k)} = \frac{\widetilde{R}_{(k)}\xi_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}, \quad \xi_{(k)} = \frac{\widetilde{R}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}. \tag{B.56}$$

From (B.56), we have

$$\frac{\bar{Z}\xi_{(k)}}{\sqrt{T}} = \frac{\bar{Z}}{\sqrt{T}}\frac{\widetilde{R}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} = \frac{\bar{Z}\bar{V}^{\intercal}}{T}\frac{\widetilde{\beta}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}} + \frac{\bar{Z}\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} = \frac{\bar{Z}\bar{V}^{\intercal}}{T}h_{k2} + \frac{\bar{Z}\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}}.$$

Using Lemma 7(ii) and the assumptions on $q$, we have

$$\left\| \frac{\bar{Z}\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| = o_p(T^{-1/2}), \quad \left\| \frac{\varsigma_{(k)}^{\intercal}\widetilde{u}_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}} \right\| = o_p(T^{-1/2}).$$

Then, along with (B.55) and Lemma 1(vi), the above equations lead to

$$\left\| \widehat{\eta} - \eta H_1 - \frac{ZV}{T} H_2 \right\| = o_p(T^{-1/2}), \tag{B.57}$$

and

$$\|\widehat{\gamma} - H_2^{\intercal}\gamma - H_2^{\intercal}\bar{v}\| = o_p(T^{-1/2}). \tag{B.58}$$

Combining (B.57) and (B.58), with $\|H_1\| \lesssim_p 1$, $\|H_2\| \lesssim_p 1$ from Lemma 9 and Assumptions A.1, A.2, we have

$$\left\| \widehat{\eta}\widehat{\gamma} - \eta H_1 H_2^{\intercal}(\gamma + \bar{v}) - \frac{ZV^{\intercal}}{T} H_2 H_2^{\intercal}\gamma \right\| = o_p(T^{-1/2}). \tag{B.59}$$

As shown in Lemma 3(iv), under the assumption that $\lambda_p(\eta^{\intercal}\eta) \gtrsim 1$, we have $\tilde{p} = p$. Together with $\mathrm{P}(\widehat{p} = \tilde{p}) \to 1$, we can impose that $\widehat{p} = p$ for derivations below. To analyze $H_1 H_2^{\intercal}$ and $H_2 H_2^{\intercal}$ in (B.59), using Lemma 9 and the assumptions on $q$, we have

$$\|H_2^{\intercal} H_2 - \mathbb{I}_p\| \leq \|H_1^{\intercal} H_2 - \mathbb{I}_p\| + \|H_1 - H_2\| \, \|H_2\| \lesssim_p T^{-1/2}. \tag{B.60}$$

Then, for the term $H_2 H_2^{\intercal}$, we have

$$\|H_2 H_2^{\intercal} - \mathbb{I}_p\| = \max_{1 \leq i \leq p} |\lambda_i(H_2 H_2^{\intercal}) - 1| = \max_{1 \leq i \leq p} |\lambda_i(H_2^{\intercal} H_2) - 1| = \|H_2^{\intercal} H_2 - \mathbb{I}_p\| \lesssim_p T^{-1/2} \tag{B.61}$$

since $H_2$ is a $p \times p$ matrix.

For the term $H_1 H_2^{\intercal}$, by Lemma 9 and the assumptions on $q$, we have

$$\|H_1^{\intercal} H_2 - \mathbb{I}_p\| = o_p(T^{-1/2}). \tag{B.62}$$

In addition, we have

$$\sigma_p(H_2) \, \|H_2 H_1^{\intercal} - \mathbb{I}_p\| \leq \|(H_2 H_1^{\intercal} - \mathbb{I}_p) H_2\| = \|H_2(H_1^{\intercal} H_2 - \mathbb{I}_p)\| \leq \|H_2\| \, \|H_1^{\intercal} H_2 - \mathbb{I}_p\| \, . \tag{B.63}$$

Since (B.60) implies that $\sigma_1(H_2)/\sigma_p(H_2) = \lambda_1(H_2 H_2^{\intercal})^{1/2} / \lambda_p(H_2 H_2^{\intercal})^{1/2} \lesssim_p 1$, (B.62) and (B.63) give

$$\|H_1 H_2^{\intercal} - \mathbb{I}_p\| = \|H_2 H_1^{\intercal} - \mathbb{I}_p\| \leq \frac{\sigma_1(H_2)}{\sigma_p(H_2)} \, \|H_1^{\intercal} H_2 - \mathbb{I}_p\| = o_p(T^{-1/2}). \tag{B.64}$$

Combining (B.59), (B.61), and (B.64), we obtain $\left\|\widehat{\eta}\widehat{\gamma} - \eta(\gamma+\bar{v}) - T^{-1}ZV^{\intercal}\gamma\right\| = o_p(T^{-1/2})$. Using Delta method and Assumption A.9, it is straightforward to obtain: $\sqrt{T}\left(\widehat{\eta}\widehat{\gamma} - \eta\gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \Phi\right)$, where $\Phi$ is as defined in Theorem 2. □

## B.8 Proof of Theorem 3

*Proof.* As in the proof of Theorem 2, we have

$$\left\|\widehat{\eta}\widehat{\gamma} - \eta(\gamma+\bar{v}) - T^{-1}ZV^{\intercal}\gamma\right\| = o_p(T^{-1/2})$$

from (B.59), (B.61), and (B.64). Together with $\widehat{\alpha}_g = \bar{g} - \widehat{\eta}\widehat{\gamma} = \alpha_g + \eta\gamma + \eta\bar{v} + \bar{z} - \widehat{\eta}\widehat{\gamma}$, we have

$$\left\|\widehat{\alpha}_g - \alpha_g - \bar{z} + T^{-1}ZV^{\intercal}\gamma\right\| = \left\|\widehat{\eta}\widehat{\gamma} - \eta(\gamma+\bar{v}) - T^{-1}ZV^{\intercal}\gamma\right\| = o_p(T^{-1/2})$$

Using Delta method and the CLT Assumption in Theorem 3 , we have $\sqrt{T}(\widehat{\alpha}_g - \alpha_g) \xrightarrow{d} \mathcal{N}(0, \widetilde{\Phi})$ where $\widetilde{\Phi}$ is as defined in Theorem 3. □

## B.9 Proof of Theorem 4

*Proof.* As shown in the proof of Theorem 2, we have $\mathrm{P}(\widehat{p} = p) \to 1$ and $\mathbb{P}(\widehat{I}_k = I_k) \to 1$ for $k \leq p$. Thus, we impose $\widehat{p} = \tilde{p} = p$ and $\widehat{I}_k = I_k$ below. Using the same notation as in the proof of Theorem 2 and (B.58), we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}|m_t - \widehat{m}_t|^2 = \frac{1}{T}\left\|\widehat{V}^{\intercal}\widehat{\gamma} - V^{\intercal}\gamma\right\|^2 &= \frac{1}{T}\left\|\sqrt{T}\xi(H_2^{\intercal}\gamma + O_p(T^{-1/2})) - V^{\intercal}\gamma\right\|^2 \\ &= \frac{1}{T}\left\|\sqrt{T}\xi H_2^{\intercal}\gamma - \bar{V}^{\intercal}\gamma\right\|^2 + O_p\left(T^{-1}\right), \end{aligned} \tag{B.65}$$

where $\xi = (\xi_{(1)}, \ldots, \xi_{(p)})$.

Using (B.56), we can write

$$\sqrt{T}\xi_{(k)}h_{k2}^{\intercal} = \frac{\widetilde{R}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}h_{k2}^{\intercal} = \frac{\bar{V}^{\intercal}\widetilde{\beta}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}h_{k2}^{\intercal} + \frac{\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}h_{k2}^{\intercal}. \tag{B.66}$$

Using Lemma 7(i), Lemma 9(i) and $\widehat{\lambda}_{(k)} \asymp_p |I_k|$, $|I_k| = qN$, we can derive from (B.66) that

$$\sqrt{T}\xi_{(k)}h_{k2}^{\intercal} = \bar{V}^{\intercal}h_{k2}h_{k2}^{\intercal} + O_p\left(q^{-1/2}N^{-1/2}T^{1/2} + T^{-1/2}\right).$$

That is,

$$\sqrt{T}\xi H_2^{\intercal} = \bar{V}^{\intercal}H_2H_2^{\intercal} + O_p\left(q^{-1/2}N^{-1/2}T^{1/2} + T^{-1/2}\right). \tag{B.67}$$

Therefore, using (B.67), (B.61) and the assumptions on $q$, we have

$$\begin{aligned} T^{-1/2}\left\|\sqrt{T}\xi H_2^\intercal\gamma - \bar{V}^\intercal\gamma\right\| &\lesssim_p T^{-1/2}\left\|\bar{V}^\intercal H_2 H_2^\intercal - \bar{V}^\intercal\right\|\|\gamma\| + q^{-1/2}N^{-1/2} + T^{-1} \\ &\lesssim_p T^{-1/2}\left\|\bar{V}\right\|\|H_2 H_2^\intercal - \mathbb{I}_p\| + q^{-1/2}N^{-1/2} + T^{-1} \\ &\lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}. \end{aligned}$$

Therefore, it follows from (B.65) that

$$\frac{1}{T}\sum_{t=1}^{T}|m_t - \widehat{m}_t|^2 = \frac{1}{T}\left\|\widehat{V}^\intercal\widehat{\gamma} - V^\intercal\gamma\right\|^2 \lesssim_p \frac{1}{T} + \frac{1}{qN}.$$

In light of the assumptions on $q$, we can choose $q$ such that $qN \gtrsim N_0/\log N_0$, which leads to

$$\frac{1}{T}\sum_{t=1}^{T}|m_t - \widehat{m}_t|^2 \lesssim_p \frac{1}{T} + \frac{\log N_0}{N_0}.$$

□

## B.10 Proof of Proposition 6

*Proof.* Write $\widetilde{\beta} = \Sigma_u^{-1/2}\beta\Sigma_v^{1/2}$, then by definition $\widetilde{m}_t$ can be written as

$$\widetilde{m}_t = 1 - \gamma^\intercal\beta^\intercal\Sigma_r^{-1}(\beta v_t + u_t) = 1 - \gamma^\intercal\Sigma_v^{-1/2}\tilde{\beta}^\intercal\left(\widetilde{\beta}\widetilde{\beta}^\intercal + \mathbb{I}_N\right)^{-1}(\widetilde{\beta}\Sigma_v^{-1/2}v_t + \Sigma_u^{-1/2}u_t), \tag{B.68}$$

or in matrix form

$$\widetilde{M} = 1 - \gamma^\intercal\beta^\intercal\Sigma_r^{-1}(\beta V + U) = 1 - \gamma^\intercal\Sigma_v^{-1/2}\tilde{\beta}^\intercal\left(\widetilde{\beta}\widetilde{\beta}^\intercal + \mathbb{I}_N\right)^{-1}(\widetilde{\beta}\Sigma_v^{-1/2}V + \Sigma_u^{-1/2}U), \tag{B.69}$$

where $\widetilde{M} = (\widetilde{m}_1, \ldots, \widetilde{m}_T)$, $V = (v_1, \ldots, v_T)$ and $U = (u_1, \ldots, u_t)$. Suppose that the SVD of $\widetilde{\beta}$ can be written as $\widetilde{\beta} = B\Lambda^{1/2}\Gamma$, where $B \in \mathbb{R}^{N\times p}$ and $\Gamma \in \mathbb{R}^{p\times p}$ are matrices of left and right singular vectors, $\Lambda^{1/2} = \mathrm{diag}(\widetilde{\lambda}_1^{1/2}, \cdots, \widetilde{\lambda}_p^{1/2})$ is a diagonal matrix and $\widetilde{\lambda}_i$ is the $i$th eigenvalue of $\widetilde{\beta}^\intercal\widetilde{\beta}$. Write $B = (b_1, \cdots, b_p)$, then $b_i^\intercal b_j = 0$ for $i \neq j$. Using the SVD of $\widetilde{\beta}$, we have

$$\widetilde{\beta}^\intercal\left(\widetilde{\beta}\widetilde{\beta}^\intercal + \mathbb{I}_N\right)^{-1} = \Gamma^\intercal\Lambda^{1/2}(\Lambda + \mathbb{I}_p)^{-1}B^\intercal.$$

Hence, we have

$$\left\|\widetilde{\beta}^\intercal\left(\widetilde{\beta}\widetilde{\beta}^\intercal + \mathbb{I}_N\right)^{-1}\widetilde{\beta} - \mathbb{I}_p\right\| = \left\|\Gamma^\intercal\Lambda^{1/2}(\Lambda + \mathbb{I}_p)^{-1}\Lambda^{1/2}\Gamma - \mathbb{I}_p\right\| = \left\|\Lambda^{1/2}(\Lambda + \mathbb{I}_p)^{-1}\Lambda^{1/2} - \mathbb{I}_p\right\| \lesssim_p \widetilde{\lambda}_p^{-1}, \tag{B.70}$$

and

$$\left\|\widetilde{\beta}^{\intercal}\left(\widetilde{\beta}\widetilde{\beta}^{\intercal}+\mathbb{I}_N\right)^{-1}\Sigma_u^{-1/2}U\right\| = \left\|\Gamma^{\intercal}\Lambda^{1/2}(\Lambda+\mathbb{I}_p)^{-1}B^{\intercal}\Sigma_u^{-1/2}U\right\| \lesssim_p \left(\widetilde{\lambda}_p^{-1/2}\right)\left\|B^{\intercal}\Sigma_u^{-1/2}U\right\|. \quad \text{(B.71)}$$

Since $\mathrm{Cov}(B^{\intercal}\Sigma_u^{-1/2}u_t)=\mathbb{I}_p$, we have $\mathrm{E}\left(\left\|B^{\intercal}\Sigma_u^{-1/2}U\right\|_{\mathrm{F}}^2\right)=pT$, which leads to

$$\left\|B^{\intercal}\Sigma_u^{-1/2}U\right\| \le \left\|B^{\intercal}\Sigma_u^{-1/2}U\right\|_{\mathrm{F}} \lesssim_p T^{1/2}. \quad \text{(B.72)}$$

For the same reason, we have $\left\|\Sigma_v^{-1/2}V\right\| \lesssim_p T^{1/2}$. Then, with Assumption A.11, (B.69), (B.70), (B.71), and (B.72), we have

$$\begin{aligned}\sqrt{\sum_{t=1}^{T}|m_t-\widetilde{m}_t|^2} &\le \left\|\gamma^{\intercal}\Sigma_v^{-1/2}\left(\widetilde{\beta}^{\intercal}\left(\widetilde{\beta}\widetilde{\beta}^{\intercal}+\mathbb{I}_N\right)^{-1}\widetilde{\beta}-\mathbb{I}_p\right)\Sigma_v^{-1/2}V\right\| + \left\|\gamma^{\intercal}\Sigma_v^{-1}\widetilde{\beta}^{\intercal}\left(\widetilde{\beta}\widetilde{\beta}^{\intercal}+\mathbb{I}_N\right)^{-1}\Sigma_u^{-1/2}U\right\| \\ &\lesssim \left\|\widetilde{\beta}^{\intercal}\left(\widetilde{\beta}\widetilde{\beta}^{\intercal}+\mathbb{I}_N\right)^{-1}\widetilde{\beta}-\mathbb{I}_p\right\|\left\|\Sigma_v^{-1/2}V\right\| + \left\|\widetilde{\beta}^{\intercal}\left(\widetilde{\beta}\widetilde{\beta}^{\intercal}+\mathbb{I}_N\right)^{-1}\Sigma_u^{-1/2}U\right\| \\ &\lesssim_p T^{1/2}\widetilde{\lambda}_p^{-1/2},\end{aligned}$$

which in turn leads to

$$\frac{1}{T}\sum_{t=1}^{T}|m_t-\widetilde{m}_t|^2 \lesssim_p \widetilde{\lambda}_p^{-1},$$

where

$$\widetilde{\lambda}_p = \lambda_p\left(\Sigma_v^{1/2}\beta^{\intercal}\Sigma_u^{-1}\beta\Sigma_v^{1/2}\right) \ge \lambda_p(\beta\Sigma_v\beta^{\intercal})\lambda_{\min}(\Sigma_u^{-1}) \asymp_p \lambda_p(\beta^{\intercal}\beta)\lambda_{\max}^{-1}(\Sigma_u) \gtrsim \lambda_p(\beta^{\intercal}\beta),$$

which concludes the proof. □

### B.11 Proof of Theorem 5(a)

*Proof.* For Ridge SDF estimator $\widehat{m}_t$, we have

$$\frac{1}{T}\sum_{t=1}^{T}|m_t-\widehat{m}_t|^2 = \frac{1}{T}\left\|\bar{R}^{\intercal}(\widehat{\Sigma}+\mu\mathbb{I}_N)^{-1}\bar{r}-V^{\intercal}\gamma\right\|^2. \quad \text{(B.73)}$$

Recall that in the proof of Proposition 3, we have a condensed form of SVD on $\bar{R}$:

$$\bar{R} = \sqrt{T}\varsigma\widehat{\Lambda}^{1/2}\xi^{\intercal} + \sqrt{T}\varsigma_*\widehat{\Lambda}_*^{1/2}\xi_*^{\intercal},$$

where $\widehat{\Lambda}^{1/2}$ is the diagonal matrix of the first $p$ singular values of $T^{-1/2}\bar{R}$ and $\widehat{\Lambda} = \mathrm{diag}\{\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p\}$, $\varsigma$, $\xi$ are the corresponding left and right singular vectors, and $\varsigma_* \in \mathbb{R}^{N\times K}$, $\xi_* \in \mathbb{R}^{T\times K}$ are the singular vectors corresponding to the remaining $K$ nonzero singular values in $\widehat{\Lambda}_*^{1/2} \in \mathbb{R}^{K\times K}$, where $K = \min\{N, T-1\} - p$. Using this representation, (B.73) becomes

$$\begin{aligned}\sqrt{\sum_{t=1}^{T}|m_t - \widehat{m}_t|^2} &= \left\|(\bar{V}^\intercal\beta^\intercal + \bar{U}^\intercal)\varsigma(\widehat{\Lambda}+\mu I)^{-1}\varsigma^\intercal\bar{r} - V^\intercal\gamma + (\bar{V}^\intercal\beta^\intercal + \bar{U}^\intercal)\varsigma_*(\widehat{\Lambda}_*+\mu I)^{-1}\varsigma_*^\intercal\bar{r}\right\| \\ &\le \left\|\bar{V}^\intercal\beta^\intercal\varsigma(\widehat{\Lambda}+\mu I)^{-1}\varsigma^\intercal\beta\gamma - \bar{V}^\intercal\gamma\right\| + \left\|\bar{V}^\intercal\beta^\intercal\varsigma(\widehat{\Lambda}+\mu I)^{-1}\varsigma^\intercal(\beta\bar{v}+\bar{u})\right\| \\ &\quad + \left\|\bar{U}^\intercal\varsigma(\widehat{\Lambda}+\mu I)^{-1}\varsigma^\intercal\bar{r}\right\| + \left\|\bar{V}^\intercal\beta^\intercal\varsigma_*(\widehat{\Lambda}_*+\mu I)^{-1}\varsigma_*^\intercal\bar{r}\right\| \\ &\quad + \left\|\bar{U}^\intercal\varsigma_*(\widehat{\Lambda}_*+\mu I)^{-1}\varsigma_*^\intercal\bar{r}\right\| + \left\|V^\intercal\gamma - \bar{V}^\intercal\gamma\right\| \end{aligned} \tag{B.74}$$

We analyze these terms one-by-one. Firstly, we consider the properties of $\varsigma$ and $\xi$. Let $\varsigma_k$ and $\xi_k$ denote the $k$th columns of $\varsigma$ and $\xi$, respectively. Note that $\varsigma_k$ and $\xi_k$ can be regarded as the $\varsigma_{(k)}$ and $\xi_{(k)}$ in our SPCA procedure with $I_k = [N]$, where $\varsigma_k$ and $\xi_k$ are the singular vectors at the $k$th stage. This means we can reuse some of the proofs without repeating essentially the same arguments therein.

Similar to (B.54), we define

$$\tilde{h}_{k1} = T^{-1/2}\bar{V}\xi_k, \quad \tilde{h}_{k2} = \widehat{\lambda}_k^{-1/2}\beta^\intercal\varsigma_k, \tag{B.75}$$

and $\tilde{H}_1 = (\tilde{h}_{11}, \ldots, \tilde{h}_{p1})$, $\tilde{H}_2 = (h_{12}, \ldots, \tilde{h}_{p2})$. Using Lemma 14, we can obtain

$$\left\|\tilde{H}_1\tilde{H}_2^\intercal - \mathbb{I}_p\right\| \lesssim_p T^{-1} + \lambda_p^{-1}(T^{-1}N + 1), \quad \left\|\tilde{H}_1 - \tilde{H}_2\right\| \lesssim_p T^{-1/2} + \lambda_p^{-1}(T^{-1}N+1). \tag{B.76}$$

Using (B.76) and Lemma 14(i), we have $\left\|\tilde{H}_2\tilde{H}_2^\intercal - \mathbb{I}_p\right\| \le \left\|\tilde{H}_1\tilde{H}_2^\intercal - \mathbb{I}_p\right\| + \left\|\tilde{H}_1 - \tilde{H}_2\right\|\left\|\tilde{H}_2\right\| \lesssim_p T^{-1/2} + \lambda_p^{-1}(T^{-1}N+1)$, which, by (B.75), is equivalent to

$$\left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1}\varsigma^\intercal\beta - \mathbb{I}_p\right\| \lesssim_p \frac{1}{\sqrt{T}} + \frac{N+T}{T\lambda_p}. \tag{B.77}$$

Consequently, with Lemma 11 and $\left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1/2}\right\| = \left\|\tilde{H}_2\right\| \lesssim_p 1$, we have

$$\begin{aligned}\left\|\beta^\intercal\varsigma\left(\widehat{\Lambda}+\mu I\right)^{-1}\varsigma^\intercal\beta - \mathbb{I}_p\right\| &\le \left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1/2}\left(\widehat{\Lambda}^{1/2}\left(\widehat{\Lambda}+\mu I\right)^{-1}\widehat{\Lambda}^{1/2} - \mathbb{I}_p\right)\widehat{\Lambda}^{-1/2}\varsigma^\intercal\beta\right\| + \left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1}\varsigma^\intercal\beta - \mathbb{I}_p\right\| \\ &\le \left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1/2}\right\|^2\left\|\widehat{\Lambda}^{1/2}\left(\widehat{\Lambda}+\mu I\right)^{-1}\widehat{\Lambda}^{1/2} - \mathbb{I}_p\right\| + \left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1}\varsigma^\intercal\beta - \mathbb{I}_p\right\| \\ &\lesssim_p \frac{1}{\sqrt{T}} + \frac{N+T}{T\lambda_p} + \frac{\mu}{\lambda_p}, \end{aligned} \tag{B.78}$$

where we use $\left\|\widehat{\Lambda}^{1/2}\left(\widehat{\Lambda}+\mu I\right)^{-1}\widehat{\Lambda}^{1/2}-\mathbb{I}_p\right\| = \max_{j\leq p}(\widehat{\lambda}_j+\mu)^{-1}\mu \lesssim_p \lambda_p^{-1}\mu$ in the last step.

With $\left\|\bar{V}\right\| \lesssim_p T^{1/2}$ from Lemma 1, it implies from (B.78) that the first term in (B.74) can be bounded:

$$\left\|\bar{V}^\intercal\beta^\intercal\varsigma(\widehat{\Lambda}+\mu I)^{-1}\varsigma^\intercal\beta\gamma-\bar{V}^\intercal\gamma\right\| \lesssim_p 1+\frac{N+T}{\sqrt{T}\lambda_p}+\frac{\mu\sqrt{T}}{\lambda_p}.$$

For the second term in (B.74), using Lemma 11, we have

$$\left\|\bar{V}^\intercal\beta^\intercal\varsigma(\widehat{\Lambda}+\mu I)^{-1}\varsigma^\intercal(\beta\bar{v}+\bar{u})\right\| \leq \left\|\bar{V}\right\|\left\|\beta^\intercal\varsigma\widehat{\Lambda}^{-1/2}\right\|\left\|\widehat{\Lambda}^{1/2}(\widehat{\Lambda}+\mu I)^{-1}\right\|\left\|\beta\bar{v}+\bar{u}\right\| \lesssim_p \sqrt{\frac{N}{\lambda_p}}. \tag{B.79}$$

Next, recall that $\varsigma_*$ and $\xi_*$ are singular vectors of $\bar{R}$, we have

$$\bar{V}^\intercal\beta^\intercal\varsigma_*+\bar{U}^\intercal\varsigma_* = \bar{R}^\intercal\varsigma_* = \sqrt{T}\xi_*\widehat{\Lambda}_*^{1/2}. \tag{B.80}$$

By Weyl's theorem and Assumption A.4, we have

$$|\sigma_j(T^{-1/2}\bar{R})-\sigma_j(T^{-1/2}\beta\bar{V})| \leq T^{-1/2}\left\|\bar{R}-\beta\bar{V}\right\| = T^{-1/2}\left\|\bar{U}\right\| \lesssim_p \sqrt{\frac{N}{T}}+1, \tag{B.81}$$

for $j\leq\min\{N,T\}$. Since $\mathrm{Rank}(T^{-1/2}\beta\bar{V})\leq p$, we have $\sigma_j(T^{-1/2}\beta\bar{V})=0$ for $j>p$ and thus

$$\left\|\widehat{\Lambda}_*^{1/2}\right\| = \sigma_{p+1}(T^{-1/2}\bar{R}) \lesssim_p \sqrt{\frac{N}{T}}+1. \tag{B.82}$$

Left multiplying (B.80) by $\bar{V}$, we obtain

$$\bar{V}\bar{V}^\intercal\beta^\intercal\varsigma_* = \sqrt{T}\bar{V}\xi_*\widehat{\Lambda}_*^{1/2}-\bar{V}\bar{U}^\intercal\varsigma_*. \tag{B.83}$$

Together with (B.82) and Assumption A.6, we have

$$\left\|\beta^\intercal\varsigma_*\right\| \leq \left\|\left(\bar{V}\bar{V}^\intercal\right)^{-1}\right\|\left(\sqrt{T}\left\|\bar{V}\right\|\left\|\widehat{\Lambda}_*^{1/2}\right\|+\left\|\bar{V}\bar{U}^\intercal\right\|\right) \lesssim_p \sqrt{\frac{N}{T}}+1, \tag{B.84}$$

and consequently,

$$\left\|\varsigma_*^\intercal\bar{r}\right\| \leq \left\|\varsigma_*^\intercal\beta\right\|\left\|\gamma+\bar{v}\right\|+\left\|\varsigma_*^\intercal\bar{u}\right\| \lesssim_p \sqrt{\frac{N}{T}}+1. \tag{B.85}$$

Using (B.84), (B.85), Lemma 13(iv) and $\left\|\bar{U}\right\| \lesssim_p N^{1/2}+T^{1/2}$, we have

$$\left\|\beta^\intercal\varsigma_*(\widehat{\Lambda}_*+\mu I)^{-1}\varsigma_*^\intercal\bar{r}\right\| \leq \left\|\beta^\intercal\varsigma_*\right\|\left\|(\widehat{\Lambda}_*+\mu I)^{-1}\right\|\left\|\varsigma_*^\intercal\bar{r}\right\| \lesssim_p \frac{N+T}{\mu T}, \tag{B.86}$$

and

$$\left\|\bar{U}^{\intercal}\varsigma_*(\widehat{\Lambda}_* + \mu I)^{-1}\varsigma_*^{\intercal}\bar{r}\right\| \leq \left\|\bar{U}\right\| \left\|(\widehat{\Lambda}_* + \mu I)^{-1}\right\| \left\|\varsigma_*^{\intercal}\bar{r}\right\| \lesssim_p \frac{N+T}{\mu\sqrt{T}}. \tag{B.87}$$

Using Lemma 13(iii), we have

$$\left\|\widehat{\Lambda}^{-1/2}\varsigma^{\intercal}\bar{r}\right\| \lesssim_p \left\|\widehat{\Lambda}^{-1/2}\varsigma^{\intercal}\beta\right\| + \left\|\widehat{\Lambda}^{-1/2}\varsigma^{\intercal}\bar{u}\right\| \lesssim_p 1 + \frac{N+T}{T\lambda_p} \lesssim_p 1,$$

where we use $\left\|\widehat{\Lambda}^{-1/2}\varsigma^{\intercal}\beta\right\| = \left\|\tilde{H}_2\right\| \lesssim_p 1$. Then, with Lemma 13(iv), we have

$$\left\|\bar{U}^{\intercal}\varsigma(\widehat{\Lambda} + \mu I)^{-1}\varsigma^{\intercal}\bar{r}\right\| \leq \left\|\bar{U}^{\intercal}\varsigma\right\| \left\|(\widehat{\Lambda} + \mu I)^{-1}\widehat{\Lambda}^{1/2}\right\| \left\|\widehat{\Lambda}^{-1/2}\varsigma^{\intercal}\bar{r}\right\| \lesssim_p \sqrt{\frac{T}{\lambda_p}} + \frac{N+T}{\sqrt{T}\lambda_p}. \tag{B.88}$$

Plugging (B.78), (B.79), (B.86), (B.87) and (B.88) into (B.74) and using $\left\|\bar{V} - V\right\| \lesssim_p 1$, we obtain

$$\frac{1}{T}\sum_{t=1}^{T}|m_t - \widehat{m}_t|^2 \lesssim_p \frac{\mu^2}{\lambda_p^2} + \frac{1}{T} + \frac{N+T}{T\lambda_p} + \frac{N^2+T^2}{\mu^2 T^2}.$$

With $\mu^2 \asymp T^{-1}\lambda_p(N+T)$, we achieve the best rate from the above bound:

$$\frac{1}{T}\sum_{t=1}^{T}|m_t - \widehat{m}_t|^2 \lesssim_p \frac{1}{T} + \frac{N+T}{T\lambda_p}.$$

□

### B.12 Proof of Theorem 5(b)

*Proof.* i. (Slow rate) Note that (13) is equivalent to a constrained optimization problem:

$$\widehat{b} = \arg\min_b \left\|\widehat{\Sigma}^{-1/2}\bar{r} - \widehat{\Sigma}^{1/2}b\right\|^2, \quad \text{subject to } \|b\|_1 \leq \mu,$$

for some tuning parameter $\mu$. This implies that the vector of the true SDF loadings, $b$, satisfies that

$$\left\|\widehat{\Sigma}^{-1/2}\bar{r} - \widehat{\Sigma}^{1/2}\widehat{b}\right\|^2 \leq \left\|\widehat{\Sigma}^{-1/2}\bar{r} - \widehat{\Sigma}^{1/2}b\right\|^2 \quad \text{and} \quad \left\|\widehat{b}\right\|_1 \leq \mu, \text{ for some } \mu \geq s.$$

Equivalently, expanding the left- and right-hand sides of the above we have

$$\widehat{b}^{\intercal}\widehat{\Sigma}\widehat{b} - b^{\intercal}\widehat{\Sigma}b \leq 2(\widehat{b} - b)^{\intercal}\bar{r},$$

which leads to

$$(\widehat{b}-b)^\intercal\widehat{\Sigma}(\widehat{b}-b)\leq 2(\widehat{b}-b)^\intercal(\bar{r}-\widehat{\Sigma}b)\leq 2\left\|\widehat{b}-b\right\|_1\left\|\bar{r}-\widehat{\Sigma}b\right\|_\infty.$$

With a tuning parameter $\mu\asymp s$, we have

$$(\widehat{b}-b)^\intercal\widehat{\Sigma}(\widehat{b}-b)\lesssim s\left\|\bar{r}-\widehat{\Sigma}b\right\|_\infty. \tag{B.89}$$

With Lemma 15, we have

$$\left\|\widehat{\Sigma}^{1/2}(\widehat{b}-b)\right\|^2\lesssim_p s\sqrt{\frac{\log N}{T}}. \tag{B.90}$$

Therefore, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T\|\widehat{m}_t-\widetilde{m}_t\|^2=&\frac{1}{T}\sum_{t=1}^T\left\|\widehat{b}^\intercal(r_t-\bar{r})-b^\intercal(r_t-\mathrm{E}(r_t))\right\|^2\\ \leq&\frac{2}{T}\sum_{t=1}^T\left\|(\widehat{b}-b)^\intercal(r_t-\bar{r})\right\|^2+\frac{2}{T}\sum_{t=1}^T\|b^\intercal(\bar{r}-\mathrm{E}(r_t))\|^2\\ \leq&2\left\|\widehat{\Sigma}^{1/2}(\widehat{b}-b)\right\|^2+2\|b\|_1^2\|\bar{r}-\mathrm{E}(r_t)\|_\infty^2\\ \lesssim_p&s\sqrt{\frac{\log N}{T}}+s^2\frac{\log N}{T}.\end{aligned}$$

Since $s\asymp\mu\gtrsim\|b\|_1$, plugging in the optimal rate choice $s\asymp\|b\|_1$, we complete the proof.
ii. (Fast rate) Since $\widehat{b}$ is the optimal solution of the minimization problem, it implies that

$$b^\intercal\widehat{\Sigma}b-2b^\intercal\bar{r}+b^\intercal\widehat{\Sigma}b+\mu\|b\|_1\geq\widehat{b}^\intercal\widehat{\Sigma}\widehat{b}-2\widehat{b}^\intercal\bar{r}+\widehat{b}^\intercal\widehat{\Sigma}\widehat{b}+\mu\|\widehat{b}\|_1. \tag{B.91}$$

Rewrite (B.91) as

$$(\widehat{b}-b)^\intercal\widehat{\Sigma}(\widehat{b}-b)\leq 2(\widehat{b}-b)^\intercal(\bar{r}-\widehat{\Sigma}b)+\mu(\|b\|_1-\|\widehat{b}\|_1). \tag{B.92}$$

If $\mu\geq 4\left\|\bar{r}-\widehat{\Sigma}b\right\|_\infty$, (B.92) becomes

$$\begin{aligned}\left\|\widehat{\Sigma}^{1/2}(\widehat{b}-b)\right\|^2\leq&2\left\|\widehat{b}-b\right\|_1\left\|\bar{r}-\widehat{\Sigma}b\right\|_\infty+\mu(\|b\|_1-\|\widehat{b}\|_1)\\ \leq&\frac{1}{2}\mu\left\|\widehat{b}-b\right\|_1+\mu(\|b\|_1-\|\widehat{b}\|_1).\end{aligned} \tag{B.93}$$

Let $J$ denote the support of $\widehat{b}$, then (B.93) can be written as

$$\left\|\widehat{\Sigma}^{1/2}(\widehat{b}-b)\right\|^2\leq\frac{1}{2}\mu\left\|\widehat{b}_J-b_J\right\|_1+\frac{1}{2}\mu\left\|\widehat{b}_{J^c}\right\|_1+\mu\left\|\widehat{b}_J-b_J\right\|_1-\mu\left\|\widehat{b}_{J^c}\right\|_1$$

$$= \frac{3}{2}\mu \left\| \widehat{b}_J - b_J \right\|_1 - \frac{1}{2}\mu \left\| \widehat{b}_{J^c} \right\|_1 . \tag{B.94}$$

Define $b^* = \widehat{b} - b$, then (B.94) implies that $3\left\|b_J^*\right\|_1 \geq \left\|b_{J^c}^*\right\|_1$, and we have

$$\frac{b^{*\intercal}(\Sigma - \widehat{\Sigma})b^*}{\left\|b^*\right\|^2} \leq \left\|\Sigma - \widehat{\Sigma}\right\|_{MAX} \frac{\left\|b^*\right\|_1^2}{\left\|b^*\right\|^2} \lesssim_p \sqrt{\frac{\log N}{T}} \left( \frac{4\left\|b_J^*\right\|_1}{\left\|b_J^*\right\|} \right)^2 \lesssim_p |J| \sqrt{\frac{\log N}{T}}.$$

Consequently, with the assumption $|J|\sqrt{\log N/T} \to 0$ and $\lambda_{\min}(\Sigma) \gtrsim 1$, we have

$$\frac{b^{*\intercal}\widehat{\Sigma}b^*}{\left\|b^*\right\|^2} = \frac{b^{*\intercal}\Sigma b^*}{\left\|b^*\right\|^2} + \frac{b^{*\intercal}(\Sigma - \widehat{\Sigma})b^*}{\left\|b^*\right\|^2} \gtrsim_p 1.$$

Therefore, we have

$$\left\|\widehat{\Sigma}^{1/2}(\widehat{b} - b)\right\|^2 = b^{*\intercal}\widehat{\Sigma}b^* \gtrsim_p \left\|b^*\right\|^2 \geq \left\|b_J^*\right\|^2 \geq |J|^{-1}\left\|b_J^*\right\|_1^2 = |J|^{-1}\left\|\widehat{b}_J - b_J\right\|_1^2. \tag{B.95}$$

Plugging (B.95) into (B.94), we have

$$\left\|\widehat{\Sigma}^{1/2}(\widehat{b} - b)\right\|^2 \leq \frac{3}{2}\mu \left\|\widehat{b}_J - b_J\right\|_1 \lesssim_p \mu |J|^{1/2} \left\|\widehat{\Sigma}^{1/2}(\widehat{b} - b)\right\|.$$

Thus,

$$\left\|\widehat{\Sigma}^{1/2}(\widehat{b} - b)\right\|^2 \lesssim_p \mu^2 |J|. \tag{B.96}$$

Choosing $\mu = 4\left\|\bar{r} - \widehat{\Sigma}b\right\|_\infty$ and by Lemma 15, we obtain

$$\left\|\widehat{\Sigma}^{1/2}(\widehat{b} - b)\right\|^2 \lesssim_p |J| \frac{\log N}{T}. \tag{B.97}$$

Similar to the slow rate case, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} \left\|\widehat{m}_t - \widetilde{m}_t\right\|^2 =& \frac{1}{T}\sum_{t=1}^{T} \left\|\widehat{b}^\intercal(r_t - \bar{r}) - b^\intercal(r_t - \mathrm{E}(r_t))\right\|^2 \\
\leq& \frac{2}{T}\sum_{t=1}^{T} \left\|(\widehat{b} - b)^\intercal(r_t - \bar{r})\right\|^2 + \frac{2}{T}\sum_{t=1}^{T} \left\|b^\intercal(\bar{r} - \mathrm{E}(r_t))\right\|^2 \\
\leq& 2\left\|\widehat{\Sigma}^{1/2}(\widehat{b} - b)\right\|^2 + 2\left\|b^\intercal(\bar{r} - \mathrm{E}(r_t))\right\|^2 \\
\lesssim_p& \left\|b\right\|_0 \frac{\log N}{T}.
\end{aligned}$$

□

### B.13 Technical Lemmas and Their Proofs

Without loss of generality, we assume that $\Sigma_v = \mathbb{I}_p$ in the following lemmas. Also, except for Lemma 4, we assume that $\widehat{p} = \tilde{p}$ and $\widehat{I}_k = I_k$ for $k = 1, \ldots, \tilde{p}$, which hold with probability approaching one as we will show in Lemma 4.

**Lemma 1.** *Under Assumptions A.1-A.7, for any $I \subset [N]$, we have the following results:*

*(i)* $\left\|T^{-1}\bar{V}\bar{V}^\intercal - \mathbb{I}_p\right\| \lesssim_p T^{-1/2}$.

*(ii)* $\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal \bar{U}_{[I]}\right\| \lesssim_p T^{1/2}$.

*(iii)* $\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal \bar{U}_{[I]}\bar{V}^\intercal\right\| \lesssim_p T^{1/2}$, $\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal \bar{U}_{[I]}\bar{Z}^\intercal\right\| \lesssim_p T^{1/2}$.

*(iv)* $\left\|\bar{U}\right\|_{\text{MAX}} \lesssim_p (\log NT)^{1/2}$, $\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2}T^{1/2}$, $\left\|\bar{U}\bar{Z}^\intercal\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2}T^{1/2}$.

*(v)* $\left\|\bar{U}_{[I]}\right\| \lesssim_p |I|^{1/2} + T^{1/2}$, $\left\|\bar{U}_{[I]}\bar{V}^\intercal\right\| \lesssim_p |I|^{1/2}T^{1/2}$, $\left\|\bar{U}_{[I]}\bar{Z}^\intercal\right\| \lesssim_p |I|^{1/2}T^{1/2}$.

*(vi)* $\left\|\bar{V}\right\| \lesssim_p T^{1/2}$, $\left\|\bar{Z}\right\| \lesssim_p T^{1/2}$, $\left\|\bar{V}\bar{Z}^\intercal\right\| \lesssim_p T^{1/2}$, $\left\|\bar{V}\bar{Z}^\intercal - VZ^\intercal\right\| \lesssim_p 1$

*Proof.* (i) Using Assumption A.1, we have

$$\left\|\frac{\bar{V}\bar{V}^\intercal}{T} - \mathbb{I}_p\right\| \leq \left\|\frac{VV^\intercal}{T} - \mathbb{I}_p\right\| + \left\|\frac{V\iota_T\iota_T^\intercal V^\intercal}{T^2}\right\| = \left\|\frac{VV^\intercal}{T} - \mathbb{I}_p\right\| + \|\bar{v}\|^2 \lesssim_p T^{-1/2}.$$

(ii) Using Assumption A.5, we have

$$\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal \bar{U}_{[I]}\right\| \leq \left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal U_{[I]}\right\| + T^{-1}\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal U_{[I]}\iota_T\iota_T^\intercal\right\| \lesssim_p T^{1/2}.$$

(iii) By Assumptions A.1, A.5 and A.6, we have

$$\begin{aligned}\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal \bar{U}_{[I]}\bar{V}^\intercal\right\| &\leq \left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal U_{[I]}V^\intercal\right\| + T^{-1}\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal U_{[I]}\iota_T\iota_T^\intercal V\right\| \\ &\leq \left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal U_{[I]}V^\intercal\right\| + \left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal U_{[I]}\iota_T\right\| \|\bar{v}\| \lesssim_p T^{1/2}.\end{aligned}$$

Replacing $\bar{V}$ by $\bar{Z}$ in the above proof, with Assumptions A.2, A.5 and A.7, we also have

$$\left\|\left(\beta_{[I]}^\intercal \beta_{[I]}\right)^{-\frac{1}{2}} \beta_{[I]}^\intercal \bar{U}_{[I]}\bar{Z}^\intercal\right\| \lesssim_p T^{1/2}.$$

(iv) Using Assumption A.4, we have

$$\left\|\bar{U}\right\|_{\text{MAX}} \leq \|U\|_{\text{MAX}} + T^{-1}\left\|U\iota_T\iota_T^\intercal\right\|_{\text{MAX}} \leq \|U\|_{\text{MAX}} + \|\bar{u}\|_{\text{MAX}}\|\iota_T\| \lesssim_p (\log N)^{1/2} + (\log T)^{1/2}.$$

Using Assumptions A.1, A.4, A.6, we have

$$\left\|\bar{U}\bar{V}^{\intercal}\right\|_{\text{MAX}} \leq \|UV^{\intercal}\|_{\text{MAX}} + T^{-1}\left\|U\iota_T\iota_T^{\intercal}V^{\intercal}\right\|_{\text{MAX}} \leq \|UV^{\intercal}\|_{\text{MAX}} + T\left\|\bar{u}\right\|_{\text{MAX}}\|\bar{v}\| \lesssim_p (\log N)^{1/2}T^{1/2}.$$

Replacing $\bar{V}$ by $\bar{Z}$ in the above proof, with Assumptions A.2, A.4 and A.7, we also have

$$\left\|\bar{U}\bar{Z}^{\intercal}\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2}T^{1/2}.$$

(v) Using Assumption A.4 , we have

$$\left\|\bar{U}_{[I]}\right\| \leq \left\|U_{[I]}\right\| + T^{-1}\left\|U_{[I]}\iota_T\iota_T^{\intercal}\right\| \leq \left\|U_{[I]}\right\| + \left\|\bar{u}_{[I]}\right\|\|\iota_T\| \lesssim_p |I|^{1/2} + T^{1/2}.$$

Using Assumptions A.1, A.4, A.6, we have

$$\left\|\bar{U}_{[I]}\bar{V}^{\intercal}\right\| \leq \left\|U_{[I]}V^{\intercal}\right\| + T^{-1}\left\|U_{[I]}\iota_T\iota_T^{\intercal}V^{\intercal}\right\| \leq \left\|U_{[I]}V^{\intercal}\right\| + T\left\|\bar{u}_{[I]}\right\|\|\bar{v}\| \lesssim_p |I|^{1/2}T^{1/2}.$$

Replacing $\bar{V}$ by $\bar{Z}$ in the above proof, with Assumptions A.2, A.4 and A.7, we also have

$$\left\|\bar{U}_{[I]}\bar{Z}^{\intercal}\right\| \lesssim_p |I|^{1/2}T^{1/2}.$$

(vi) Using Assumption A.1, we have

$$\left\|\bar{V}\right\| \leq \|V\| + T^{-1}\left\|V\iota_T\iota_T^{\intercal}\right\| \leq \|V\| + \|\bar{v}\|\|\iota_T\| \lesssim_p T^{1/2}.$$

Using Assumption A.2, we have

$$\left\|\bar{Z}\right\| \leq \|Z\| + T^{-1}\left\|Z\iota_T\iota_T^{\intercal}\right\| \leq \|Z\| + \|\bar{z}\|\|\iota_T\| \lesssim_p T^{1/2}.$$

Using Assumptions A.1 and A.2, we have

$$\left\|\bar{V}\bar{Z}^{\intercal}\right\| \leq \|VZ\| + T^{-1}\left\|V\iota_T\iota_T^{\intercal}Z\right\| \leq \|V\| + T\|\bar{v}\|\|\bar{z}\| \lesssim_p T^{1/2},$$

and

$$\left\|\bar{V}\bar{Z}^{\intercal} - VZ^{\intercal}\right\| = \left\|T^{-1}V\iota_T\iota_T^{\intercal}Z\right\| = T\|\bar{v}\|\|\bar{z}\| \lesssim_p 1.$$

□

**Lemma 2.** *The singular vectors $\xi_{(k)}$s we obtain from Algorithm 5 satisfy $\xi_{(j)}^{\intercal}\xi_{(k)} = \delta_{jk}$ for $j, k \leq \widehat{p}$.*

*Proof.* If $j = k$, this result holds from the definition of $\xi_{(k)}$. If $j < k$, recall that $\widetilde{R}_{(k)}$ is defined in

(B.39) and $\xi_{(k)}$ is the first right singular vector of $\widetilde{R}_{(k)}$, we have

$$\widetilde{R}_{(k)} = \bar{R}_{[I_k]} \prod_{i<k} \left( \mathbb{I}_T - \xi_{(i)} \xi_{(i)}^\intercal \right) \quad \text{and} \quad \xi_{(k)} = \arg\max_{\alpha} \frac{\left\| \widetilde{R}_{(k)} \alpha \right\|}{\|\alpha\|}.$$

If $\xi_{(k)}^\intercal \xi_{(j)} = c_0 \neq 0$ for some $j < k$, then

$$\left\| \widetilde{R}_{(k)} (\xi_{(k)} - c_0 \xi_{(j)}) \right\| = \left\| \widetilde{R}_{(k)} \xi_{(k)} - c_0 \widetilde{R}_{(k)} \xi_{(j)} \right\| = \left\| \widetilde{R}_{(k)} \xi_{(k)} \right\|, \tag{B.98}$$

since the definition of $\widetilde{R}_{(k)}$ implies that $\widetilde{R}_{(k)} \xi_{(j)} = 0$ for $j < k$.

On the other hand, since $\xi_{(k)}^\intercal \xi_{(j)} = c_0 \neq 0$, we have $(\xi_{(k)} - c_0 \xi_{(j)})^\intercal \xi_{(j)} = 0$, and consequently,

$$\|\xi_{(k)}\|^2 = \|\xi_{(k)} - c_0 \xi_{(j)}\|^2 + \|c_0 \xi_{(j)}\|^2 > \|\xi_{(k)} - c_0 \xi_{(j)}\|^2. \tag{B.99}$$

Apparently, if $\left\| \widetilde{R}_{(k)} \right\| = 0$, the process will stop so we have $\left\| \widetilde{R}_{(k)} \right\| > 0$ for $k \leq \widehat{p}$. Together with (B.98) and (B.99), we have

$$\left\| \widetilde{R}_{(k)} \right\| = \frac{\left\| \widetilde{R}_{(k)} \xi_{(k)} \right\|}{\|\xi_{(k)}\|} \leq \frac{\left\| \widetilde{R}_{(k)} (\xi_{(k)} - c_0 \xi_{(j)}) \right\|}{\|\xi_{(k)} - c_0 \xi_{(j)}\|},$$

which contradicts with the definition of $\xi_{(k)}$. Therefore, $\xi_{(k)}^\intercal \xi_{(j)} = 0$ for $j < k$. This completes the proof. □

**Lemma 3.** *Under Assumption A.3, if $c \to 0$, $qN/N_0 \to 0$ then $b_k$, $\beta_{(k)}$ and $\tilde{p}$ defined in Section A satisfy*

*(i)* $\langle b_j, b_k \rangle = \delta_{jk}$ *for* $j \leq k \leq \tilde{p}$.

*(ii)* $\left\| \beta_{(k)} \right\| \asymp q^{1/2} N^{1/2}$.

*(iii)* $\tilde{p} \leq p$.

*(iv)* $\tilde{p} = p$, *if we further have* $\lambda_p(\eta^\intercal \eta) \gtrsim 1$.

*Proof.* (i) Recall that $b_k$ is the first right singular vector of $\beta_{(k)}$ and $\beta_{(k)} = \beta_{[I_k]} \prod_{j<k} \mathbb{M}_{b_j}$. Using the same arguments as in the proof of Lemma 2, we have $\langle b_j, b_k \rangle = \delta_{jk}$ for $j \leq k \leq \tilde{p}$.

(ii) The selection rule at $k$th step implies that

$$\frac{1}{|I_k|} \sum_{i \in I_k} \left\| \beta_{[i]} \prod_{j<k} \mathbb{M}_{b_j} \eta^\intercal \right\|_{\text{MAX}}^2 \geq \frac{1}{N_0} \sum_{i \in I_0} \left\| \beta_{[i]} \prod_{j<k} \mathbb{M}_{b_j} \eta^\intercal \right\|_{\text{MAX}}^2. \tag{B.100}$$

For any matrix $A \in \mathbb{R}^{N\times d}$ and set $I \subset [N]$, we have

$$\sum_{i\in I} \left\|A_{[i]}\right\|_{\mathrm{MAX}}^2 \le \|A\|_{\mathrm{F}}^2 \le d \sum_{i\in I} \left\|A_{[i]}\right\|_{\mathrm{MAX}}^2,$$

and

$$\|A\|^2 \le \|A\|_{\mathrm{F}}^2 \le d\,\|A\|^2,$$

we thereby have

$$\|A\|^2 \asymp \sum_{i\in I} \left\|A_{[i]}\right\|_{\mathrm{MAX}}^2. \tag{B.101}$$

Using this result, (B.100) becomes

$$\frac{1}{|I_k|}\left\|\beta_{[I_k]}\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal\right\|^2 \gtrsim \frac{1}{N_0}\left\|\beta_{[I_0]}\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal\right\|^2.$$

Then, we have

$$\frac{1}{\sqrt{|I_k|}}\left\|\beta_{(k)}\right\|\left\|\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal\right\| \ge \frac{1}{\sqrt{|I_k|}}\left\|\beta_{[I_k]}\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal\right\| \gtrsim \frac{1}{\sqrt{N_0}}\left\|\beta_{[I_0]}\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal\right\| \ge \frac{1}{\sqrt{N_0}}\sigma_p\left(\beta_{[I_0]}\right)\left\|\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal\right\|, \tag{B.102}$$

where we use $\beta_{[I_k]}\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal = \beta_{[I_k]}(\prod_{j<k}\mathbb{M}_{b_j})^2\eta^\intercal = \beta_{(k)}\prod_{j<k}\mathbb{M}_{b_j}\eta^\intercal$ in the first inequality. With $\sigma_p(\beta_{[I_0]}) \gtrsim \sqrt{N_0}$ from Assumption A.3, (B.102) leads to $\left\|\beta_{(k)}\right\| \gtrsim |I_k|^{1/2}$. In addition, $\|\beta\|_{\mathrm{MAX}} \lesssim 1$ from Assumption A.3 leads to $\left\|\beta_{(k)}\right\| \lesssim |I_k|^{1/2}$. Therefore, we have $\left\|\beta_{(k)}\right\| \asymp |I_k|^{1/2} \asymp q^{1/2}N^{1/2}$.

(iii) From (i), we have shown that $b_k$'s are pairwise orthogonal for $k \le \tilde{p}$. It is impossible to have more than $p$ pairwise orthogonal $p$ dimensional vectors. Thus, $\tilde{p} \le p$.

(iv) Recall that $\tilde{p}$ is defined in Section A. Since the procedure in its definition stops at $\tilde{p}+1$, we have at most $qN-1$ rows of $\beta$ satisfying $\left\|\beta_{[i]}\prod_{j\le\tilde{p}}\mathbb{M}_{b_j}\eta^\intercal\right\|_{\mathrm{MAX}} \ge c$, which implies

$$\left\|\beta_{[I_0]}\prod_{j\le\tilde{p}}\mathbb{M}_{b_j}\eta^\intercal\right\|^2 \lesssim qN + (N_0 - qN)c^2 = o(N_0),$$

where we use (B.101) and the assumptions $c \to 0$, $qN/N_0 \to 0$. With $\sigma_p(\beta_{[I_0]}) \gtrsim \sqrt{N_0}$ from

Assumption A.3, we have

$$\left\| \eta \prod_{j \le \tilde{p}} \mathbb{M}_{b_j} \right\| \le \sigma_p(\beta_{[I_0]})^{-1} \left\| \beta_{[I_0]} \prod_{j \le \tilde{p}} \mathbb{M}_{b_j} \eta^\intercal \right\| = o(1). \tag{B.103}$$

If $\tilde{p} \le p-1$, using (i), we have

$$\eta \prod_{j \le \tilde{p}} \mathbb{M}_{b_j} = \eta - \eta \sum_{j \le \tilde{p}} b_j b_j^\intercal,$$

which implies that

$$\sigma_p(\eta) \le \sigma_1 \left( \eta \prod_{j \le \tilde{p}} \mathbb{M}_{b_j} \right) + \sigma_p \left( \eta \sum_{j \le \tilde{p}} b_j b_j^\intercal \right). \tag{B.104}$$

Since

$$\text{Rank} \left( \eta \sum_{j \le \tilde{p}} b_j b_j^\intercal \right) \le \tilde{p} \le p-1, \tag{B.105}$$

we have $\sigma_p \left( \eta \sum_{j \le \tilde{p}} b_j b_j^\intercal \right) \le 0$ and thus (B.104) and (B.103) lead to $\sigma_p(\eta) \lesssim \sigma_1 \left( \eta \prod_{j \le \tilde{p}} \mathbb{M}_{b_j} \right) \longrightarrow 0$. This contradicts with the assumption that $\lambda_p(\eta^\intercal \eta) \gtrsim 1$. Therefore, we have $\tilde{p} \ge p$. Together with the result in (iii), we have $\tilde{p} = p$. □

**Lemma 4.** *Suppose Assumptions A.1-A.8 hold. If* $c^{-1} \log(NT)^{1/2} \left( q^{-1/2} N^{-1/2} + T^{-1/2} \right) \to 0$ *and* $c \to 0$, *then for* $k \le \tilde{p}$ *and for* $I_k$, $\tilde{p}$ *and* $\beta_{(k)}$ *defined in Section A, we have*

*(i)* $\mathrm{P}(\widehat{I}_k = I_k) \to 1$.

*(ii)* $\left\| \widetilde{R}_{(k)} - \beta_{(k)} \bar{V} \right\| \lesssim_p q^{1/2} N^{1/2} + T^{1/2}$.

*(iii)* $|\widehat{\lambda}_{(k)}^{1/2} / \left\| \beta_{(k)} \right\| - 1| \lesssim_p q^{-1/2} N^{-1/2} + T^{-1/2}$.

*(iv)* $\left\| \mathbb{P}_{\widehat{V}_{(k)}^\intercal} - T^{-1} \bar{V}^\intercal \mathbb{P}_{b_k} \bar{V} \right\| \lesssim_p q^{-1/2} N^{-1/2} + T^{-1/2}$.

*(v)* $\mathrm{P}(\widehat{p} = \tilde{p}) \to 1$.

*Proof.* We prove (i)-(iv) by induction. First, we show that (i)-(iv) hold when $k = 1$:

(i) Recall that $\widehat{I}_1$ is selected based on $T^{-1} \bar{R} \bar{G}^\intercal$ and $I_1$ based on $\beta \eta^\intercal$. With simple algebra, we have

$$T^{-1} \bar{R} \bar{G}^\intercal - \beta \eta^\intercal = \beta \left( T^{-1} \bar{V} \bar{V}^\intercal - \mathbb{I}_p \right) \eta^\intercal + T^{-1} \bar{U} \bar{V}^\intercal \eta^\intercal + T^{-1} \beta \bar{V} \bar{Z}^\intercal + T^{-1} \bar{U} \bar{Z}^\intercal.$$

With Assumptions A.1, A.2, A.3, A.6 A.7, we have

$$\begin{aligned}\left\|T^{-1}\bar{R}\bar{G}^{\intercal}-\beta\eta^{\intercal}\right\|_{\text{MAX}} \lesssim & \|\beta\|_{\text{MAX}}\left\|T^{-1}\bar{V}\bar{V}^{\intercal}-\mathbb{I}_p\right\|\|\eta\|+T^{-1}\left\|\bar{U}\bar{V}^{\intercal}\right\|_{\text{MAX}}\|\eta\| \\ & +T^{-1}\|\beta\|_{\text{MAX}}\left\|\bar{V}\bar{Z}^{\intercal}\right\|+T^{-1}\left\|\bar{U}\bar{Z}^{\intercal}\right\|_{\text{MAX}} \lesssim_p (\log N)^{1/2}T^{-1/2}.\end{aligned}$$

From Assumption A.8, we have $c_{qN}^{(1)}-c_{qN+1}^{(1)} \gtrsim c_{qN}^{(1)}$ and the the definition of $\tilde{p}$ implies that $c_{qN}^{(k)} \geq c$ for $k \leq \tilde{p}$. Thus, we have $c_{qN}^{(1)}-c_{qN+1}^{(1)} \gtrsim c$. Define the events

$$\begin{aligned}A_1 :=& \left\{ \left\|T^{-1}\bar{R}_{[i]}\bar{G}^{\intercal}\right\|_{\text{MAX}} > (c_{qN}^{(1)}+c_{qN+1}^{(1)})/2 \text{ for all } i \in I_1\right\}, \\ A_2 :=& \left\{ \left\|T^{-1}\bar{R}_{[i]}\bar{G}^{\intercal}\right\|_{\text{MAX}} < (c_{qN}^{(1)}+c_{qN+1}^{(1)})/2 \text{ for all } i \in I_1^c\right\}, \\ A_3 :=& \left\{ \left\|T^{-1}\bar{R}_{[i]}\bar{G}^{\intercal}-\beta_{[i]}\eta^{\intercal}\right\|_{\text{MAX}} \geq (c_{qN}^{(1)}-c_{qN+1}^{(1)})/2 \text{ for some } i \in [N]\right\}. \end{aligned} \tag{B.106}$$

It is easy to observe that $\{\widehat{I}_1 = I_1\} \supset A_1 \cap A_2$. In addition, from the definition of $I_1$, we have $\left\|\beta_{[i]}\eta^{\intercal}\right\|_{\text{MAX}} \geq c_{qN}^{(1)}$ for all $i \in I_1$ and $\left\|\beta_{[i]}\eta^{\intercal}\right\|_{\text{MAX}} \leq c_{qN+1}^{(1)}$ for all $i \in I_1^c$. Therefore, if $A_1^c$ occurs, we have

$$\left\|T^{-1}\bar{R}_{[i]}\bar{G}^{\intercal}-\beta_{[i]}\eta^{\intercal}\right\|_{\text{MAX}} \geq (c_{qN}^{(1)}-c_{qN+1}^{(1)})/2,$$

for some $i \in I_1$, which implies $A_1^c \subset A_3$. Similarly, we have $A_2^c \subset A_3$. Using $\{\widehat{I}_1 = I_1\} \supset A_1 \cap A_2$ and $A_1^c \cup A_2^c \subset A_3$, we have

$$\mathrm{P}(\widehat{I}_1 = I_1) \geq \mathrm{P}(A_1 \cap A_2) = 1-\mathrm{P}(A_1^c \cup A_2^c) \geq 1-\mathrm{P}(A_3). \tag{B.107}$$

Using $c^{-1}(\log N)^{1/2}T^{-1/2} \to 0$ and $c_{qN}^{(1)}-c_{qN+1}^{(1)} \gtrsim c$, we have $\mathrm{P}(A_3) \to 0$ and consequently, $\mathrm{P}(\widehat{I}_1 = I_1) \to 1$.

(ii) Since $\widehat{I}_1 = I_1$ with high probability, we impose $\widehat{I}_1 = I_1$ below. Then, we have $\widetilde{R}_{(1)} = \bar{R}_{[I_1]}$ and Assumption A.13 gives $\left\|\widetilde{R}_{(1)}-\beta_{(1)}\bar{V}\right\| = \left\|\bar{U}_{[I_1]}\right\| \lesssim_p q^{1/2}N^{1/2}+T^{1/2}$.

(iii) From Lemma 10, we have $\sigma_j(\beta_{(1)}\bar{V})/\sigma_j(\beta_1) = T^{1/2}+O_p(1)$. The result in (ii) implies that

$$\left|\left\|\widetilde{R}_{(1)}\right\|-\left\|\beta_{(1)}\bar{V}\right\|\right| \leq \left\|\widetilde{R}_{(1)}-\beta_{(1)}\bar{V}\right\| \lesssim_p q^{1/2}N^{1/2}+T^{1/2}.$$

Together with $\left\|\beta_{(1)}\right\| \asymp qN$ from Lemma 3, we have

$$\left|\frac{\widehat{\lambda}_{(1)}^{1/2}}{\left\|\beta_{(k)}\right\|}-1\right| = \left|\frac{\left\|\widetilde{R}_{(1)}\right\|}{T^{1/2}\left\|\beta_{(1)}\right\|}-1\right| \leq \frac{\left|\left\|\widetilde{R}_{(1)}\right\|-\left\|\beta_{(1)}\bar{V}\right\|\right|}{T^{1/2}\left\|\beta_{(1)}\right\|}+\frac{\left|\left\|\beta_{(1)}\bar{V}\right\|-T^{1/2}\left\|\beta_{(1)}\right\|\right|}{T^{1/2}\left\|\beta_{(1)}\right\|} \lesssim_p q^{-1/2}N^{-1/2}+T^{-1/2}.$$

(iv) Let $\tilde{\xi}_{(1)} \in \mathbb{R}^{T\times 1}$ denote the first right singular vector of $\beta_{(1)}\bar{V}$. From Lemma 10, we have

$$\left\|\mathbb{P}_{\tilde{\xi}_{(1)}}-T^{-1}\bar{V}^{\intercal}\mathbb{P}_{b_k}\bar{V}\right\| \lesssim_p T^{-1/2} \tag{B.108}$$

and $\sigma_j(\beta_{(1)}\bar{V})/\sigma_j(\beta_{(1)}) = T^{1/2} + O_p(1)$ for $j \le p$, which leads to

$$\sigma_1(\beta_{(1)}\bar{V}) - \sigma_2(\beta_{(1)}\bar{V}) = T^{1/2}(\sigma_1(\beta_{(1)}) - \sigma_2(\beta_{(1)})) + O_p(\sigma_1(\beta_{(1)})) \asymp_p T^{1/2}\sigma_1(\beta_{(1)}), \tag{B.109}$$

where we use the assumption that $\sigma_2(\beta_{(1)}) \le (1+\delta)^{-1}\sigma_1(\beta_{(1)})$ in the last equation.

Using $\left\|\widetilde{R}_{(1)} - \beta_{(1)}\bar{V}\right\| \lesssim_p q^{1/2}N^{1/2} + T^{1/2}$ as proved in (ii), (B.109), Lemma 3 and Wedin's sin-theta theorem for singular vectors in Wedin (1972), we have

$$\left\|\mathbb{P}_{\widehat{V}_{(k)}^\intercal} - \mathbb{P}_{\tilde{\xi}_{(1)}}\right\| \lesssim_p \frac{q^{1/2}N^{1/2} + T^{1/2}}{\sigma_1(\beta_{(1)}\bar{V}) - \sigma_2(\beta_{(1)}\bar{V})} \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}, \tag{B.110}$$

In light of (B.108) and (B.110), we have that (iv) holds for $k = 1$.

So far, we have proved that (i)-(iv) hold for $k = 1$. Now, assuming that (i)-(iv) hold for $j \le k-1$, we will show that (i)-(iv) continue to hold for $j = k$.

(i) Again, we show the difference between the sample covariances and their population counterparts introduced in the SPCA procedure are tiny. At the $k$th step, the difference can be written as

$$\begin{aligned}
&\left\|\beta\prod_{j=1}^{k-1}\mathbb{M}_{b_j}\eta^\intercal - T^{-1}(\beta\bar{V}+\bar{U})\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}(\eta\bar{V}+\bar{Z})^\intercal\right\|_{\text{MAX}} \\
\le &\left\|\beta\prod_{j=1}^{k-1}\mathbb{M}_{b_j}\eta^\intercal - T^{-1}\beta\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{V}^\intercal\eta^\intercal\right\|_{\text{MAX}} + T^{-1}\left\|\beta\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{Z}^\intercal\right\|_{\text{MAX}} \\
&+ T^{-1}\left\|\bar{U}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{V}^\intercal\eta^\intercal\right\|_{\text{MAX}} + T^{-1}\left\|\bar{U}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{Z}^\intercal\right\|_{\text{MAX}}
\end{aligned} \tag{B.111}$$

Since (iv) holds for $j \le k-1$, we have

$$\left\|\sum_{j=1}^{k-1}\mathbb{P}_{\widehat{V}_{(j)}^\intercal} - T^{-1}\bar{V}^\intercal\sum_{j=1}^{k-1}\mathbb{P}_{b_j}\bar{V}\right\| = \left\|\sum_{j=1}^{k-1}\left(\mathbb{P}_{\widehat{V}_{(j)}^\intercal} - T^{-1}\bar{V}^\intercal\mathbb{P}_{b_j}\bar{V}\right)\right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}. \tag{B.112}$$

Using Lemma 2 and Lemma 3(i), we have

$$\prod_{j=1}^{k-1}\mathbb{M}_{b_j} = \mathbb{I}_p - \sum_{j=1}^{k-1}\mathbb{P}_{b_j}, \quad \text{and} \quad \prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}} = \mathbb{I}_T - \sum_{j=1}^{k-1}\mathbb{P}_{\widehat{V}_{(j)}}.$$

Using the above equations, (B.112), and $\left\|T^{-1}\bar{V}\bar{V}^\intercal - \mathbb{I}_p\right\| \lesssim_p T^{-1/2}$, we have

$$T^{-1/2}\left\|\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal} - \prod_{j=1}^{k-1}\mathbb{M}_{b_j}\bar{V}\right\| = T^{-1/2}\left\|\bar{V}\sum_{j=1}^{k-1}\mathbb{P}_{\widehat{V}_{(j)}^\intercal} - \sum_{j=1}^{k-1}\mathbb{P}_{b_j}\bar{V}\right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}. \tag{B.113}$$

Similarly, right multiplying $\bar{V}^\intercal$ to the term inside the $\|\cdot\|$ of (B.113), we have

$$\left\|T^{-1}\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{V}^\intercal - \prod_{j=1}^{k-1}\mathbb{M}_{b_j}\right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}. \tag{B.114}$$

Then, we analyze these four terms in (B.111) one by one. For the first term, using (B.114) and Assumption A.3, we have

$$\begin{aligned}\left\|\beta\prod_{j=1}^{k-1}\mathbb{M}_{b_j}\eta^\intercal - T^{-1}\beta\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{V}^\intercal\eta^\intercal\right\|_{\text{MAX}} &\lesssim \|\beta\|_{\text{MAX}}\left\|\prod_{j=1}^{k-1}\mathbb{M}_{b_j} - T^{-1}\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{V}^\intercal\right\|\|\eta\| \\ &\lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}.\end{aligned}$$

For the second term, using (B.113), Lemma 1 and Assumptions A.3 and A.2, we have

$$\begin{aligned}T^{-1}\left\|\beta\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{Z}^\intercal\right\|_{\text{MAX}} &\lesssim T^{-1}\|\beta\|_{\text{MAX}}\left\|\prod_{j=1}^{k-1}\mathbb{M}_{b_j}\right\|\left\|\bar{V}\bar{Z}^\intercal\right\| + T^{-1}\|\beta\|_{\text{MAX}}\left\|\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal} - \prod_{j=1}^{k-1}\mathbb{M}_{b_j}\bar{V}\right\|\|\bar{Z}\| \\ &\lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}.\end{aligned}$$

For the third term, using (B.113) and Lemma 1, we have

$$\begin{aligned}T^{-1}\left\|\bar{U}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{V}^\intercal\eta^\intercal\right\|_{\text{MAX}} &\lesssim T^{-1}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}}\left\|\prod_{j=1}^{k-1}\mathbb{M}_{b_j}\right\|\|\eta\| \\ &\quad + T^{-1}\|\bar{U}\|_{\text{MAX}}T^{1/2}\left\|\bar{V}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal} - \prod_{j=1}^{k-1}\mathbb{M}_{b_j}\bar{V}\right\|\|\eta\| \\ &\lesssim_p (\log NT)^{1/2}\left(q^{-1/2}N^{-1/2} + T^{-1/2}\right).\end{aligned}$$

For the forth term, using (B.112) and Lemma 1, we have

$$T^{-1}\left\|\bar{U}\prod_{j=1}^{k-1}\mathbb{M}_{\widehat{V}_{(j)}^\intercal}\bar{Z}^\intercal\right\|_{\text{MAX}} \lesssim T^{-1}\left\|\bar{U}\bar{Z}^\intercal\right\|_{\text{MAX}} + T^{-2}\left\|\bar{U}\bar{V}^\intercal\right\|_{\text{MAX}}\left\|\sum_{j=1}^{k-1}\mathbb{P}_{b_j}\right\|\left\|\bar{V}\bar{Z}^\intercal\right\|$$

$$+ T^{-1/2} \left\|\bar{U}\right\|_{\text{MAX}} \left\| T^{-1} \bar{V}^\intercal \sum_{j=1}^{k-1} \mathbb{P}_{b_j} \bar{V} - \sum_{j=1}^{k-1} \mathbb{P}_{\widehat{V}_{(j)}^\intercal} \right\| \left\|\bar{Z}\right\|$$
$$\lesssim_p (\log NT)^{1/2} \left( q^{-1/2} N^{-1/2} + T^{-1/2} \right).$$

Hence, we have

$$\left\| T^{-1} \bar{R} \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal} \bar{G}^\intercal - \beta \prod_{j=1}^{k-1} \mathbb{M}_{b_j} \eta^\intercal \right\|_{\text{MAX}} \lesssim_p (\log NT)^{1/2} \left( q^{-1/2} N^{-1/2} + T^{-1/2} \right). \tag{B.115}$$

As in the case of $k = 1$, from Assumption A.8, we have $c_{qN}^{(k)} - c_{qN+1}^{(k)} \gtrsim c_{qN}^{(k)}$. In addition, since the stopping rule for the procedure in Section A is $c_{qN}^{(\tilde{p}+1)} < c$, we have $c_{qN}^{(k)} \geq c$ for $k \leq \tilde{p}$. With the assumption that

$$c^{-1} (\log NT)^{1/2} \left( q^{-1/2} N^{-1/2} + T^{-1/2} \right) \to 0,$$

we can reuse the arguments for (B.106) and (B.107) in the case of $k = 1$ and obtain $\mathrm{P}(\widehat{I}_k = I_k) \to 1$.

(ii) We impose $\widehat{I}_k = I_k$ below. Then, we have $\widetilde{R}_{(k)} = \bar{R}_{[I_k]} \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal}$ and thus

$$\widetilde{R}_{(k)} - \beta_{(k)} \bar{V} = \bar{R}_{[I_k]} \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal} - \beta_{(k)} \bar{V} = \bar{\beta}_{[I_k]} \left( \bar{V} \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal} - \prod_{j=1}^{k-1} \mathbb{M}_{b_j} \bar{V} \right) + \bar{U}_{[I_k]} \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal}.$$

Hence, using Assumptions A.3, Lemma 1, and (B.113), we have

$$\left\| \widetilde{R}_{(k)} - \beta_{(k)} \bar{V} \right\| \leq \left\| \beta_{[I_k]} \right\| \left\| \bar{V} \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal} - \prod_{j=1}^{k-1} \mathbb{M}_{b_j} \bar{V} \right\| + \left\| \bar{U}_{[I_k]} \right\| \left\| \prod_{j=1}^{k-1} \mathbb{M}_{\widehat{V}_{(j)}^\intercal} \right\| \lesssim_p q^{1/2} N^{1/2} + T^{1/2}.$$

(iii) The proof of (iii) is analogous to the case $k = 1$. Rewrite the proof of the case $k = 1$ by replacing $\widetilde{R}_{(1)}$ and $\beta_{(1)}$ by $\widetilde{R}_{(k)}$ and $\beta_{(k)}$. We have $\left| \widehat{\lambda}_{(k)}^{1/2} / \left\| \beta_{(k)} \right\| - 1 \right| \lesssim_p q^{-1/2} N^{-1/2} + T^{-1/2}$.

(iv) The proof of (iv) is analogous to the case $k = 1$. Let $\tilde{\xi}_{(k)}$ denote the first right singular vector of $\beta_{(k)} \bar{V}$, then we have $\left\| \mathbb{M}_{\tilde{\xi}_{(k)}} - T^{-1} \bar{V}^\intercal \mathbb{M}_{b_k} \bar{V} \right\| \lesssim_p T^{-1/2}$ from Lemma 10. Since we have $\left\| \widetilde{R}_{(k)} - \beta_{(k)} \bar{V} \right\| \lesssim_p q^{-1/2} N^{-1/2} + T^{-1/2}$ from (ii), using the same proof as in the case $k = 1$, we have

$$\left\| \mathbb{M}_{\widehat{V}_{(k)}^\intercal} - \mathbb{M}_{\tilde{\xi}_{(k)}} \right\| \lesssim_p q^{-1/2} N^{-1/2} + T^{-1/2},$$

by Wedin's sin-theta theorem. Combining these two inequalities completes the proof.

To sum up, by induction, we have shown that (i)-(iv) hold for $k \leq \tilde{p}$.

(v) Recall that $\tilde{p}$ is determined by $\beta_{[i]} \prod_{j<k} \mathbb{M}_{b_j} \eta^\intercal$ whereas $\widehat{p}$ is determined by $T^{-1} \bar{R}_{[i]} \prod_{j<k} \mathbb{M}_{\widehat{V}_{(j)}^\intercal} \bar{G}^\intercal$. Since (iv) holds for $j \leq \tilde{p}$ as shown above, using the same proof for (B.115),

we have

$$\left\| T^{-1}\bar{R}\prod_{j=1}^{\tilde{p}}\mathbb{M}_{\widehat{V}_{(j)}^{\intercal}}\bar{G}^{\intercal} - \beta\prod_{j=1}^{\tilde{p}}\mathbb{M}_{b_j}\eta^{\intercal}\right\|_{\text{MAX}} \lesssim_p (\log NT)^{1/2}\left(q^{-1/2}N^{-1/2}+T^{-1/2}\right). \tag{B.116}$$

The assumption $c_{qN}^{(\tilde{p}+1)} \leq (1+\delta)^{-1}c$ in Assumption A.8 implies that $c - c_{qN}^{(\tilde{p}+1)} \asymp c$. Together with

$$c^{-1}(\log NT)^{1/2}\left(q^{-1/2}N^{-1/2}+T^{-1/2}\right) \to 0,$$

we can reuse the arguments for (B.106) and (B.107) with events

$$\begin{aligned} B_1 :&= \left\{ \left\| T^{-1}\bar{R}_{[i]}\prod_{j=1}^{\tilde{p}}\mathbb{M}_{\widehat{V}_{(j)}^{\intercal}}\bar{G}^{\intercal}\right\|_{\text{MAX}} > (c+c_{qN}^{(\tilde{p}+1)})/2 \text{ for at most } qN-1 \text{ rows } i\in[N]\right\}, \\ B_2 :&= \left\{ \left\| T^{-1}\bar{R}_{[i]}\prod_{j=1}^{\tilde{p}}\mathbb{M}_{\widehat{V}_{(j)}^{\intercal}}\bar{G}^{\intercal} - \beta_{[i]}\prod_{j=1}^{\tilde{p}}\mathbb{M}_{b_j}\eta^{\intercal}\right\|_{\text{MAX}} \geq (c-c_{qN}^{(\tilde{p}+1)})/2 \text{ for some } i\in[N]\right\}, \end{aligned} \tag{B.117}$$

to obtain $\mathrm{P}(\widehat{p}=\tilde{p}) \geq \mathrm{P}(B_1) = 1-\mathrm{P}(B_1^c) \geq 1-\mathrm{P}(B_2) \to 1$. □

**Lemma 5.** *Suppose that $\Gamma_{(k)} \in \mathbb{R}^{|I_k|\times|I_k|}$ is an orthogonal matrix with the first $p$ rows equals to $\left(\beta_{[I_k]}^{\intercal}\beta_{[I_k]}\right)^{-\frac{1}{2}}\beta_{[I_k]}^{\intercal}$ and we define*

$$\begin{pmatrix} s_{(k)}^1 \\ s_{(k)}^2 \end{pmatrix} := \Gamma_{(k)}\varsigma_{(k)} \quad \text{and} \quad \begin{pmatrix} \widetilde{U}_{(k)}^1 \\ \widetilde{U}_{(k)}^2 \end{pmatrix} := \Gamma_{(k)}\bar{U}_{[I_k]},$$

*where $s_{(k)}^1 \in \mathbb{R}^{p\times 1}$ and $\widetilde{U}_{(k)}^1 \in \mathbb{R}^{p\times T}$ are the first $p$ rows of $\Gamma_{(k)}\varsigma_{(k)}$ and $\Gamma_{(k)}\bar{U}_{[I_k]}$, respectively. Then, under Assumptions A.1-A.8, we have*

*(i)* $\left\|s_{(k)}^2\right\| \lesssim_p T^{-1/2}\widehat{\lambda}_{(k)}^{-1/2}(|I_k|^{1/2}+T^{1/2})$.

*(ii)* $\left\|\widetilde{U}_{(k)}^1\right\| \lesssim_p T^{1/2}$, $\left\|\widetilde{U}_{(k)}^1\bar{V}^{\intercal}\right\| \lesssim_p T^{1/2}$, $\left\|\widetilde{U}_{(k)}^1\bar{Z}^{\intercal}\right\| \lesssim_p T^{1/2}$.

*Proof.* (i) The assumption $\widehat{I}_k = I_k$ and the definition (B.39) of $\widetilde{R}_{(k)}$ together lead to

$$\widetilde{R}_{(k)} = \bar{R}_{[I_k]}\prod_{i<k}\left(\mathbb{I}_T - \xi_{(i)}\xi_{(i)}^{\intercal}\right).$$

Then, with (B.56) and Lemma 2, we have $\varsigma_{(k)} = \bar{R}_{[I_k]}\xi_{(k)}/\sqrt{T\widehat{\lambda}_{(k)}}$. From the construction of $\Gamma_{(k)}$, we have

$$\Gamma_{(k)}\bar{R}_{(k)} = \begin{pmatrix} \left(\beta_{[I_k]}^{\intercal}\beta_{[I_k]}\right)^{\frac{1}{2}}\bar{V} + \widetilde{U}_{(k)}^1 \\ \widetilde{U}_{(k)}^2 \end{pmatrix},$$

which in turn gives

$$\begin{pmatrix} s^1_{(k)} \\ s^2_{(k)} \end{pmatrix} = \Gamma_{(k)}\varsigma_{(k)} = \frac{1}{\sqrt{T\widehat{\lambda}_{(k)}}} \begin{pmatrix} \left(\beta^\intercal_{[I_k]}\beta_{[I_k]}\right)^{\frac{1}{2}} \bar{V} + \widetilde{U}^1_{(k)} \\ \widetilde{U}^2_{(k)} \end{pmatrix} \xi_{(k)}.$$

With Lemma 1(v), we have

$$\left\| s^2_{(k)} \right\| = \left\| \frac{\widetilde{U}^2_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \leq \left\| \frac{\bar{U}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \lesssim_p T^{-1/2}\widehat{\lambda}^{-1/2}_{(k)}(|I_k|^{1/2} + T^{1/2}).$$

(ii) With Lemma 1(ii)(iii) and the definition of $\Gamma_{(k)}$, these results follow immediately. □

**Lemma 6.** *Under Assumptions A.1-A.8, if $\widehat{\lambda}_{(j)} \asymp_p |I_j|$ and $|I_j| \asymp qN$ for $j \leq \tilde{p}$, then for $k \leq \tilde{p}$, we have*

*(i)* $\left\| \frac{\bar{U}^\intercal_{[I_k]}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1}.$

*(ii)* $\left\| \frac{\bar{V}\bar{U}^\intercal_{[I_k]}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1}N^{-1} + T^{-1}$, $\left\| \frac{\bar{Z}\bar{U}^\intercal_{[I_k]}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1}N^{-1} + T^{-1}$, $|\frac{\varsigma^\intercal_{(k)}\bar{u}_{[I_k]}}{\sqrt{\widehat{\lambda}_{(k)}}}| \lesssim_p q^{-1}N^{-1} + T^{-1}$.

*Proof.* (i) Using the equation $\varsigma^\intercal_{(k)}\bar{U}_{[I_k]} = (s^1_{(k)})^\intercal\widetilde{U}^1_{(k)} + (s^2_{(k)})^\intercal\widetilde{U}^2_{(k)}$ and Lemma 5, we have

$$\begin{aligned} \left\| \varsigma^\intercal_{(k)}\bar{U}_{[I_k]} \right\| &\leq \left\| s^1_{(k)} \right\| \left\| \widetilde{U}^1_{(k)} \right\| + \left\| s^2_{(k)} \right\| \left\| \widetilde{U}^2_{(k)} \right\| \leq \left\| s^1_{(k)} \right\| \left\| \widetilde{U}^1_{(k)} \right\| + \left\| s^2_{(k)} \right\| \left\| \bar{U}_{[I_k]} \right\| \\ &\lesssim_p \sqrt{T} + \frac{|I_k| + T}{\sqrt{T\widehat{\lambda}_{(k)}}}, \end{aligned} \tag{B.118}$$

which leads to

$$\left\| \frac{\bar{U}^\intercal_{[I_k]}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \lesssim_p \frac{1}{\sqrt{\widehat{\lambda}_{(k)}}} + \frac{|I_k| + T}{T\widehat{\lambda}_{(k)}} \lesssim_p q^{-1/2}N^{-1/2} + T^{-1}.$$

(ii) From Lemmas 1 and 5, we have

$$\begin{aligned} \left\| \bar{V}\bar{U}^\intercal_{[I_k]}\varsigma^{(k)} \right\| &\leq \left\| \bar{V}\left(\widetilde{U}^1_{(k)}\right)^\intercal s^1_{(k)} \right\| + \left\| \bar{V}\left(\widetilde{U}^2_{(k)}\right)^\intercal s^2_{(k)} \right\| \leq \left\| \bar{V}\left(\widetilde{U}^1_{(k)}\right)^\intercal \right\| + \left\| \bar{V}\bar{U}^\intercal_{[I_k]} \right\| \left\| s^2_{(k)} \right\| \\ &\lesssim_p \sqrt{T} + \frac{|I_k| + T}{\sqrt{\widehat{\lambda}_{(k)}}}, \end{aligned}$$

which leads to

$$\left\| \frac{\bar{V}\bar{U}^\intercal_{[I_k]}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| \lesssim_p \frac{1}{\sqrt{T\widehat{\lambda}_{(k)}}} + \frac{|I_k| + T}{T\widehat{\lambda}_{(k)}} \lesssim_p q^{-1}N^{-1} + T^{-1}.$$

Replacing $\bar{V}$ by $\bar{Z}$ and $\iota_T^\intercal$ in the above proof and using Lemmas 1 and 5, we have similar results:

$$\left\| \frac{\bar{Z}\bar{U}_{[I_k]}^\intercal \varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1}N^{-1} + T^{-1}, \quad \text{and} \quad |\frac{\bar{u}_{[I_k]}^\intercal \varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}| \lesssim_p q^{-1}N^{-1} + T^{-1}. \tag{B.119}$$

□

**Lemma 7.** *Under Assumptions A.1-A.8, if $\widehat{\lambda}_{(j)} \asymp_p |I_j|$ and $|I_j| \asymp qN$ for $j \le \tilde{p}$, then for $k, l \le \tilde{p}$, we have*

*(i)* $\left\| \frac{\widetilde{U}_{(k)}^\intercal \varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1}, \quad \left\| \frac{\widetilde{U}_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}.$

*(ii)* $\left\| \frac{\bar{V}\widetilde{U}_{(k)}^\intercal \varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1}N^{-1}+T^{-1}, \quad \left\| \frac{\bar{Z}\widetilde{U}_{(k)}^\intercal \varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}} \right\| \lesssim_p q^{-1}N^{-1}+T^{-1}, \quad |\frac{\varsigma_{(k)}^\intercal \widetilde{u}_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}| \lesssim_p q^{-1}N^{-1}+T^{-1}.$

*(iii)* $|\frac{\xi_{(l)}^\intercal \widetilde{U}_{(k)}^\intercal \varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}| \lesssim_p q^{-1}N^{-1} + T^{-1}.$

*Proof.* (i) Recall that in the definition of $U_{(k)}$ in (B.40), we have

$$\widetilde{U}_{(k)} = \bar{U}_{[I_k]} - \sum_{i=1}^{k-1} \frac{\bar{R}_{[I_k]}\xi_{(i)}}{\sqrt{T}} \frac{\varsigma_{(i)}^\intercal \widetilde{U}_{(i)}}{\sqrt{\widehat{\lambda}_{(i)}}}. \tag{B.120}$$

Then, a direct multiplication of $\varsigma_{(k)}^\intercal/\sqrt{T\widehat{\lambda}_{(k)}}$ from the left side of (B.120) leads to

$$\frac{\varsigma_{(k)}^\intercal \widetilde{U}_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} = \frac{\varsigma_{(k)}^\intercal \bar{U}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}} - \sum_{i=1}^{k-1} \frac{\varsigma_{(k)}^\intercal \bar{R}_{[I_k]}\xi_{(i)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \frac{\varsigma_{(i)}^\intercal \widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}.$$

Consequently, with Lemma 6(i) we have

$$\begin{aligned} \left\| \frac{\varsigma_{(k)}^\intercal \widetilde{U}_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| &\le \left\| \frac{\varsigma_{(k)}^\intercal \bar{U}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| + \sum_{i=1}^{k-1} \left\| \frac{\bar{R}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}} \right\| \left\| \frac{\varsigma_{(i)}^\intercal \widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \right\| \lesssim_p \frac{1}{\sqrt{\widehat{\lambda}_{(k)}}} + \frac{|I_k|+T}{T\widehat{\lambda}_{(k)}} + \sqrt{\frac{|I_k|}{\widehat{\lambda}_{(k)}}} \sum_{i=1}^{k-1} \left\| \frac{\varsigma_{(i)}^\intercal \widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \right\| \\ &\lesssim_p q^{-1/2}N^{-1/2} + T^{-1} + \sum_{i=1}^{k-1} \left\| \frac{\varsigma_{(i)}^\intercal \widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}} \right\|. \end{aligned} \tag{B.121}$$

If $\left\| T^{-1/2}\widehat{\lambda}_{(i)}^{-1/2}\varsigma_{(i)}^\intercal \widetilde{U}_{(i)} \right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1}$ holds for $i \le k-1$, then (B.121) implies that this inequality also holds for $k$. In addition, when $k = 1$, $\widetilde{U}_{(1)} = \bar{U}_{[I_1]}$ and this equation is implied from Lemma 6(i). Therefore, we have $\left\| T^{-1/2}\widehat{\lambda}_{(k)}^{-1/2}\varsigma_{(k)}^\intercal \widetilde{U}_{(k)} \right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1}$ for $k \le \tilde{p}$ by induction.

Using (B.120) again, with Assumption A.4, we have

$$\left\|\frac{\widetilde{U}_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}\right\| \leq \left\|\frac{\bar{U}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}}\right\| + \sum_{i=1}^{k-1}\left\|\frac{\bar{R}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}}\right\|\left\|\frac{\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2} + \sum_{i=1}^{k-1}\left\|\frac{\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\|. \tag{B.122}$$

When $k=1$, Assumption A.4 implies that $\left\|T^{-1/2}\widehat{\lambda}_{(k)}^{-1/2}\widetilde{U}_{(k)}\right\| \lesssim_p q^{-1/2}N^{-1/2} + T^{-1/2}$. Then, using the same induction argument with (B.122), we have this ineqaulity holds for $k \leq \tilde{p}$.

(ii) Similarly, by simple multiplication of $\bar{V}^\intercal$ from the right side of (B.120), we have

$$\frac{\varsigma_{(k)}^\intercal\widetilde{U}_{(k)}\bar{V}^\intercal}{T\sqrt{\widehat{\lambda}_{(k)}}} = \frac{\varsigma_{(k)}^\intercal\bar{U}_{[I_k]}\bar{V}^\intercal}{T\sqrt{\widehat{\lambda}_{(k)}}} - \sum_{i=1}^{k-1}\frac{\varsigma_{(k)}^\intercal\bar{R}_{[I_k]}\xi_{(i)}}{\sqrt{T\widehat{\lambda}_{(k)}}}\frac{\varsigma_{(i)}^\intercal\widetilde{U}_{(i)}\bar{V}^\intercal}{T\sqrt{\widehat{\lambda}_{(i)}}}.$$

Consequently, we have

$$\begin{aligned}\left\|\frac{\varsigma_{(k)}^\intercal\widetilde{U}_{(k)}\bar{V}^\intercal}{T\sqrt{\widehat{\lambda}_{(k)}}}\right\| &\leq \left\|\frac{\varsigma_{(k)}^\intercal\bar{U}_{[I_k]}\bar{V}^\intercal}{T\sqrt{\widehat{\lambda}_{(k)}}}\right\| + \sum_{i=1}^{k-1}\left\|\frac{\bar{R}_{[I_k]}}{\sqrt{T\widehat{\lambda}_{(k)}}}\right\|\left\|\frac{\varsigma_{(i)}^\intercal\widetilde{U}_{(i)}\bar{V}^\intercal}{T\sqrt{\widehat{\lambda}_{(i)}}}\right\| \\ &\lesssim_p \frac{1}{\sqrt{T\widehat{\lambda}_{(k)}}} + \frac{|I_k|+T}{T\widehat{\lambda}_{(k)}} + \sqrt{\frac{|I_k|}{\widehat{\lambda}_{(k)}}}\sum_{i=1}^{k-1}\left\|\frac{\varsigma_{(i)}^\intercal\widetilde{U}_{(i)}\bar{V}^\intercal}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\| \\ &\lesssim_p q^{-1}N^{-1} + T^{-1} + \sum_{i=1}^{k-1}\left\|\frac{\varsigma_{(i)}^\intercal\widetilde{U}_{(i)}\bar{V}^\intercal}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\|.\end{aligned} \tag{B.123}$$

When $k=1$, $\left\|T^{-1}\widehat{\lambda}_{(k)}^{-1/2}\varsigma_{(k)}^\intercal\widetilde{U}_{(k)}\bar{V}^\intercal\right\| \lesssim_p q^{-1}N^{-1} + T^{-1}$ is a result of Lemma 6(ii). Then, a direct induction argument using (B.123) leads to this inequality for $k \leq \tilde{p}$.

Replacing $\bar{V}$ by $\bar{Z}$ and $\iota_T^\intercal$ in the above proof, and using Lemma 6(ii), we have the following results:

$$\left\|\frac{\bar{Z}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}}\right\| \lesssim_p q^{-1}N^{-1} + T^{-1} \quad \text{and} \quad \left|\frac{\widetilde{u}_{(k)}^\intercal\varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}}\right| \lesssim_p q^{-1}N^{-1} + T^{-1}.$$

(iii) Recall that $\widetilde{R}_{(k)} = \widetilde{\beta}_{(k)}\bar{V} + \widetilde{U}_{(k)}$ as defined in (B.39), we have

$$|\varsigma_{(l)}^\intercal\widetilde{R}_{(l)}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}| \leq |\varsigma_{(l)}^\intercal\widetilde{\beta}_{(l)}\bar{V}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}| + |\varsigma_{(l)}^\intercal\widetilde{U}_{(l)}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}| \leq \left\|\varsigma_{(l)}^\intercal\widetilde{\beta}_{(l)}\right\|\left\|\bar{V}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}\right\| + \left\|\varsigma_{(l)}^\intercal\widetilde{U}_{(l)}\right\|\left\|\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}\right\|.$$

Using (B.56), we have

$$|\frac{\xi_{(k)}^{\intercal}\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}| = |\frac{\varsigma_{(l)}^{\intercal}\widetilde{R}_{(l)}\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}\widehat{\lambda}_{(l)}}}| \leq \left\|\frac{\varsigma_{(l)}^{\intercal}\widetilde{\beta}_{(l)}}{\sqrt{\widehat{\lambda}_{(l)}}}\right\| \left\|\frac{\bar{V}\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}}\right\| + \left\|\frac{\widetilde{U}_{(k)}^{\intercal}\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}\right\| \left\|\frac{\widetilde{U}_{(l)}^{\intercal}\varsigma_{(l)}}{\sqrt{T\widehat{\lambda}_{(l)}}}\right\|. \tag{B.124}$$

With Lemma 1 and (i), we have

$$T^{1/2}\left\|\widetilde{\beta}_{(k)}\right\| \lesssim_p \sigma_p(\bar{V})\left\|\widetilde{\beta}_{(k)}\right\| \leq \left\|\widetilde{\beta}_{(k)}\bar{V}\right\| \leq \left\|\widetilde{U}_{(k)}\right\| + \left\|\widetilde{R}_{(k)}\right\| \leq \left\|\widetilde{U}_{(k)}\right\| + \|\bar{R}_{[I_k]}\| \lesssim_p T^{1/2}q^{1/2}N^{1/2}, \tag{B.125}$$

which leads to $\left\|\widehat{\lambda}_{(k)}^{-1/2}\varsigma_{(k)}^{\intercal}\widetilde{\beta}_{(k)}\right\| \lesssim_p q^{-1/2}N^{-1/2}\left\|\widetilde{\beta}_{(k)}\right\| \lesssim_p 1$. Using this inequality and results of (i) and (ii) in (B.124) completes the proof. □

**Lemma 8.** *Under Assumptions A.1-A.8, if $\widehat{\lambda}_{(j)} \asymp_p |I_j|$ and $|I_j| \asymp qN$ for $j \leq \tilde{p}$, then for $k \leq \tilde{p}+1$, we have*

*(i)* $\left\|\widetilde{Z}_{(k)}\bar{V}^{\intercal}\right\| \lesssim_p T^{1/2} + Tq^{-1}N^{-1}$.

*(ii)* $\left\|\widetilde{Z}_{(k)}\bar{U}_{[I_0]}^{\intercal}\right\| \lesssim_p N_0^{1/2}T^{1/2} + Tq^{-1/2}N^{-1/2}$.

*Proof.* (i) From the definition (B.44) of $\widetilde{Z}_{(k)}$, we have

$$\widetilde{Z}_{(k)}\bar{V}^{\intercal} = \bar{Z}\bar{V}^{\intercal} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}\bar{V}^{\intercal}}{\sqrt{T\widehat{\lambda}_{(i)}}}.$$

Then, with Lemma 7(ii), we have

$$\left\|\widetilde{Z}_{(k)}\bar{V}^{\intercal}\right\| \leq \|\bar{Z}\bar{V}^{\intercal}\| + \sum_{i=1}^{k-1}\|\bar{G}\xi_{(i)}\|\left\|\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}\bar{V}^{\intercal}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\| \lesssim_p T^{1/2} + T\left(q^{-1}N^{-1} + T^{-1}\right) \lesssim_p T^{1/2} + Tq^{-1}N^{-1}.$$

(ii) With (B.44) again, we have

$$\widetilde{Z}_{(k)}\bar{U}_{[I_0]}^{\intercal} = \bar{Z}\bar{U}_{[I_0]}^{\intercal} - \sum_{i=1}^{k-1}\bar{G}\xi_{(i)}\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}\bar{U}_{[I_0]}^{\intercal}}{\sqrt{T\widehat{\lambda}_{(i)}}},$$

which, along with Lemma 7(i) and the assumptions on $q$, lead to

$$\begin{aligned}\left\|\widetilde{Z}_{(k)}\bar{U}_{[I_0]}^{\intercal}\right\| &\leq \left\|\bar{Z}\bar{U}_{[I_0]}^{\intercal}\right\| + \sum_{i=1}^{k-1}\|\bar{G}\xi_{(i)}\|\left\|\frac{\varsigma_{(i)}^{\intercal}\widetilde{U}_{(i)}}{\sqrt{T\widehat{\lambda}_{(i)}}}\right\|\|\bar{U}_{[I_0]}\| \\ &\lesssim_p N_0^{1/2}T^{1/2} + \left(q^{-1/2}N^{-1/2} + T^{-1}\right)\left(N_0^{1/2}T^{1/2} + T\right)\end{aligned}$$

$$\lesssim_p N_0^{1/2} T^{1/2} + T q^{-1/2} N^{-1/2}.$$

□

**Lemma 9.** *Suppose that Assumptions A.1-A.8 hold. If $\widehat{\lambda}_{(j)} \asymp_p |I_j|$ and $|I_j| \asymp qN$ for $j \le \tilde{p}$, then $H_1$, $H_2$ defined by (B.54) satisfy*

*(i)* $\|H_1\| \lesssim_p 1$, $\|H_2\| \lesssim_p 1$.

*(ii)* $\|H_1^\intercal H_2 - \mathbb{I}_{\tilde{p}}\| \lesssim_p T^{-1} + q^{-1}N^{-1}$.

*(iii)* $\|H_1 - H_2\| \lesssim_p T^{-1/2} + q^{-1}N^{-1}$.

*Proof.* (i) Using the definition (B.54) of $H_1$ and Lemma 1, we have

$$\|h_{k1}\| = \left\| \frac{\bar{V}\xi_{(k)}}{\sqrt{T}} \right\| \le T^{-1/2} \left\|\bar{V}\right\| \lesssim_p 1,$$

which leads to $\|H_1\| \lesssim_p 1$.

Using the definition (B.54) of $H_2$, we have

$$\|h_{k2}\| = \left\| \frac{\widetilde{\beta}_{(k)}^\intercal \varsigma_{(k)}}{\sqrt{\widehat{\lambda}_{(k)}}} \right\| \le q^{-1/2} N^{-1/2} \left\|\widetilde{\beta}_{(k)}\right\|. \tag{B.126}$$

With Lemma 1 and Lemma 7(i), we have

$$T^{1/2} \left\|\widetilde{\beta}_{(k)}\right\| \lesssim_p \sigma_p(\bar{V}) \left\|\widetilde{\beta}_{(k)}\right\| \le \left\|\widetilde{\beta}_{(k)}\bar{V}\right\| \le \left\|\widetilde{U}_{(k)}\right\| + \left\|\widetilde{R}_{(k)}\right\| \le \left\|\widetilde{U}_{(k)}\right\| + \|\bar{R}_{[I_k]}\| \lesssim_p T^{1/2} q^{1/2} N^{1/2}. \tag{B.127}$$

Combining (B.126) and (B.127), we have $\|h_{k2}\| \lesssim_p 1$ and thus $\|H_2\| \lesssim_p 1$.

(ii) By (B.56) and Lemma 2, we have

$$\delta_{lk} = \xi_{(l)}^\intercal \xi_{(k)} = \frac{\xi_{(l)}^\intercal \bar{V}^\intercal \widetilde{\beta}_{(k)}^\intercal \varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} + \frac{\xi_{(l)}^\intercal \widetilde{U}_{(k)}^\intercal \varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}} = h_{l1}^\intercal h_{k2} + \frac{\xi_{(l)}^\intercal \widetilde{U}_{(k)}^\intercal \varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}.$$

By Lemma 7(iii), we have

$$|h_{l1}^\intercal h_{k2} - \delta_{lk}| \lesssim_p q^{-1}N^{-1} + T^{-1},$$

and thus $\|H_1^\intercal H_2 - \mathbb{I}_{\tilde{p}}\| \lesssim_p q^{-1}N^{-1} + T^{-1}$.

(iii) Using (B.56), we have

$$\bar{V}\xi_{(k)} = \frac{\bar{V}\bar{V}^\intercal\widetilde{\beta}_{(k)}^\intercal}{\sqrt{T\widehat{\lambda}_{(k)}}}\varsigma_{(k)} + \frac{\bar{V}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}}{\sqrt{T\widehat{\lambda}_{(k)}}}.$$

With the definition of $h_{k1}$ and $h_{k2}$, it becomes

$$h_{k1} = \frac{\bar{V}\bar{V}^\intercal}{T}h_{k2} + \frac{\bar{V}\widetilde{U}_{(k)}^\intercal\varsigma_{(k)}}{T\sqrt{\widehat{\lambda}_{(k)}}}. \tag{B.128}$$

With $\|h_{k2}\| \lesssim_p 1$, Lemma 1 and Lemma 7(ii), (B.128) leads to

$$h_{k1} - h_{k2} \lesssim_p T^{-1/2} + q^{-1}N^{-1}.$$

This completes the proof. □

**Lemma 10.** *For any $N \times p$ matrix $\beta$, if $\left\|T^{-1}\bar{V}\bar{V}^\intercal - \mathbb{I}_p\right\| \lesssim_p T^{-1/2}$, we have*

*(i) $\sigma_j(\beta\bar{V})/\sigma_j(\beta) = T^{1/2} + O_p(1)$ for $j \le p$.*

*(ii) If $\sigma_1(\beta) - \sigma_2(\beta) \asymp \sigma_1(\beta)$, then $\left\|\mathbb{P}_{\tilde{\xi}} - T^{-1}\bar{V}^\intercal\mathbb{P}_b\bar{V}\right\| \lesssim_p T^{-1/2}$, where $b$ is the first right singular vector of $\beta$ and $\tilde{\xi}$ is the first right singular vector of $\beta\bar{V}$.*

*Proof.* (i) For $j \le p$, $\sigma_j(\beta\bar{V})^2 = \lambda_j(\beta\bar{V}\bar{V}^\intercal\beta^\intercal) = \lambda_j(\beta^\intercal\beta\bar{V}\bar{V}^\intercal)$ which implies

$$\lambda_j(\beta^\intercal\beta)\lambda_p(\bar{V}\bar{V}^\intercal) \le \sigma_j(\beta\bar{V})^2 \le \lambda_j(\beta^\intercal\beta)\lambda_1(\bar{V}\bar{V}^\intercal).$$

With the assumption $\left\|T^{-1}\bar{V}\bar{V} - \mathbb{I}_p\right\| \lesssim_p T^{-1/2}$, we have $T^{-1/2}\sigma_j(\beta\bar{V})/\sigma_j(\beta) = 1 + O_p\left(T^{-1/2}\right)$ by sin-theta theorem.

(ii) Let $\varsigma$ and $\tilde{\varsigma}$ be the first singular vectors of $\beta$ and $\beta\bar{V}$, respectively. Equivalently, $\varsigma$ and $\tilde{\varsigma}$ are the eigenvectors of $\beta\beta^\intercal$ and $T^{-1}\beta\bar{V}\bar{V}^\intercal\beta^\intercal$. Since $\left\|\beta\beta^\intercal - T^{-1}\beta\bar{V}\bar{V}^\intercal\beta^\intercal\right\| \le \|\beta\|^2\left\|T^{-1}\bar{V}\bar{V}^\intercal - \mathbb{I}_p\right\| \lesssim_p \sigma_1(\beta)^2T^{-1/2}$ and $\sigma_1(\beta) - \sigma_2(\beta) \asymp \sigma_1(\beta)$, by sin-theta theorem we have

$$\|\varsigma\varsigma^\intercal - \tilde{\varsigma}\tilde{\varsigma}^\intercal\| \lesssim \frac{\left\|\beta\beta^\intercal - T^{-1}\beta\bar{V}\bar{V}^\intercal\beta^\intercal\right\|}{\sigma_1(\beta)^2 - \sigma_2(\beta)^2 - O(\left\|\beta\beta^\intercal - T^{-1}\beta\bar{V}\bar{V}^\intercal\beta^\intercal\right\|)} \lesssim_p T^{-1/2}.$$

Using the relationship between left and right singular vectors, we have

$$b^\intercal = \frac{\varsigma^\intercal\beta}{\sigma_1(\beta)}, \quad \tilde{\xi}^\intercal = \frac{\tilde{\varsigma}^\intercal\beta\bar{V}}{\left\|\beta\bar{V}\right\|}.$$

Therefore,

$$\left\|\mathbb{P}_{\tilde{\xi}}-\frac{\sigma_1(\beta)^2}{\left\|\beta\bar{V}\right\|^2}\bar{V}^\intercal\mathbb{P}_b\bar{V}\right\|=\left\|\tilde{\xi}\tilde{\xi}^\intercal-\frac{\bar{V}^\intercal\beta^\intercal\varsigma\varsigma^\intercal\beta\bar{V}}{\left\|\beta\bar{V}\right\|^2}\right\|=\left\|\frac{\bar{V}^\intercal\beta^\intercal\tilde{\varsigma}\tilde{\varsigma}^\intercal\beta\bar{V}}{\left\|\beta\bar{V}\right\|^2}-\frac{\bar{V}^\intercal\beta^\intercal\varsigma\varsigma^\intercal\beta\bar{V}}{\left\|\beta\bar{V}\right\|^2}\right\|\lesssim_p T^{-1/2}. \quad \text{(B.129)}$$

By Weyl's inequality, we have $T^{-1}\left\|\beta\bar{V}\right\|^2=\lambda_1(T^{-1}\beta\bar{V}\bar{V}^\intercal\beta^\intercal)=\lambda_1(\beta\beta^\intercal)+O_p(\sigma_1(\beta)^2T^{-1/2})=\sigma_1(\beta)^2+O_p(\sigma_1(\beta)^2T^{-1/2})$. Plugging this result into (B.129), we have $\left\|\mathbb{P}_{\tilde{\xi}}-T^{-1}\bar{V}^\intercal\mathbb{P}_b\bar{V}\right\|\lesssim_p T^{-1/2}$. □

Lemmas 11-13 below are concerned with the singular values and singular vectors of $T^{-1/2}\bar{R}$. We use $\varsigma_j$, $\xi_j$ and $\widehat{\lambda}_j^{1/2}$, $j\leq p$ to denote them throughout Lemmas 11-13.

**Lemma 11.** *Under the assumptions of Theorem 5(a), we have*

$$\frac{\widehat{\lambda}_j}{\lambda_j}-1\lesssim_p\lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)+T^{-1/2},$$

*where* $\lambda_j=\lambda_j(\beta^\intercal\beta)$ *and* $\widehat{\lambda}_j=\lambda_j(T^{-1}\bar{R}\bar{R}^\intercal)$.

*Proof.* Since $\lambda_j\left(\beta\bar{V}\bar{V}^\intercal\beta^\intercal\right)=\lambda_j\left(\beta^\intercal\beta\bar{V}\bar{V}^\intercal\right)$, we have

$$\lambda_j\left(\beta^\intercal\beta\right)\lambda_p\left(\frac{\bar{V}\bar{V}^\intercal}{T}\right)\leq\frac{\lambda_j\left(\beta^\intercal\beta\bar{V}\bar{V}^\intercal\right)}{T}\leq\lambda_j\left(\beta^\intercal\beta\right)\lambda_1\left(\frac{\bar{V}\bar{V}^\intercal}{T}\right). \quad \text{(B.130)}$$

By Lemma 1(i) and Weyl's inequality, we have $\lambda_j\left(T^{-1}\bar{V}\bar{V}^\intercal\right)-1\lesssim_p T^{-1/2}$ for $j\leq p$. Then, (B.130) becomes

$$\frac{\lambda_j\left(\beta\bar{V}\bar{V}^\intercal\beta^\intercal\right)}{T\lambda_j(\beta^\intercal\beta)}-1\lesssim_p T^{-1/2},$$

which is equivalent to

$$\frac{\sigma_j\left(\beta\bar{V}\right)}{\sqrt{T}\sigma_j(\beta)}-1\lesssim_p T^{-1/2}. \quad \text{(B.131)}$$

Using Weyl's inequality again, we have $\left|\sigma_j\left(\bar{R}\right)-\sigma_j\left(\beta\bar{V}\right)\right|\leq\left\|\bar{U}\right\|\lesssim_p N^{1/2}+T^{1/2}$, which is equivalent to

$$\frac{\widehat{\lambda}_j^{1/2}}{\lambda_j^{1/2}}-\frac{\sigma_j(\beta\bar{V})}{\sqrt{T}\sigma_j(\beta)}\lesssim_p\frac{1}{\sqrt{T}}+\frac{\sqrt{N}+\sqrt{T}}{\sqrt{T\lambda_j}}. \quad \text{(B.132)}$$

Combine (B.131) and (B.132), we complete the proof. □

**Lemma 12.** *Suppose that the SVD of $\beta$ is given by:*

$$\beta = \Gamma^\intercal \begin{pmatrix} \Lambda^{\frac{1}{2}} \\ 0 \end{pmatrix} H, \tag{B.133}$$

*where $\Gamma \in \mathbb{R}^{N\times N}$, $H \in \mathbb{R}^{p\times p}$ are orthogonal matrices, and $\Lambda$ is a diagonal matrix of the eigenvalues of $\beta^\intercal\beta$. If we write $\Gamma_{\varsigma_j} = (s_{j1}^\intercal, s_{j2}^\intercal)^\intercal$, where $s_{j1} \in \mathbb{R}^p$, $s_{j2} \in \mathbb{R}^{N-p}$. Then under the assumptions of Theorem 5(a), we have*

*(i)* $\left\|(\Lambda/\lambda_j)^{1/2}\left(s_{j1} - \langle s_{j1}, e_{j1}\rangle e_{j1}\right)\right\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)$, *where $e_{i1}$ is a $p\times 1$ unit vector with the $i$th entry being equal to 1.*

*(ii)* $\|s_{j1} - \langle s_{j1}, e_{j1}\rangle e_{j1}\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)$.

*(iii)* $\left\|(\Lambda/\lambda_j)^{1/2} s_{j1}\right\| \lesssim_p 1$.

*(iv)* $\|s_{j2}\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)$.

*Proof.* With the orthogonal matrix $\Gamma$ defined above, we can write

$$\widetilde{U} = \Gamma\bar{U} = \begin{pmatrix} \widetilde{U}_{1_{p\times T}} \\ \widetilde{U}_{2_{(N-p)\times T}} \end{pmatrix}, \tag{B.134}$$

so that

$$\Gamma\bar{R} = \begin{pmatrix} \Lambda^{\frac{1}{2}} \\ 0 \end{pmatrix}\bar{V} + \tilde{U} = \begin{pmatrix} \Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1 \\ \widetilde{U}_2 \end{pmatrix}.$$

The relationship between singular vectors $\varsigma_j$ and $\xi_j$ can be written as

$$\Gamma\varsigma_j = \frac{(\Gamma\bar{R})\,\xi_j}{\sqrt{T\widehat{\lambda}_j}}, \quad \xi_j = \frac{(\Gamma\bar{R})^\intercal(\Gamma\varsigma_j)}{\sqrt{T\widehat{\lambda}_j}}. \tag{B.135}$$

Specifically, we have

$$s_{j1} = \frac{\left(\Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1\right)\xi_j}{\sqrt{T\widehat{\lambda}_j}}, \quad s_{j2} = \frac{\widetilde{U}_2\xi_j}{\sqrt{T\widehat{\lambda}_j}}, \quad \xi_j = \frac{\left(\Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1\right)^\intercal s_{j1} + \widetilde{U}_2^\intercal s_{j2}}{\sqrt{T\widehat{\lambda}_j}}. \tag{B.136}$$

From (B.136), we have

$$\left(\Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1\right)\left(\Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1\right)^\intercal s_{j1} + \left(\Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1\right)\widetilde{U}_2^\intercal s_{j2} = T\widehat{\lambda}_j s_{j1}. \tag{B.137}$$

We can rewrite (B.137) as

$$\left(\mathbb{I}_p - \frac{\Lambda}{\lambda_j}\right) s_{j1} = \frac{1}{T\lambda_j}\left(\Lambda^{\frac{1}{2}}\bar{V} + \widetilde{U}_1\right)\widetilde{U}_2^\intercal s_{j2} + \frac{1}{\lambda_j}\Lambda^{\frac{1}{2}}\left(\frac{\bar{V}\bar{V}^\intercal}{T} - I\right)\Lambda^{\frac{1}{2}} s_{j1} + \frac{\Lambda^{\frac{1}{2}}\bar{V}\widetilde{U}_1^\intercal}{T\lambda_j} s_{j1} + \frac{\widetilde{U}_1\bar{V}^\intercal\Lambda^{\frac{1}{2}}}{T\lambda_j} s_{j1} + \frac{\widetilde{U}_1\widetilde{U}_1^\intercal}{T\lambda_j} s_{j1} - \left(\frac{\widehat{\lambda}_j}{\lambda_j} - 1\right) s_{j1}. \tag{B.138}$$

Define $L = \mathrm{diag}(l_1, \ldots, l_p)$, where $l_i$ is equal to $\lambda_j/(\lambda_j - \lambda_i)$ if $i \neq j$ and 0 otherwise.

By left multiplying $L$ to both sides of (B.138), we have

$$s_{j1} - \langle s_{j1}, e_{j1}\rangle e_{j1} = \frac{1}{T\lambda_j} L\Lambda^{\frac{1}{2}}\bar{V}\frac{\widetilde{U}_2^\intercal\widetilde{U}_2}{\sqrt{T\widehat{\lambda}_j}}\xi_j + \frac{1}{T\lambda_j} L\widetilde{U}_1\frac{\widetilde{U}_2^\intercal\widetilde{U}_2}{\sqrt{T\widehat{\lambda}_j}}\xi_j + \frac{1}{\lambda_j} L\Lambda^{\frac{1}{2}}\left(\frac{\bar{V}\bar{V}^\intercal}{T} - \mathbb{I}_p\right)\Lambda^{\frac{1}{2}} s_{j1} + \frac{L\Lambda^{\frac{1}{2}}\bar{V}\widetilde{U}_1^\intercal}{T\lambda_j} s_{j1} + L\frac{\widetilde{U}_1\bar{V}^\intercal\Lambda^{\frac{1}{2}}}{T\lambda_j} s_{j1} + L\frac{\widetilde{U}_1\widetilde{U}_1^\intercal}{T\lambda_j} s_{j1} - \left(\frac{\widehat{\lambda}_j}{\lambda_j} - 1\right) L s_{j1}. \tag{B.139}$$

Now left multiplying $\left(\frac{\Lambda}{\lambda_j}\right)^{\frac{1}{2}}$ again, we have

$$\begin{aligned}\left(\frac{\Lambda}{\lambda_j}\right)^{\frac{1}{2}}\left(s_{j1} - \langle s_{j1}, e_{j1}\rangle e_{j1}\right) =& \frac{1}{T\lambda_j^{3/2}}\Lambda^{\frac{1}{2}} L\Lambda^{\frac{1}{2}}\bar{V}\frac{\widetilde{U}_2^\intercal\widetilde{U}_2}{\sqrt{T\widehat{\lambda}_j}}\xi_j + \frac{1}{T\lambda_j^{3/2}}\Lambda^{\frac{1}{2}} L\widetilde{U}_1\frac{\widetilde{U}_2^\intercal\widetilde{U}_2}{\sqrt{T\widehat{\lambda}_j}}\xi_j \\ &+ \frac{1}{\lambda_j}\Lambda^{\frac{1}{2}} L\Lambda^{\frac{1}{2}}\left(\frac{\bar{V}\bar{V}^\intercal}{T} - \mathbb{I}_p\right)\left(\frac{\Lambda}{\lambda_j}\right)^{\frac{1}{2}} s_{j1} + \Lambda^{\frac{1}{2}} L\Lambda^{\frac{1}{2}}\frac{\bar{V}\widetilde{U}_1^\intercal}{T\lambda_j^{3/2}} s_{j1} \\ &+ \Lambda^{\frac{1}{2}} L\frac{\widetilde{U}_1\bar{V}^\intercal}{T\lambda_j}\left(\frac{\Lambda}{\lambda_j}\right)^{\frac{1}{2}} s_{j1} + \Lambda^{\frac{1}{2}} L\frac{\widetilde{U}_1\widetilde{U}_1^\intercal}{T\lambda_j^{3/2}} s_{j1} - \left(\frac{\widehat{\lambda}_j}{\lambda_j} - 1\right)\left(\frac{\Lambda}{\lambda_j}\right)^{\frac{1}{2}} L s_{j1} \\ =& K_1 + K_2 + K_3 + K_4 + K_5 + K_6 + K_7.\end{aligned} \tag{B.140}$$

Before we analyze these seven terms in (B.140), we first analyze $\|L\|$, $\left\|L\Lambda^{1/2}\right\|$ and $\|L\Lambda\|$. Since $L$ and $\Lambda$ are diagonal matrices, by Assumption A.13 we can easily show that

$$\|L\| \lesssim 1, \quad \left\|L\Lambda^{1/2}\right\| \lesssim \lambda_j^{1/2}, \quad \|L\Lambda\| \lesssim \lambda_j. \tag{B.141}$$

In addition, Lemma 1(ii)(iii)(v) imply that

$$\left\|\widetilde{U}_1\right\| = \left\|(\beta^\intercal\beta)^{-1/2}\beta^\intercal\bar{U}\right\| \lesssim_p T^{1/2}, \quad \left\|\widetilde{U}_1\bar{V}^\intercal\right\| = \left\|(\beta^\intercal\beta)^{-1/2}\beta^\intercal\bar{U}\bar{V}^\intercal\right\| \lesssim_p T^{1/2}, \quad \left\|\widetilde{U}_2\right\| \leq \left\|\bar{U}\right\| \lesssim_p N^{1/2} + T^{1/2}. \tag{B.142}$$

Using Lemma 1(i)(vi), Lemma 11, (B.141) and (B.142), we analyze these seven terms in (B.140) one

by one. For the first term, we have

$$\|K_1\| \leq T^{-3/2}\lambda_j^{-3/2}\widehat{\lambda}_j^{-1/2}\|L\Lambda\|\left\|\bar{V}\right\|\left\|\widetilde{U}_2^\intercal\widetilde{U}_2\right\|\|\xi_j\| \lesssim_p \lambda_j^{-1}(T^{-1}N+1),$$

where we also use $\left\|\widetilde{U}_2^\intercal\widetilde{U}_2\right\| \leq \left\|\bar{U}^\intercal\bar{U}\right\| \lesssim_p N+T$ in the last equation. For the second term, we have

$$\|K_2\| \leq T^{-3/2}\lambda_j^{-3/2}\widehat{\lambda}_j^{-1/2}\left\|\Lambda^{1/2}L\right\|\left\|\widetilde{U}_1\right\|\left\|\widetilde{U}_2^\intercal\widetilde{U}_2\right\|\|\xi_j\| \lesssim_p \lambda_j^{-3/2}(T^{-1}N+1).$$

For the third term, we have

$$\|K_3\| \leq \lambda_j^{-1}\|L\Lambda\|\left\|T^{-1}\bar{V}\bar{V}^\intercal-\mathbb{I}_p\right\|\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\| \lesssim_p T^{-1/2}\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\|.$$

For the forth term, we have

$$\|K_4\| \leq T^{-1}\lambda_j^{-3/2}\|L\Lambda\|\left\|\bar{V}\widetilde{U}_1^\intercal\right\| \lesssim_p \lambda_j^{-1/2}T^{-1/2},$$

where we use $\left\|\bar{V}\widetilde{U}_1^\intercal\right\| \lesssim_p T^{1/2}$ from Lemma 1. For the fifth term, we have

$$\|K_5\| \leq T^{-1}\lambda_j^{-1}\left\|L\Lambda^{1/2}\right\|\left\|\widetilde{U}_1\bar{V}^\intercal\right\|\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\| \lesssim_p \lambda_j^{-1/2}T^{-1/2}\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\|.$$

For the sixth term, we have

$$\|K_6\| \leq T^{-1}\lambda_j^{-3/2}\left\|L\Lambda^{1/2}\right\|\left\|\widetilde{U}_1\widetilde{U}_1^\intercal\right\| \lesssim_p \lambda_j^{-1},$$

where we use $\left\|\widetilde{U}_1\widetilde{U}_1^\intercal\right\| \lesssim_p T$ as shown in Lemma 1. For the last term, we have

$$\|K_7\| \leq \lambda_j^{-2}|\widehat{\lambda}_j-\lambda_j|\left\|L\Lambda^{1/2}\right\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)+T^{-1/2}.$$

To sum up, (B.140) gives

$$\left\|(\Lambda/\lambda_j)^{1/2}\left(s_{j1}-\langle s_{j1},e_{j1}\rangle e_{j1}\right)\right\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)+T^{-1/2}+T^{-1/2}\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\|. \tag{B.143}$$

Note that

$$\begin{aligned}\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\| &\leq \left\|(\Lambda/\lambda_j)^{1/2}\left(s_{j1}-\langle s_{j1},e_{j1}\rangle e_{j1}\right)\right\|+\left\|(\Lambda/\lambda_j)^{1/2}\langle s_{j1},e_{j1}\rangle e_{j1}\right\| \\ &\leq \left\|(\Lambda/\lambda_j)^{1/2}\left(s_{j1}-\langle s_{j1},e_{j1}\rangle e_{j1}\right)\right\|+|\langle s_{j1},e_{j1}\rangle|\sqrt{\lambda_j^{-1}e_{j1}^\intercal\Lambda e_{j1}} \\ &= \left\|(\Lambda/\lambda_j)^{1/2}\left(s_{j1}-\langle s_{j1},e_{j1}\rangle e_{j1}\right)\right\|+O_p(1).\end{aligned}$$

Plugging this into (B.143), we have

$$\left\|(\Lambda/\lambda_j)^{1/2}\left(s_{j1}-\langle s_{j1},e_{j1}\rangle e_{j1}\right)\right\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)+T^{-1/2}, \tag{B.144}$$

which in turn leads to $\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\| \lesssim_p 1$ as by assumption $\lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)\to 0$. Similarly, we can analyze corresponding terms in (B.139), and obtain

$$\|s_{j1}-\langle s_{j1},e_{j1}\rangle e_{j1}\| \lesssim_p T^{-1/2}\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\|+\lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1) \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1)+T^{-1/2}.$$

From (B.136), we have

$$\|s_{j2}\| \le \left\|\frac{\widetilde{U}_2}{\sqrt{T\lambda_j}}\right\|\left\|\left(\frac{\lambda_j}{\widehat{\lambda}_j}\right)^{\frac{1}{2}}\right\|\|\xi_j\| \lesssim_p \lambda_j^{-1/2}(T^{-1/2}N^{1/2}+1). \tag{B.145}$$

This concludes the proof. □

**Lemma 13.** *Under the assumptions of Theorem 5(a), we have*

*(i)* $\left\|\frac{\xi_i^\intercal \bar{U}^\intercal \varsigma_j}{\sqrt{T\widehat{\lambda}_j}}\right\| \lesssim_p \frac{1}{T}+\frac{N+T}{T\lambda_i}+\frac{N+T}{T\lambda_j}$.

*(ii)* $\left\|\frac{\bar{V}\bar{U}^\intercal\varsigma_i}{T\sqrt{\widehat{\lambda}_i}}\right\| \lesssim_p \frac{1}{T}+\frac{N+T}{T\lambda_i}$, $\left|\frac{\varsigma_i^\intercal \bar{u}}{\sqrt{\widehat{\lambda}_i}}\right| \lesssim_p \frac{1}{T}+\frac{N+T}{T\lambda_i}$.

*(iv)* $\left\|\frac{\varsigma_i^\intercal \bar{U}}{\sqrt{T\widehat{\lambda}_i}}\right\| \lesssim_p \frac{1}{\sqrt{\lambda_i}}+\frac{N+T}{T\lambda_i}$.

*Proof.* (i) From (B.135), we have

$$\frac{\xi_i^\intercal \bar{U}^\intercal\varsigma_j}{\sqrt{T\widehat{\lambda}_j}} = \frac{\varsigma_i^\intercal \bar{R}\bar{U}^\intercal\varsigma_j}{T\sqrt{\widehat{\lambda}_i\widehat{\lambda}_j}}.$$

Using the orthogonal matrix $\Gamma$ and the notations in Lemma 11 and Lemma 12, we have

$$\begin{aligned}\varsigma_i^\intercal \bar{R}\bar{U}^\intercal\varsigma_j = s_i^\intercal\left(\Gamma\beta\bar{V}+\widetilde{U}\right)\widetilde{U}^\intercal s_j =& s_{i1}^\intercal\left(\Lambda^{\frac{1}{2}}\bar{V}+\widetilde{U}_1\right)\widetilde{U}_1^\intercal s_{j1}+s_{i2}^\intercal\widetilde{U}_2\widetilde{U}_1^\intercal s_{j1}\\ &+s_{i1}^\intercal\left(\Lambda^{\frac{1}{2}}\bar{V}+\widetilde{U}_1\right)\widetilde{U}_2^\intercal s_{j2}+s_{i2}^\intercal\widetilde{U}_2\widetilde{U}_2^\intercal s_{j2}\\ =&K_1+K_2+K_3+K_4.\end{aligned}$$

Recall that from Lemma 12, we have $\left\|(\Lambda/\lambda_j)^{1/2}s_{j1}\right\| \lesssim_p 1$. Using this result and Lemma 1, we analyze these four terms one by one. For the first term, we have

$$\|K_1\| \le \left\|s_{i1}^\intercal\Lambda^{\frac{1}{2}}\right\|\left\|\bar{V}\widetilde{U}_1^\intercal\right\|\|s_{j1}\|+\|s_{i1}\|\left\|\widetilde{U}_1\widetilde{U}_1^\intercal\right\|\|s_{j1}\| \lesssim_p \sqrt{\lambda_i T}+T.$$

For the second term, we have

$$\|K_2\| \le \|s_{i2}\| \left\|\widetilde{U}_2\right\| \left\|\widetilde{U}_1\right\| \lesssim_p \sqrt{\frac{N+T}{T\lambda_i}} \left(\sqrt{N}+\sqrt{T}\right)\sqrt{T} \lesssim_p \lambda_i^{-1/2}(N+T).$$

For the third term, we have

$$\|K_3\| \le \left(\left\|s_{i1}^\intercal \Lambda^{\frac{1}{2}}\right\| \|\bar{V}\| + \left\|\widetilde{U}_1\right\|\right) \left\|\widetilde{U}_2\right\| \|s_{j2}\| \lesssim_p \sqrt{\lambda_i T}\left(\sqrt{N}+\sqrt{T}\right)\sqrt{\frac{N+T}{T\lambda_j}} = \lambda_j^{-1/2}\lambda_i^{1/2}(N+T).$$

For the last term, we have

$$\|K_4\| \le \left\|\widetilde{U}_2\widetilde{U}_2^\intercal\right\| \|s_{i2}\| \|s_{j2}\| \lesssim_p \lambda_i^{-1/2}\lambda_j^{-1/2}T^{-1}(N+T)^2.$$

Using above equations and Lemma 11, we get

$$\left\|\frac{\xi_i^\intercal \bar{U}^\intercal \varsigma_j}{\sqrt{T\widehat{\lambda}_j}}\right\| = \left\|\frac{\varsigma_i^\intercal \bar{R}\bar{U}^\intercal \varsigma_j}{T\sqrt{\widehat{\lambda}_i\widehat{\lambda}_j}}\right\| \lesssim_p \frac{1}{T} + \frac{N+T}{T\lambda_i} + \frac{N+T}{T\lambda_j}.$$

(ii) Using $\bar{U}^\intercal\varsigma_i = \widetilde{U}_1^\intercal s_{i1} + \widetilde{U}_2^\intercal s_{i2}$ and (B.142), we have

$$\left\|\bar{V}\bar{U}^\intercal\varsigma_i\right\| \le \left\|\bar{V}\widetilde{U}_1^\intercal s_{i1}\right\| + \left\|\bar{V}\widetilde{U}_2^\intercal s_{i2}\right\| \le \left\|\bar{V}\widetilde{U}_1^\intercal\right\| + \|\bar{V}\|\,\|\bar{U}\|\,\|s_{i2}\| \lesssim_p \sqrt{T} + \frac{N+T}{\sqrt{\lambda_i}}.$$

Then, with Lemma 11, we have $\left\|T^{-1}\widehat{\lambda}_i^{-1/2}\bar{V}\bar{U}^\intercal\varsigma_i\right\| \lesssim_p T^{-1} + \lambda_i^{-1}(T^{-1}N+1)$.

Replace $\bar{V}$ in the above proof by $\iota_T^\intercal$, we can get $\left\|\widehat{\lambda}_i^{-1/2}\bar{u}^\intercal\varsigma_i\right\| \lesssim_p T^{-1} + \lambda_i^{-1}(T^{-1}N+1)$.

(iii) Using $\bar{U}^\intercal\varsigma_i = \widetilde{U}_1^\intercal s_{i1} + \widetilde{U}_2^\intercal s_{i2}$ and (B.142), we have

$$\left\|\varsigma_i^\intercal\bar{U}\right\| \le \left\|s_{i1}^\intercal\widetilde{U}_1\right\| + \left\|s_{i2}^\intercal\widetilde{U}_2\right\| \le \left\|\widetilde{U}_1\right\| + \|\bar{U}\| \lesssim_p \sqrt{T} + \frac{N+T}{\sqrt{T\lambda_i}}.$$

Applying Lemma 11 again completes the proof. □

**Lemma 14.** *Under the assumptions of Theorem 5(a), $\tilde{H}_1$, $\tilde{H}_2$ defined by (B.75) satisfy*

*(i)* $\left\|\tilde{H}_1\right\| \lesssim_p 1$, $\left\|\tilde{H}_2\right\| \lesssim_p 1$.

*(ii)* $\left\|\tilde{H}_1^\intercal\tilde{H}_2 - \mathbb{I}_{\tilde{p}}\right\| \lesssim_p T^{-1} + \lambda_p^{-1}(T^{-1}N+1)$.

*(iii)* $\left\|\tilde{H}_1 - \tilde{H}_2\right\| \lesssim_p T^{-1/2} + \lambda_p^{-1}(T^{-1}N+1)$.

*Proof.* (i) Using the definition of $\tilde{H}_1$ in (B.75) and Lemma 1, we have

$$\left\|\tilde{h}_{k1}\right\| = \left\|\frac{\bar{V}\xi_k}{\sqrt{T}}\right\| \le T^{-1/2}\|\bar{V}\| \lesssim_p 1,$$

which leads to $\left\|\tilde{H}_1\right\| \lesssim_p 1$.

Using $\Gamma\varsigma_k = (s_{k1}^\intercal, s_{k2}^\intercal)^\intercal$, the SVD of $\beta$ in (B.133), the definition of $\tilde{H}_2$ in (B.75), Lemma 11 and Lemma 12(iii), we have

$$\left\|\tilde{h}_{k2}\right\| = \left\|\frac{\beta^\intercal \varsigma_k}{\sqrt{\widehat{\lambda}_k}}\right\| = \left\|\frac{\Lambda^{1/2} s_{k1}}{\sqrt{\widehat{\lambda}_k}}\right\| \lesssim_p 1, \tag{B.146}$$

which leads to $\left\|\tilde{H}_2\right\| \lesssim_p 1$.

(ii) By (B.135) and Lemma 2, for $l, k \leq p$, we have

$$\delta_{lk} = \xi_l^\intercal \xi_k = \frac{\xi_l^\intercal \bar{V}^\intercal \beta^\intercal \varsigma_k}{\sqrt{T\widehat{\lambda}_k}} + \frac{\xi_l^\intercal \bar{U}^\intercal \varsigma_k}{\sqrt{T\widehat{\lambda}_k}} = \tilde{h}_{l1}^\intercal \tilde{h}_{k2} + \frac{\xi_l^\intercal \bar{U}^\intercal \varsigma_k}{\sqrt{T\widehat{\lambda}_k}}.$$

By Lemma 13(i), we have

$$|\tilde{h}_{l1}^\intercal \tilde{h}_{k2} - \delta_{lk}| \lesssim_p \frac{1}{T} + \frac{N+T}{T\min\{\lambda_l, \lambda_k\}} \leq \frac{1}{T} + \frac{N+T}{T\lambda_p},$$

and thus $\left\|\tilde{H}_1^\intercal \tilde{H}_2 - \mathbb{I}_p\right\| \lesssim_p T^{-1} + \lambda_p^{-1}(T^{-1}N + 1)$.

(iii) Using (B.135), we have

$$\bar{V}\xi_k = \frac{\bar{V}\bar{V}^\intercal \beta^\intercal}{\sqrt{T\widehat{\lambda}_k}}\varsigma_k + \frac{\bar{V}\bar{U}^\intercal \varsigma_k}{\sqrt{T\widehat{\lambda}_k}}.$$

With the definition of $h_{k1}$ and $h_{k2}$, it becomes

$$\tilde{h}_{k1} = \frac{\bar{V}\bar{V}^\intercal}{T}\tilde{h}_{k2} + \frac{\bar{V}\bar{U}^\intercal \varsigma_k}{T\sqrt{\widehat{\lambda}_k}}. \tag{B.147}$$

With $\left\|\tilde{h}_{k2}\right\| \lesssim_p 1$, Lemma 1 and Lemma 13(ii), (B.147) leads to

$$\left\|\tilde{h}_{k1} - \tilde{h}_{k2}\right\| \leq \left\|T^{-1}\bar{V}\bar{V}^\intercal - \mathbb{I}_p\right\| \left\|\tilde{h}_{k2}\right\| + \left\|\frac{\bar{V}\bar{U}^\intercal \varsigma_k}{T\sqrt{\widehat{\lambda}_k}}\right\| \lesssim_p T^{-1/2} + \lambda_p^{-1}(T^{-1}N + 1),$$

which concludes the proof of (iii). □

**Lemma 15.** *Under Assumption A.13, we have*

$$\left\|\bar{r} - \widehat{\Sigma} b\right\|_\infty \lesssim_p \sqrt{\frac{\log N}{T}}, \qquad \left\|b^\intercal(\bar{r} - \mathrm{E}(r_t))\right\| \lesssim_p \frac{1}{\sqrt{T}}.$$

*Proof.* For the first inequality, we have

$$\left\|\bar{r}-\widehat{\Sigma}b\right\|_{\infty}\leq\left\|\bar{r}-\mathrm{E}(r)\right\|_{\infty}+\left\|\Sigma b-\widehat{\Sigma}b\right\|_{\infty}\lesssim_{p}\sqrt{\frac{\log N}{T}},$$

where we use large deviation inequalities in Assumption A.12:

$$\left\|\bar{r}-\mathrm{E}(r_t)\right\|_{\infty}\lesssim_{p}\sqrt{\frac{\log N}{T}},\quad\text{and}\quad\left\|\Sigma b-\widehat{\Sigma}b\right\|_{\infty}=\left\|\frac{1}{T}\bar{R}\bar{R}^{\intercal}b-\mathrm{Cov}(r_t,r_t^{\intercal}b)\right\|_{\infty}\lesssim_{p}\sqrt{\frac{\log N}{T}}.$$

The second inequality follows immediately from Assumption A.12:

$$\left\|b^{\intercal}(\bar{r}-\mathrm{E}(r_t))\right\|=\left|\frac{1}{T}\sum_{t=1}^{T}m_t-\mathrm{E}(m_t)\right|\lesssim_{p}\frac{1}{\sqrt{T}}.$$

□

## References


Bai, Z. and J. W. Silverstein (2009). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis 7*(1), 1–46.

Fan, J., Y. Liao, and M. Mincheva (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics 39*(6), 3320–3356.

Giglio, S. W. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Wang, W. and J. Fan (2017, 06). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist. 45*(3), 1342–1374.

Wedin, P.-Å. (1972). Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics 12*(1), 99–111.