Supplementary material on "Identifying Cointegration by Eigenanalysis"

Rongmao Zhang[⋆] Peter Robinson[†] Qiwei Yao[‡]

[⋆]School of Mathematics, Zhejiang University, Hangzhou, 310058, China

[†]Department of Economics, London School of Economics
London, WC2A 2AE, U.K.

[‡]Department of Statistics, London School of Economics
London, WC2A 2AE, U.K.

S.1 Proof for Section 3.1

Let

$$\begin{aligned}\mathbf{\Sigma}_j^x &= \operatorname{diag}\left[\left(\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\overline{\mathbf{x}}_1)(\mathbf{x}_{t1}-\overline{\mathbf{x}}_1)'\right), \left(\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2}-\overline{\mathbf{x}}_2)(\mathbf{x}_{t2}-\overline{\mathbf{x}}_2)'\right)\right] \\ &\equiv \operatorname{diag}(\mathbf{\Sigma}_{j1}^x, \mathbf{\Sigma}_{j2}^x),\end{aligned}$$

$\mathbf{W}^x = \sum_{j=0}^{j_0} \mathbf{\Sigma}_j^x(\mathbf{\Sigma}_j^x)' =: \operatorname{diag}(\mathbf{D}_1^x, \mathbf{D}_2^x)$ and $\mathbf{\Gamma}_x$ be the $p\times p$ orthogonal matrix such that

$$\mathbf{W}^x\mathbf{\Gamma}_x = \mathbf{\Gamma}_x\mathbf{\Lambda}_x,$$

where $\mathbf{\Lambda}_x$ is the diagonal matrix of eigenvalues of $\mathbf{W}^x$. Since $\mathbf{x}_{t1}$ is nonstationary and $\mathbf{x}_{t2}$ is stationary, intuitively $\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\overline{\mathbf{x}}_1)\ (\mathbf{x}_{t1}-\overline{\mathbf{x}}_1)'$ and $\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2}-\overline{\mathbf{x}}_2)(\mathbf{x}_{t2}-\overline{\mathbf{x}}_2)'$ do not share the same eigenvalues, so $\mathbf{\Gamma}_x$ must be block-diagonal. Define $\mathbf{W}^y = \mathbf{A}\mathbf{W}^x\mathbf{A}'$, then

$$\mathbf{W}^y = \mathbf{A}\mathbf{W}^x\mathbf{A}' = \mathbf{A}\mathbf{\Gamma}_x\mathbf{\Lambda}_x\mathbf{\Gamma}_x'\mathbf{A}'.$$

This implies that the columns of $\mathbf{A}\mathbf{\Gamma}_x$ are just the orthogonal eigenvectors of $\mathbf{W}^y$. Since $\mathbf{\Gamma}_x$ is block-diagonal, it follows that $\mathcal{M}(\mathbf{A}_2)$ is the same as the space spanned by the eigenvectors corresponding to the smallest r eigenvalues of $\mathbf{W}^y$. As a result, to show the distance between the cointegration space and its estimate is small, we only need to show that the space spanned by the eigenvectors of $\mathbf{W}^y$ can be approximated by that of $\widehat{\mathbf{W}}$. This question is usually solved by perturbation matrix theory. In particular, let

$$\widehat{\mathbf{W}} = \mathbf{W}^y + \Delta\mathbf{W}^y,\ \ \Delta\mathbf{W}^y = \widehat{\mathbf{W}} - \mathbf{W}^y,$$

and

$$\operatorname{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x) = \min_{\lambda \in \lambda(\mathbf{D}_1^x),\, \mu \in \lambda(\mathbf{D}_2^x)} |\lambda - \mu|,$$

where $\lambda(\mathbf{A})$ denotes the set of eigenvalues of a matrix $\mathbf{A}$. When $||\Delta \mathbf{W}^y|| = o_p(\operatorname{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x))$, one can use the perturbation results of Golub and Loan (1996) to establish the bound of Theorems 3.1, 3.3 and 4.1, see also Lam and Yao (2012) or Chang, Guo and Yao (2017). However, in our setting $\operatorname{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x)$ can be of smaller order than $||\Delta \mathbf{W}^y||$, i.e., $\operatorname{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x)/||\Delta \mathbf{W}^y|| \xrightarrow{p} 0$ as $n \to \infty$ and the above method will not work.

To fix this problem, we adopt the perturbation results of Dopico, Moro and Molera (2000) instead. A similar idea was used by Chen and Hurvich (2006) to recover their fractional cointegration spaces via the periodogram matrix, using a random diagonal block matrix instead. However, because of the quadratic form of $\mathbf{W}^x$ $(= \sum_{j=1}^{j_0} \mathbf{\Sigma}_j^x (\mathbf{\Sigma}_j^x)')$, we cannot find a normalizing constant matrix $\mathbf{C}_n$ such that $\mathbf{C}_n \mathbf{W}^x \mathbf{C}_n = O_e(1)$ or $\mathbf{C}_n \mathbf{W}^y \mathbf{C}_n = O_e(1)$, the argument of Chen and Hurvich (2006) based on the perturbation bound of Barlow and Slapnicar (2002) cannot be used. We first establish some lemmas (i.e. Lemmas 1-4 below), which will be used to prove Theorem 3.1.

For $1 \le i \le p - r$, set $f_0^i(t) = W^i(t)$, $f_{d_i}^i(t) = \int_0^t f_{d_i-1}^i(s)\, dt$, $\mu_i = \mathrm{E} z_t^i$ and define

$$F^i(t) \quad = \quad f_{d_i}^i(t) - \int_0^1 f_{d_i}^i(t)\, dt, \quad G_d(t) = \frac{\prod_{j=0}^{d-1}(t+j)}{d!}, \quad \bar{G}_d = \frac{1}{n}\sum_{t=1}^n G_d(t).$$

Then, we have the following weak convergence result for the sample autocovariance.

Lemma 1. *Let $L_d(t) = G_d(t) - \bar{G}_d$. Suppose $x_t^i \sim I(d_i)$, $1 \le i \le p - r$, then under Condition 1,*

$$\Big(\frac{x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)}{n^{d_i - 1/2}}, 1 \le i \le p - r\Big) \xrightarrow{d} \Big(F^i(t), 1 \le i \le p - r\Big) \text{ and } \tag{S.1}$$

$$\Big(\frac{1}{n^{d_i + 1/2}} \sum_{t=1}^n (x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_t^j - \mathrm{E} x_t^j), i \le p - r, p - r + 1 \le j \le p\Big) \xrightarrow{p} \mathbf{0}. \tag{S.2}$$

Proof. For any $I(d_l)$ process x_t^l, we can write

$$\nabla^{d_l} x_t^l = \mathrm{E} z_t^l + (z_t^l - \mathrm{E} z_t^l) =: \mu_l + \zeta_t^l.$$

Let $U_t^l(0) = \zeta_t^l$, $V_t^l(0) = \mu_l$ and

$$U_t^l(j) = \sum_{s=1}^t U_s^l(j-1), \quad V_t^l(j) = \sum_{s=1}^t V_s^l(j-1).$$

Then

$$x_t^l = U_t^l(d_l) + V_t^l(d_l) = \sum_{j=1}^t U_j^l(d_l - 1) + \sum_{j=1}^t V_j^l(d_l - 1). \tag{S.3}$$

By induction, we have

$$V_t^l(d_l) = \mu_l \prod_{j=0}^{d_l-1} (t+j)/d_l! = \mu_l G_{d_l}(t). \tag{S.4}$$

On the other hand, since $\mathrm{E}\zeta_t^l = 0$, by (i) of Condition 1 and continuous mapping theorem, it follows that

$$U_{[ns]}^l(d_l)/n^{d_l-1/2} \overset{J_1}{\Longrightarrow} f_{d_l}^l(s), \text{ on } D[0,1]. \tag{S.5}$$

Thus, by (S.3)–(S.5),

$$(x_{[ns]}^l - \mu_l G_{d_l}([ns]))/n^{d_l-1/2} \overset{J_1}{\Longrightarrow} f_{d_l}^l(s), \text{ on } D[0,1]. \tag{S.6}$$

Since $S_n^i(t_i)$, $1 \le i \le p$ converge to their limiting distribution jointly, (S.1) follows from (S.6) and the continuous mapping theorem.

As for conclusion (S.2), by the joint convergence condition (see (i) of Condition 1) and (S.1),

$$\left(\frac{x_{[nt_i]}^i - \bar{x}^i - \mu_i L_{d_i}([nt_i])}{n^{d_i-1/2}}, \frac{1}{\sqrt{n}} \sum_{s=1}^{[nt_j]} (x_s^j - \mathrm{E}x_s^j),\, 1 \le i \le p-r,\, p-r+1 \le j \le p \right)$$
$$\overset{J_1}{\Longrightarrow} \left(F^i(t_i), W^j(t_j) \right)_{ij},$$

on $D[0,1] \times D[0,1]$. (ii) of Condition 1 implies that $\mathrm{E}|x_s^j|^2 < \infty$. This gives

$$\max_{1 \le s \le n} |x_s^j - \mathrm{E}x_s^j|/\sqrt{n} = o_p(1), \text{ and } \frac{1}{n} \sum_{s=1}^{n} |x_s^j - \mathrm{E}x_s^j| = O_p(1). \tag{S.7}$$

Thus, by Theorem 3.1 of Ling and Li (1998), we have

$$\frac{1}{n^{d_i+1/2}} \sum_{t=1}^{n} (x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_t^j - \mathrm{E}x_1^j) \xrightarrow{p} 0.$$

Since p is fixed, we have (S.2) as desired. $\square$

Next, we establish a bound for the eigenvalues of $\mathbf{\Sigma}_j^x$ and $\mathbf{A}'\widehat{\mathbf{\Sigma}}_j\mathbf{A} =: \widehat{\mathbf{\Sigma}}_j^x$.

Without loss of generality, we assume the first r_1 components of $\mathbf{x}_{t1}$ are $I(a_1)$, the next r_2 components are $I(a_2)$ and the last r_q components of $\mathbf{x}_{t1}$ are $I(a_q)$, that is,

$$\boldsymbol{x}_{t1} = (\overbrace{x_t^1, \cdots, x_t^{r_q}}^{I(a_q)},\ \overbrace{x_t^{r_q+1}, \cdots, x_t^{r_q+r_{q-1}}}^{I(a_{q-1})}, \cdots, \overbrace{x_t^{\sum_{j=2}^q r_j+1}, \cdots, x_t^{\sum_{j=1}^q r_j}}^{I(a_1)})',$$

where $a_1 < a_2 < \cdots < a_q$ are positive integers and $\sum_{i=1}^q r_i = p - r$. For $1 \le i \le q$, define $\nu_q = 0$ and $\nu_i = \sum_{j=i+1}^q r_j$. Then for any $\mathbf{x}_t(r_i) := (x_t^{\nu_i+1}, \cdots, x_t^{\nu_i+r_i})'$, if $\boldsymbol{\mu}_i := (\mu_{\nu_i+1}, \cdots, \mu_{\nu_i+r_i})' \ne 0$, there must exist a $r_i \times (r_i - 1)$ matrix $\mathbf{P}_i$ and $r_i \times 1$ vector $\bar{\boldsymbol{\mu}}_i$ such that $\mathbf{P}_i'\mathbf{P}_i = \mathbf{I}_{(r_i-1)}$, $(\mathbf{P}_i, \boldsymbol{\mu}_i)$ has full rank r_i, $\mathbf{P}_i'\boldsymbol{\mu}_i = 0$ and $\bar{\boldsymbol{\mu}}_i'\boldsymbol{\mu}_i = 1$, where $\mathbf{I}_a$ denotes $a \times a$ matrix. Let $\mathbf{B}_i = (\mathbf{P}_i, n^{-1/2}\bar{\boldsymbol{\mu}}_i)'$ if $\boldsymbol{\mu}_i \neq 0$ and $\mathbf{B}_i = \mathbf{I}_{r_i}$ if $\boldsymbol{\mu}_i = 0$, and $\boldsymbol{\Theta}_n = \text{diag}(\mathbf{B}_q, \cdots, \mathbf{B}_2, \mathbf{B}_1, \mathbf{I}_r)$. Define

$$\mathbf{D}_{n1} = \text{diag}\Big(\overbrace{n^{a_q-1/2}, \cdots, n^{a_q-1/2}}^{r_q}, \cdots, \overbrace{n^{a_1-1/2}, \cdots, n^{a_1-1/2}}^{r_1}\Big),\ \mathbf{D}_{n2} = (\overbrace{1, \cdots, 1}^{r}),$$

and $\mathbf{D}_n =: \text{diag}(\mathbf{D}_{n1}, \mathbf{D}_{n2})$. Let $H^d(t) = t^d/d! - 1/(d+1)!$, $F^i(t)$ be given as in Lemma 1, $\mathbf{F}_i(t) = (F^{\nu_i+1}(t), \cdots, F^{\nu_i+r_i}(t))'$, $\mathbf{M}_i(t) = (\mathbf{F}_i'(t)\mathbf{P}_i, H^{a_i}(t))'I(\boldsymbol{\mu}_i \neq 0) + \mathbf{F}_i(t)I(\boldsymbol{\mu}_i = 0)$, and $\mathbf{M}(t) = (\mathbf{M}_q'(t), \mathbf{M}_{q-1}'(t), \cdots, \mathbf{M}_1'(t))'$. Then Lemma 2 below follows from Lemma 1 and the continuous mapping theorem.

Lemma 2. *Let* $\boldsymbol{\Gamma}_j(x) = \textit{diag}\Big(\frac{1}{n}\sum_{t=1}^n(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)', \text{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}_{1,2})\Big)$. *Under Condition 1, we have*

$$\mathbf{D}_n^{-1}\boldsymbol{\Theta}_n\boldsymbol{\Gamma}_j^x\boldsymbol{\Theta}_n'\mathbf{D}_n^{-1} \xrightarrow{d} \text{diag}\Big(\int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt, \text{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}_{1,2})\Big).$$

Proof. For any $1 \le i \le l$, we define $\bar{\mathbf{x}}(s_i) = \frac{1}{n}\sum_{t=1}^n \mathbf{x}_t(s_i)$. When $\boldsymbol{\mu}_i = 0$, Lemma 1 gives

$$\frac{\mathbf{B}_i(\mathbf{x}_t(s_i) - \bar{\mathbf{x}}(s_i))}{n^{a_i-1/2}} = \frac{(\mathbf{x}_t(s_i) - \bar{\mathbf{x}}(s_i))}{n^{a_i-1/2}} \overset{J_1}{\Longrightarrow} \mathbf{F}_i(t), \quad \text{on} \quad \prod_{j=1}^{s_i} D[0,1]. \tag{S.8}$$

When $\boldsymbol{\mu}_i \neq 0$, by $\bar{\boldsymbol{\mu}}_i'\boldsymbol{\mu}_i = 1$, we have $\mathbf{B}_i\mathbf{x}_t(s_i) = ((\mathbf{P}_i'\mathbf{x}_t(s_i))', n^{-1/2}G_{d_i}(t) + n^{-1/2}\bar{\boldsymbol{\mu}}_i'\mathbf{x}_t(s_i))'$. Note that

$$\lim_{n\to\infty}\frac{1}{n^{a_i}}(G_{a_i}([nt]) + \bar{\boldsymbol{\mu}}_i'\mathbf{x}_{[nt]}(s_i)) \overset{p}{=} \lim_{n\to\infty}\frac{1}{n^{a_i}}G_{a_i}([nt]) = \lim_{n\to\infty}\frac{1}{n^{a_i}}\prod_{j=0}^{a_i-1}([nt]+j) = \frac{t^{a_i}}{a_i!}. \tag{S.9}$$

Thus, by $\mathbf{P}_i'\boldsymbol{\mu}_i = 0$, Lemma 1 and continuous mapping theorem, we have

$$\frac{\mathbf{B}_i(\mathbf{x}_t(s_i) - \bar{\mathbf{x}}(s_i))}{n^{a_i-1/2}} \overset{J_1}{\Longrightarrow} ((\mathbf{P}_i'\mathbf{F}_i(t))', H^{a_i}(t))' \quad \text{on} \quad \prod_{j=1}^{s_i} D[0,1]. \tag{S.10}$$

Combining (S.8), (S.10) and the joint convergence condition ((i) of Condition 1) yields

$$\left(\left(\frac{\mathbf{B}_i(\mathbf{x}_t(s_i) - \bar{\mathbf{x}}(s_i))}{n^{a_i-1/2}}\right)', 1 \le i \le l\right) \overset{J_1}{\Longrightarrow} (\mathbf{M}_i'(t), 1 \le i \le l) = \mathbf{M}'(t) \quad \text{on} \quad \prod_{j=1}^{p-r} D[0,1]. \tag{S.11}$$

Let $\mathbf{B} = \text{diag}(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_l) = (b_{ij})$. By (S.11) and continuous mapping theorem, we have

$$\begin{aligned}\mathbf{D}_{n1}^{-1}\mathbf{B}\left(\frac{1}{n}\sum_{t=1}^n(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)'\right)\mathbf{B}'\mathbf{D}_{n1}^{-1} &= \left(\frac{1}{n}\sum_{t=1}^n(\mathbf{D}_{n1}^{-1}\mathbf{B}(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1))(\mathbf{D}_{n1}^{-1}\mathbf{B}(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1))'\right) \\ &\xrightarrow{d} \int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt \end{aligned} \tag{S.12}$$

and complete the proof of Lemma 2. □

Let $F^i(t)$, $1 \le i \le p-r$ be defined in Lemma 1, where $W^i(t) = \sigma_{ii} B^i(t)$ and $B^i(t)$, $1 \le i \le p-r$ are independent Brownian motions. Let $\mathbf{F}(t) = (F^1(t), F^2(t), \cdots, F^{p-r}(t))'$.

Lemma 3. *Under condition 2 and $p = o(n^{1/2-\tau})$ with $0 < \tau < 1/2$,*

$$\left\| \mathbf{D}_n^{-1} \boldsymbol{\Gamma}_j^x \mathbf{D}_n^{-1} - \mathrm{diag}\Big(\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\, dt,\, \mathrm{Cov}(\mathbf{x}_{1+j,2},\, \mathbf{x}_{1,2}) \Big) \right\|_2 = o_p(1). \tag{S.13}$$

Further, $\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\, dt$ is positive definite.

Proof. First, we show (S.13). To this end, it is enough to show

$$\left\| \mathbf{D}_{n1}^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} (\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)' \right) \mathbf{D}_{n1}^{-1} - \int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\, dt \right\|_2 = o_p(1). \tag{S.14}$$

Let $\boldsymbol{\xi}_t = (\xi_t^1, \xi_t^2, \cdots, \xi_t^{p-r})'$ be an integrated process with components ξ_t^i satisfying

$$\nabla^{d_i} \xi_t^i = \sigma_{ii} v_t^i := \tilde{\varepsilon}_t^i.$$

For any given $1 \le i,\, j \le p - r$,

$$\begin{aligned}
& \frac{1}{n^{d_i+d_j}} \sum_{t=1}^{n} [(x_t^i - \bar{x}^i)(x_t^j - \bar{x}^j) - (\xi_t^i - \bar{\xi}^i)(\xi_t^j - \bar{\xi}^j)] \\
= \;& \frac{1}{n^{d_i+d_j}} \sum_{t=1}^{n} [x_t^i - \xi_t^i - (\bar{x}^i - \bar{\xi}^i)](x_t^j - \bar{x}^j) + \frac{1}{n^{d_i+d_j}} \sum_{t=1}^{n} (\xi_t^i - \bar{\xi}^i)[x_t^j - \xi_t^i - (\bar{x}^j - \bar{\xi}^j)] \\
=: \;& r_{ij}^1 + r_{ij}^2.
\end{aligned}$$

By induction, it is easy to show that under condition (10),

$$\sup_{1\le i\le p-r} \sup_{1\le t\le n} \mathrm{E}[(x_t^i - \xi_t^i)/n^{d_i-1/2}]^2 \le \sup_{1\le i\le p-r} \sup_{1\le t\le n} \frac{1}{n} \mathrm{E}(S_t^i - \sum_{l=1}^{t} \tilde{\varepsilon}_l^i)^2 = O(n^{2\tau-1}), \tag{S.15}$$

$$\sup_{1\le i\le p-r} \sup_{1\le t\le n} \mathrm{E}(\xi_t^i/n^{d_i-1/2})^2 = O(1) \quad \text{and} \quad \sup_{1\le i\le n} \sup_{1\le t\le n} \mathrm{E}(x_t^i/n^{d_i-1/2})^2 = O(1). \tag{S.16}$$

Thus, by equations (S.15), (S.16) and the independence of the components,

$$\sum_{i,j=1}^{p-r} [\mathrm{E}(r_{ij}^1)^2 + \mathrm{E}(r_{ij}^2)^2] = O(p^2 n^{2\tau-1}),$$

which implies

$$\left\| \mathbf{D}_{n1}^{-1} \left[\left(\frac{1}{n} \sum_{t=1}^{n} (\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)' \right) - \left(\frac{1}{n} \sum_{t=1}^{n} (\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})' \right) \right] \mathbf{D}_{n1}^{-1} \right\|_2 = O_p(pn^{\tau-1/2}),$$

where $\bar{\boldsymbol{\xi}} = (\bar{\xi}^1, \bar{\xi}^2, \cdots, \bar{\xi}^{p-r})'$. Thus, for the proof of (S.14), it suffices to show

$$\sup_{1\le i,j\le p-r} \left\| \frac{1}{n^{d_i+d_j}} \sum_{t=1}^{n} (\xi_t^i - \bar{\xi}^i)(\xi_t^j - \bar{\xi}^j) - \int_0^1 F^i(t)F^j(t)\, dt \right\|_2 = o_{a.s.}(n^{\tau-1/2}). \tag{S.17}$$

Note that

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\xi_t^i}{n^{d_i-1/2}}\right)\left(\frac{\xi_t^j}{n^{d_j-1/2}}\right)-\sum_{t=1}^{n}\int_{(t-1)/n}^{t/n} f^i(a)f^j(a)\,da \\
= & \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\xi_t^i}{n^{d_i-1/2}}-f^i(t/n)\right)\left(\frac{\xi_t^j}{n^{d_j-1/2}}\right)+\frac{1}{n}\sum_{t=1}^{n} f^i(t/n)\left(\frac{\xi_t^j}{n^{d_j-1/2}}-f^j(t/n)\right) \\
& -\sum_{t=1}^{n}\int_{(t-1)/n}^{t/n}(f^i(a)-f^i(t/n))f^j(a)+f^i(t/n)(f^j(a)-f^j(t/n))\,da \\
=: & J_{n1}(i,j)+J_{n2}(i,j)+J_{n3}(i,j).
\end{aligned}
$$

From the definition of $I(d)$ process, it is easy to deduce that if $\xi_t \sim I(d)$ satisfying $\nabla^d \xi_t = \varepsilon_t$, then $\xi_t = \sum_{s=1}^t \varepsilon_s$ when $d=1$ and when $d \ge 2$,

$$
\xi_t = \sum_{s=1}^{t}\left[\prod_{i=1}^{d-1}(t-s+i)/(d-1)!\right]\varepsilon_s \tag{S.18}
$$

and $f^i(t)$ can be rewritten as

$$
f^i(t) = \int_0^t (t-s)^{d_i-1}\,dW^i(s)/(d_i-1)!. \tag{S.19}
$$

By (S.19) and the continuity of $W^i(s)$, it is easy to get that

$$
\sup_{1\le i,j\le p-r}\ \sup_{(t-1)/n\le a\le t/n} |[f^i(a)-f^i(t/n)]f^j(a)| = O_{a.s.}(n^{-1/2}\log^2 n).
$$

Thus,

$$
\sup_{1\le i,j\le p-r} |J_{n3}(i,j)| = O_{a.s.}(n^{-1/2}\log^2 n). \tag{S.20}
$$

Set $\prod_{h=1}^{0}(t-s+h)/0! = 1$. Using expressions (S.18) and (S.19), we have

$$
\begin{aligned}
\xi_t^i/n^{d_i-1/2}-f^i(t/n) &= \sum_{s=1}^{t}\frac{\left[\prod_{h=1}^{d_i-1}(t-s+h)-(t-s)^{d_i-1}\right]}{n^{d_i-1/2}(d_i-1)!}\tilde{\varepsilon}_s^i \\
&\quad +\left(\sum_{s=1}^{t}\frac{(t-s)^{d_i-1}\tilde{\varepsilon}_s^i}{n^{d_i-1/2}(d_i-1)!}-\int_0^{t/n}\frac{(t/n-s)^{d_i-1}}{(d_i-1)!}\,dW^i(s)\right) \\
&=: H_{n1}^i(t)+H_{n2}^i(t).
\end{aligned}
$$

It is easy to get that

$$
\sup_{1\le i\le p-r}\ \sup_{1\le t\le n} |H_{n1}^i(t)| = O_{a.s.}(n^{-1/2}\log n). \tag{S.21}
$$

On the other hand, we have for any $1 \le t \le n$,

$$
\begin{aligned}
H_{n2}^i(t) &= \frac{1}{(d_i-1)!}\sum_{s=1}^{t}\left(\left(\frac{t}{n}-\frac{s}{n}\right)^{d_i-1}\frac{\tilde{\varepsilon}_s^i}{\sqrt{n}}-\int_{(s-1)/n}^{s/n}\left(\frac{t}{n}-a\right)^{d_i-1}dW^i(a)\right) \\
&=^d \frac{1}{(d_i-1)!}\sum_{s=1}^{t}\int_{(s-1)/n}^{s/n}\left[\left(\frac{t}{n}-\frac{s}{n}\right)^{d_i-1}-\left(\frac{t}{n}-a\right)^{d_i-1}\right]dW^i(a).
\end{aligned} \tag{S.22}
$$

This gives

$$\sup_{1\le i\le p-r}\sup_{1\le t\le n}|H_{n2}^i(t)| = nO_{a.s}(n^{-3/2}\log n) = O_{a.s.}(n^{-1/2}\log n). \tag{S.23}$$

Since the normal sequences $\{\nu_t^i\}$ are independent with respect to i, it follows that

$$\sup_{1\le j\le p-r}\sup_{1\le t\le n}|\xi_t^j| = O_{a.s.}(n^{d_j-1/2}\log n). \tag{S.24}$$

Thus, by (S.21), (S.23) and (S.24), we have

$$\sup_{1\le i,j\le p-r}|J_{n1}(i,j)| = O_{a.s.}(n^{-1/2}\log^2 n). \tag{S.25}$$

Similarly, we have

$$\sup_{1\le i,j\le p-r}|J_{n2}(i,j)| = O_{a.s.}(n^{-1/2}\log^2 n). \tag{S.26}$$

Using the same argument, we can show

$$\sup_{1\le i,j\le p-r}\left|(\bar{\xi}^i/n^{d_i-1/2})(\bar{\xi}^j/n^{d_j-1/2}) - \int_0^1 f^i(t)\,dt\int_0^1 f^i(t)\,dt\right| = O_{a.s.}(n^{-1/2}\log^2 n). \tag{S.27}$$

Combining equations (S.20), (S.25)–(S.27), we have (S.17) and conclude (S.13). The positive definiteness of $\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\,dt$ can be shown similarly to that of Lemma 3.1.1 in Chan and Wei (1988). □

Lemma 4. *Under Condition 1, or Condition 2 and $p = o(n^{1/2-\tau})$, we have*

$$\max_{0\le j\le j_0}\|\mathbf{D}_n^{-1}\boldsymbol{\Theta}_n(\boldsymbol{\Sigma}_j^x - \boldsymbol{\Gamma}_j^x)\boldsymbol{\Theta}_n'\mathbf{D}_n^{-1}\|_2 \xrightarrow{p} 0 \text{ and} \tag{S.28}$$

$$\max_{0\le j\le j_0}\|\mathbf{D}_n^{-1}\boldsymbol{\Theta}_n(\widehat{\boldsymbol{\Sigma}}_j^x - \boldsymbol{\Gamma}_j^x)\boldsymbol{\Theta}_n'\mathbf{D}_n^{-1}\|_2 \xrightarrow{p} 0. \tag{S.29}$$

Proof. We first consider the case with fixed p. To this end, we split the matrix into three parts: the nonstationary block, the cross block with elements being the product of stationary component with nonstationary component and the stationary block.

(I) As for the nonstationary block, we have for $1 \leq i, h \leq p-r$,

$$
\begin{aligned}
& \sum_{t=1}^{n-j}(x_{t+j}^i - \bar{x}^i)(x_t^h - \bar{x}^h) - \sum_{t=1}^{n}(x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) \\
= \;& -\sum_{t=1}^{j}(x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) - \sum_{t=1}^{n-j}(x_{t+j}^i - \bar{x}^i)(x_{t+j}^h - x_t^h) \\
= \;& -\sum_{t=1}^{j}(x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_t^h - \bar{x}^h - \mu_h L_{d_h}(t)) - \sum_{t=1}^{n-j}(x_{t+j}^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_{t+j}^h - x_t^h) \\
& -\mu_h \sum_{t=1}^{j} L_{d_h}(t)(x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)) - \mu_i \sum_{t=1}^{j} L_{d_i}(t)(x_t^h - \bar{x}^h - \mu_h L_{d_h}(t)) \\
& -\mu_i\mu_h \sum_{t=1}^{j} L_{d_h}(t) L_{d_i}(t) - \mu_i \sum_{t=1}^{n-j} L_{d_i}(t)(x_{t+j}^h - x_t^h) \\
=: \;& \sum_{m=1}^{6} \delta_{nm}(j, i, h).
\end{aligned}
$$

From (S.6), it follows that

$$
\begin{aligned}
\sup_{0\leq j\leq j_0} \frac{|\delta_{n1}(j,i,h)|}{n^{d_i+d_h}} &\leq \frac{j_0}{n}\left(\sup_{1\leq t\leq n}\frac{|x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)|}{n^{d_i-1/2}}\right)\left(\sup_{1\leq t\leq n}\frac{|x_t^h - \bar{x}^h - \mu_h L_{d_h}(t)|}{n^{d_h-1/2}}\right) \\
&= O_p(1/n).
\end{aligned} \tag{S.30}
$$

As for $\delta_{n2}(j, i, h)$, we have

(i) If $d_h = 1$, then $x_{t+j}^h - x_t^h = \sum_{i=t+1}^{t+j} \varepsilon_i^h$. Since $\mathrm{E}|\varepsilon_i^h| < \infty$, it follows that

$$
\begin{aligned}
\sup_{0\leq j\leq j_0} \frac{|\delta_{n2}(j,i,h)|}{n^{d_i+d_h}} &\leq \left(\sup_{t\leq n}\frac{|x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)|}{n^{d_i-1/2}}\right)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}\sum_{i=t+1}^{t+j_0}|\varepsilon_i^h|\right) \\
&= O_p(1/n^{1/2}).
\end{aligned} \tag{S.31}
$$

(ii) If $d_h \geq 2$, then $x_{t+j}^h - x_t^h = \sum_{s=t+1}^{t+j}[U_s^h(d_h-1) + V_s^h(d_h-1)]$ (see Lemma 1), it follows that

$$
\sup_{1\leq t\leq n}\sup_{1\leq j\leq j_0}\frac{|x_{t+j}^h - x_t^h|}{n^{d_h-1/2}} \leq \frac{j_0}{n^{1/2}}\sup_{1\leq s\leq n}\frac{|U_s^h(d_h-1) + V_s^h(d_h-1)|}{n^{d_h-1}} = O_p(1/n^{1/2}), \tag{S.32}
$$

which implies

$$
\begin{aligned}
\sup_{j\leq j_0} \frac{|\delta_{n2}(j,i,h)|}{n^{d_i+d_h}} &\leq \left(\sup_{t\leq n}\frac{|x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)|}{n^{d_i-1/2}}\right)\left(\sup_{t\leq n}\sup_{j\leq j_0}\frac{|x_{t+j}^h - x_t^h|}{n^{d_h-1/2}}\right) \\
&= O_p(1/n^{1/2}).
\end{aligned} \tag{S.33}
$$

Let $\mathbf{\Delta}_{nm}(j) = (\delta_{nm}(j,i,h))_{(p-r)\times(p-r)}$, $m = 1, 2, \cdots, 6$. Then by equations (S.30) – (S.33),

$$\begin{aligned}
&\sup_{1\le j\le j_0} ||\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}(\mathbf{\Delta}_{n1}(j)+\mathbf{\Delta}_{n2}(j))\mathbf{B}'\mathbf{D}_{n1}^{-1}||_2 \\
\le\ & \left(\sum_{i=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}|b_{il}b_{im}|\right) O_p(1/n^{1/2}) \\
\le\ & \left(\sum_{i=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}(b_{il}^2+b_{im}^2)\right) O_p(1/n^{1/2}) = O_p(p^2/n^{1/2}). \qquad \text{(S.34)}
\end{aligned}$$

By the definition of $\mathbf{B}$, it is easy to see that the elements of $\mathbf{B}\mathbf{\Delta}_{n3}(j)$, $\mathbf{B}\mathbf{\Delta}_{n4}(j)$ are zero except in rows $\sum_{i=1}^{j} s_i$, $j = 1, 2, \cdots, l$ and the non-zero elements have the following forms:

$$n^{-1/2}\sum_{t=1}^{j} L_{a_h}(t)(x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)).$$

Thus, by

$$\sup_{1\le j\le j_0}\left|\frac{\sum_{t=1}^{j} L_{d_h}(t)(x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))}{n^{d_h+d_i-1/2}}\right| = O_p(1),$$

we have

$$\sup_{1\le j\le j_0} ||\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}(\mathbf{\Delta}_{n3}(j)+\mathbf{\Delta}_{n4}(j))\mathbf{B}'\mathbf{D}_{n1}^{-1}||_2 \quad = \quad O_p(pn^{-1}). \qquad \text{(S.35)}$$

Similarly, by $\sup_{1\le j\le j_0}\left|\frac{\sum_{t=1}^{j} L_{d_h}(t)L_{d_i}(t)}{n^{d_i+d_h}}\right| = O_p(1)$, we have

$$\sup_{1\le j\le j_0} ||\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}\mathbf{\Delta}_{n5}(j)\mathbf{B}'\mathbf{D}_{n1}^{-1}||_2 = O_p(pn^{-1}). \qquad \text{(S.36)}$$

Further, using (S.9), similar to (S.31) and (S.32), we can show

$$\sup_{1\le j\le j_0}\frac{1}{n^{d_i+d_h+1/2}}\sum_{t=1}^{n-j} L_{d_i}(t)(x_{t+j}^h - x_t^h) = O_p(1/n^{1/2}).$$

Thus, for $\mathbf{\Delta}_{n6}(j)$ we have

$$\sup_{1\le j\le j_0} ||\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}\mathbf{\Delta}_{n6}(j)\mathbf{B}'\mathbf{D}_{n1}^{-1}||_2 = O_p(p/n^{1/2}). \qquad \text{(S.37)}$$

Combining equations (S.34), (S.35), (S.36) and (S.37) gives

$$\begin{aligned}
&\frac{\mathbf{D}_{n1}^{-1}\mathbf{B}}{n}\left|\left|\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)' - \sum_{t=1}^{n}(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)'\right|\right|_2 \mathbf{B}'\mathbf{D}_{n1}^{-1} \\
=\ & O_p(p^2/n^{1/2}). \qquad \text{(S.38)}
\end{aligned}$$

(II) As for the cross block, we first show

$$\begin{aligned}
&\left|\left|\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}\sum_{t=1}^{n}(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)'\right|\right|_2 + \left|\left|n^{-1}\sum_{t=1}^{n}(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)'\mathbf{B}'\mathbf{D}_{n1}^{-1}\right|\right|_2 \\
=\ & o_p(1). \qquad \text{(S.39)}
\end{aligned}$$

Note that for $1 \le i \le p-r$ and $p-r \le h \le p$,

$$\begin{aligned}\sum_{t=1}^{n}(x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) &= \sum_{t=1}^{n}(x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_t^h - \bar{x}^h) + \mu_i \sum_{t=1}^{n} L_{d_i}(t)(x_t^h - \bar{x}^h), \\ &=: \omega_{ih}^1 + \omega_{ih}^2. \qquad \text{(S.40)}\end{aligned}$$

Let $\boldsymbol{\Omega}_1 = (\omega_{ih}^1)_{(p-r)\times r}$ and $\boldsymbol{\Omega}_2 = (\omega_{ih}^2)_{(p-r)\times r}$. Then the elements of $\mathbf{B}\boldsymbol{\Omega}_1 = (e_{jh})$ have the following expression:

$$e_{jh} = \sum_{i=1}^{p-r} b_{ji} \sum_{t=1}^{n}(x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_t^h - \bar{x}^h) = \sum_{i=1}^{p-r} b_{ji}\omega_{ih}^1.$$

By Lemma 1, we have

$$\left|\frac{e_{jh}}{n^{d_i+1/2}}\right| \le \frac{1}{n^{d_i+1/2}} \sum_{i=1}^{p-r} |b_{ji}\omega_{ih}^1| = o_p(1). \qquad \text{(S.41)}$$

On the other hand, by the definition of $\mathbf{B}$, the elements of $\mathbf{B}\boldsymbol{\Omega}_2 = (d_{jh})$ can be represented as

$$d_{jh} = \frac{1}{n^{1/2}} \sum_{t=1}^{n} L_{a_i}(t)(x_t^h - \bar{x}^h) I(j = s_i),\ i = 1, 2, \cdots, l.$$

It is easy to get that

$$|d_{jh}|/n^{1/2+d_i} = o_p(1). \qquad \text{(S.42)}$$

Consequently, by (S.41) and (S.42), it follows that

$$\begin{aligned}\left|\left|\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}\sum_{t=1}^{n}(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)'\right|\right|_2 &\le \left|\left|\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}\boldsymbol{\Omega}_1\right|\right|_2 + \left|\left|\mathbf{D}_{n1}^{-1}\mathbf{B}n^{-1}\boldsymbol{\Omega}_2\right|\right|_2 \\ &= o_p(1). \qquad \text{(S.43)}\end{aligned}$$

Similarly,

$$\left|\left|n^{-1}\sum_{t=1}^{n}(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)'\mathbf{B}'\mathbf{D}_{n1}^{-1}\right|\right|_2 = o_p(1). \qquad \text{(S.44)}$$

(S.39) follows from (S.43) and (S.44).

Next, we show

$$\sup_{j\le j_0}\left|\left|\frac{\mathbf{D}_{n1}^{-1}\mathbf{B}}{n}\left(\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t,2} - \bar{\mathbf{x}}_2)' - \sum_{t=1}^{n}(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)'\right)\right|\right|_2 = o_p(1) \qquad \text{(S.45)}$$

and

$$\begin{aligned}&\sup_{j\le j_0}\left|\left|\frac{1}{n}\left(\sum_{t=1}^{n-j}(\mathbf{x}_{t,2} - \bar{\mathbf{x}}_2)(\mathbf{x}_{t+j,1} - \bar{\mathbf{x}}_1)' - \sum_{t=1}^{n}(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)'\right)\mathbf{B}'\mathbf{D}_{n1}^{-1}\right|\right|_2 \qquad \text{(S.46)} \\ &= o_p(1).\end{aligned}$$

As for (S.45), note that for any $1 \le i \le p-r$, $p-r+1 \le h \le p$,

$$\begin{aligned}
& \frac{1}{n^{d_i+1/2}} \left(\sum_{t=1}^{n-j} (x_{t+j}^i - \bar{x}^i)(x_t^h - \bar{x}^h) - \sum_{t=1}^{n} (x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) \right) \\
= \; & \frac{1}{n} \sum_{t=1}^{n-j} \left(\frac{x_{t+j}^i - x_t^i}{n^{d_i-1/2}} \right) (x_t^h - \mathrm{E}x_1^h) - \frac{(\bar{x}^h - \mathrm{E}x_1^h)}{n} \sum_{t=1}^{n-j} \left(\frac{x_{t+j}^i - x_t^i}{n^{d_i-1/2}} \right) \\
& - \frac{1}{n^{d_i+1/2}} \sum_{t=n-j+1}^{n} (x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) \\
= \; & L_{1n}(j,i,h) + L_{2n}(j,i,h) + L_{3n}(j,i,h). \qquad \text{(S.47)}
\end{aligned}$$

By (S.32) and $\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}|x_t^h| = O(1)$, it follows that when $d_i \ge 2$,

$$\sup_{0 \le j \le j_0} |L_{1n}(j,i,h)| = O_p(1/n^{1/2}). \qquad \text{(S.48)}$$

When $d_i = 1$, by $x_{t+j}^i - x_t^i = \sum_{s=t+1}^{t+j} \varepsilon_s^i$, we have

$$\mathrm{E} \sup_{0 \le j \le j_0} |L_{1n}(j,i,h)| \le \max_{1 \le t \le n} \frac{1}{n^{1/2}} \sum_{s=t+1}^{t+j_0} \mathrm{E}|\varepsilon_s^i (x_t^h - \mathrm{E}x_1^h)| = O(1/n^{1/2}).$$

Thus, (S.48) also holds for $d_i = 1$. Similar to $L_{1n}(j,i,h)$, we have

$$\sup_{1 \le j \le j_0} \left| \frac{1}{n} \sum_{t=1}^{n-j} \frac{x_{t+j}^i - x_t^i}{n^{d_i-1/2}} \right| = O_p(1/n^{1/2}).$$

This combining with Condition 1 show

$$\sup_{j \le j_0} |L_{2n}(j,i,h)| = n^{-1/2} \left(n^{1/2} |\bar{x}^h - \mathrm{E}x_1^h| \right) \left| \frac{1}{n} \sum_{t=1}^{n-j} \frac{x_{t+j}^i - x_t^i}{n^{d_i-1/2}} \right| = O_p(1/n). \qquad \text{(S.49)}$$

For $L_{3n}(j,i,h)$, by Lemma 1 and (S.9), we have

$$\sup_{1 \le t \le n} |x_t^i - \bar{x}^i| / n^{d_i} = O_p(1),$$

thus by $\sum_{t=n-j_0+1}^{n} \mathrm{E}|x_t^h|/n^{1/2} = O(1/n^{1/2})$, we have

$$\sup_{j \le j_0} |L_{3n}(j,i,h)| = O_p(1/n^{1/2}). \qquad \text{(S.50)}$$

Therefore, by (S.47)–(S.50),

$$\sup_{j \le j_0} \frac{1}{n^{d_i+1/2}} \left| \sum_{t=1}^{n-j} (x_{t+j}^i - \bar{x}^i)(x_t^h - \bar{x}^h) - \sum_{t=1}^{n} (x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) \right| = O_p(1/n^{1/2}), \qquad \text{(S.51)}$$

which shows (S.45).

For (S.46), note that for any $1 \le i \le p-r$, $p-r+1 \le h \le p$,

$$\begin{aligned} & \frac{1}{n^{d_i+1/2}} \left(\sum_{t=1}^{n-j} (x_t^i - \bar{x}^i)(x_{t+j}^h - \bar{x}^h) - \sum_{t=1}^{n} (x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) \right) \\ = & \frac{1}{n^{d_i+1/2}} \sum_{t=1}^{n-j} (x_{t+j}^h - x_t^h)(x_t^i - \bar{x}^i) + L_{3n}(j,i,h) =: L_{4n}(j,i,h) + L_{3n}(j,i,h). \end{aligned}$$

Let $\mathbf{L}(j) = (L_{4n}(j,i,h))'_{(p-r)\times r}$ and decompose $L_{4n}(j,i,h)$ into two terms as in (S.40). Using the same arguments as in (S.41) and (S.42), we can show

$$\sup_{j \le j_0} ||n^{-1}\mathbf{L}(j)\mathbf{B}'\mathbf{D}_{n1}^{-1}||_2 = o_p(1), \tag{S.52}$$

thus, by (S.50), we have (S.46). Combining equations (S.39) with (S.45) and (S.46) shows that the cross blocks tend to 0 in probability.

(III) As for the stationary block, let $\mathbf{\Upsilon}_j^x$ and $\widehat{\mathbf{\Upsilon}}_j^x$ be the matrixes obtained by replacing the stationary block $\frac{1}{n}\sum_{j=1}^{n-j}(\mathbf{x}_{t+j,2} - \bar{\mathbf{x}}_2)(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)'$ in $\mathbf{\Sigma}_j^x$ and $\widehat{\mathbf{\Sigma}}_j^x$ with $\mathrm{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}'_{1,2})$. By (ii) of Condition 1, we have

$$||\mathbf{\Sigma}_j^x - \mathbf{\Upsilon}_j^x||_2 = o_p(1) \text{ and } ||\widehat{\mathbf{\Sigma}}_j^x - \widehat{\mathbf{\Upsilon}}_j^x||_2 = o_p(1). \tag{S.53}$$

Thus, by (S.38) and the fact that the cross blocks tend to 0 in probability (see (II)), we have

$$\begin{aligned} ||\mathbf{D}_n^{-1}\mathbf{\Theta}_n(\mathbf{\Sigma}_j^x - \mathbf{\Gamma}_j^x)\mathbf{\Theta}_n'\mathbf{D}_n^{-1}||_2 &\le ||\mathbf{D}_n^{-1}\mathbf{\Theta}_n(\mathbf{\Sigma}_j^x - \mathbf{\Upsilon}_j^x)\mathbf{\Theta}_n'\mathbf{D}_n^{-1}||_2 + ||\mathbf{D}_n^{-1}\mathbf{\Theta}_n(\mathbf{\Upsilon}_j^x - \mathbf{\Gamma}_j^x)\mathbf{\Theta}_n'\mathbf{D}_n^{-1}||_2 \\ &= o_p(1) \end{aligned}$$

and

$$\begin{aligned} ||\mathbf{D}_n^{-1}\mathbf{\Theta}_n(\widehat{\mathbf{\Sigma}}_j^x - \mathbf{\Gamma}_j^x)\mathbf{\Theta}_n'\mathbf{D}_n^{-1}||_2 &\le ||\mathbf{D}_n^{-1}\mathbf{\Theta}_n(\widehat{\mathbf{\Sigma}}_j^x - \widehat{\mathbf{\Upsilon}}_j^x)\mathbf{\Theta}_n'\mathbf{D}_n^{-1}||_2 + ||\mathbf{D}_n^{-1}\mathbf{\Theta}_n(\widehat{\mathbf{\Upsilon}}_j^x - \mathbf{\Gamma}_j^x)\mathbf{\Theta}_n'\mathbf{D}_n^{-1}||_2 \\ &= o_p(1). \end{aligned}$$

Hence, Lemma 4 holds for finite p.

Next, consider the case: $p = o(n^{1/2-\tau})$. We still split the matrix into three parts as above.

(The nonstationary block.) Since $b_1 \le \lim_{n\to\infty} \mathrm{Var}(\sum_{s=1}^n z_s^i/\sqrt{n}) \equiv \sigma_{ii}^2 \le b_2$ for all i, it follows that as $n \to \infty$,

$$\max_{1\le t\le n} \mathrm{Var}\Big(\sum_{s=1}^t z_s^i/\sqrt{n}\Big) \le b_2 \quad \text{and} \quad \max_{1\le t\le n} \mathrm{Var}\Big(x_t^i/n^{d_i-1/2}\Big) \le b_2. \tag{S.54}$$

Let $\delta_{n1}(j,i,h), \delta_{n2}(j,i,h)$ be defined as above with $\mu_i = \mu_h = 0$. Note that the components of $\{\mathbf{z}_t\}$ are independent, by (S.54) and some elementary computation, we can show

$$\mathrm{E}\left[\sum_{i,h=1}^{p-r}\left(\sup_{j\le j_0}\frac{|\delta_{n1}(j,i,h)|}{n^{d_i+d_h}}\right)^2\right] = O(j_0^2p^2/n^2)$$

and

$$\mathrm{E}\left[\sum_{i,h=1}^{p-r}\left(\sup_{j\le j_0}\frac{|\delta_{n2}(j,i,h)|}{n^{d_i+d_h}}\right)^2\right]=O(j_0^2p^2/n).$$

Combining the above two equations yields

$$\begin{aligned}&\mathbf{D}_{n1}^{-1}n^{-1}\left\|\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)'-\sum_{t=1}^{n}(\mathbf{x}_{t,1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)'\right\|_2\mathbf{D}_{n1}^{-1}\\ =\;& O_p(pn^{-\frac{1}{2}}).\end{aligned}\tag{S.55}$$

(The cross block.) Let ω_{ih} be defined as in (S.40) with $\mu_i=0$. Since $\mathbf{z}_t$ and $\mathbf{x}_{t2}$ are independent, it follows from (S.54) that

$$\begin{aligned}\mathrm{E}\Big[\sum_{i=1}^{p-r}\sum_{h=p-r+1}^{p}\omega_{ih}^2\Big] &= \sum_{i=1}^{p-r}\sum_{h=p-r+1}^{p}\mathrm{E}(\omega_{ih}^2)\\ &= \sum_{i=1}^{p-r}\sum_{h=p-r+1}^{p}n^{-2}\sum_{t,t'=1}^{n}\mathrm{E}\Big[\frac{(x_t^i-\bar{x}^i)}{n^{d_i-1/2}}\frac{(x_{t'}^i-\bar{x}^i)}{n^{d_i-1/2}}\Big]\mathrm{E}[(x_t^h-\bar{x}^h)][(x_{t'}^h-\bar{x}^h)]\\ &= O\left(\sum_{i=1}^{p-r}\sum_{h=p-r+1}^{p}n^{-2}\sum_{t,t'=1}^{n}|\mathrm{E}[(x_t^h-\bar{x}^h)][(x_{t'}^h-\bar{x}^h)]|\right)=O(p^2/n^{1-2\tau})\end{aligned}$$

by (iii) of Condition 2. Thus,

$$\left\|\mathbf{D}_{n1}^{-1}n^{-1}\sum_{t=1}^{n}(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)'\right\|_2=\left\|\mathbf{D}_{n1}^{-1}n^{-1}\mathbf{\Omega}_1\right\|_2=O_p(pn^{-1/2+\tau}).\tag{S.56}$$

Similarly,

$$\left\|n^{-1}\sum_{t=1}^{n}(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)'\mathbf{D}_{n1}^{-1}\right\|_2=O_p(pn^{-1/2+\tau}).\tag{S.57}$$

Further, by some elementary computation, it is easy to show

$$\mathrm{E}\Big[\sum_{i=1}^{p-r}\sum_{j=p-r+1}^{p}\sum_{l=1}^{3}(L_{ln}(j,i,h))^2\Big]=O(pr/n^{1-2\tau}),$$

which gives

$$\begin{aligned}&\sup_{j\le j_0}\left\|\frac{\mathbf{D}_{n1}^{-1}}{n}\left(\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t+j,2}-\bar{\mathbf{x}}_2)'-\sum_{t=1}^{n}(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)'\right)\right\|_2\\ =\;& O_p(pn^{-\frac{1}{2}+\tau}).\end{aligned}\tag{S.58}$$

Note that $L_{4n}(j,i,h)=\sum_{t=n-j+1}^{n}(x_t^i-\bar{x}^i)x_t^h-\sum_{t=1}^{j}(x_t^i-\bar{x}^i)x_t^h-\sum_{t=1}^{n-j}(x_{t+j}^i-x_t^i)x_{t+j}^h$, it is easy to show that $\mathrm{E}\Big[\sum_{i=1}^{p-r}\sum_{j=p-r+1}^{p}(L_{4n}(j,i,h))^2\Big]=O(pr/n^{1-2\tau})$ too. As a result, we have

$$\begin{aligned}&\sup_{j\le j_0}\left\|\frac{1}{n}\left(\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2}-\bar{\mathbf{x}}_2)(\mathbf{x}_{t+j,1}-\bar{\mathbf{x}}_1)'-\sum_{t=1}^{n}(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)(\mathbf{x}_{t1}-\bar{\mathbf{x}}_1)'\right)\mathbf{D}_{n1}^{-1}\right\|_2\\ =\;& O_p(pn^{-\frac{1}{2}+\tau}).\end{aligned}\tag{S.59}$$

Consequently, by equations (S.56)–(S.59), we get that the norms of the cross blocks are $O_p(pn^{-\frac{1}{2}+\tau})$.

(THE STATIONARY BLOCK.) By (ii) of Condition 2, we also have (S.53). Thus, by (S.55) and the bound of the cross blocks (see above), we have if $p = o(n^{1/2-\tau})$,

$$\begin{aligned}||\mathbf{D}_n^{-1}(\mathbf{\Sigma}_j^x - \mathbf{\Gamma}_j^x)\mathbf{D}_n^{-1}||_2 &\leq ||\mathbf{D}_n^{-1}(\mathbf{\Sigma}_j^x - \mathbf{\Upsilon}_j^x)\mathbf{D}_n^{-1}||_2 + ||\mathbf{D}_n^{-1}(\mathbf{\Upsilon}_j^x - \mathbf{\Gamma}_j^x)\mathbf{D}_n^{-1}||_2 \\ &= o_p(1) + O_p(pn^{-1/2}) = o_p(1)\end{aligned}$$

and

$$||\mathbf{D}_n^{-1}(\widehat{\mathbf{\Sigma}}_j^x - \mathbf{\Gamma}_j^x)\mathbf{D}_n^{-1}||_2 \leq ||\mathbf{D}_n^{-1}(\widehat{\mathbf{\Sigma}}_j^x - \widehat{\mathbf{\Upsilon}}_j^x)\mathbf{D}_n^{-1}||_2 + ||\mathbf{D}_n^{-1}(\widehat{\mathbf{\Upsilon}}_j^x - \mathbf{\Gamma}_j^x)\mathbf{D}_n^{-1}||_2 = o_p(1).$$

Hence, Lemma 4 holds also for large p with $p = o(n^{1/2-\tau})$. □

Proof of Theorem 3.1. Since

$$\begin{aligned}\{D(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{A}_2))\}^2 &= \frac{1}{r}\{\text{tr}[\mathbf{A}_2'(I_p - \widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2')\mathbf{A}_2]\} \\ &\leq ||\mathbf{A}_2'(\mathbf{A}_2\mathbf{A}_2' - \widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2')\mathbf{A}_2||_2 \leq 2||\widehat{\mathbf{A}}_2 - \mathbf{A}_2||_2^2,\end{aligned}$$

it follows from Theorem I.5.5 of Stewart and Sun (1990) (see also Proposition 2.1 of Vu and Lei (2013)) that

$$D(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{A}_2)) \leq \sqrt{2}||\widehat{\mathbf{A}}_2 - \mathbf{A}_2||_2 \leq \sqrt{2}||\widehat{\mathbf{A}}_2 - \mathbf{A}_2||_F \leq 2\sqrt{2}||\sin\Theta(\widehat{\mathbf{A}}_2, \mathbf{A}_2)||_F, \qquad \text{(S.60)}$$

where $\Theta(\widehat{\mathbf{A}}_2, \mathbf{A}_2) = \arccos[(\mathbf{A}_2'\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2\mathbf{A}_2)^{1/2}]$ is the canonical angle between the column spaces of $\widehat{\mathbf{A}}_2$ and $\mathbf{A}_2$. Let $\eta = \min_{\lambda\in\lambda(\mathbf{D}_1^x),\, \mu\in\lambda(\widetilde{\mathbf{D}}_2^x)} |\lambda - \mu|/\sqrt{\lambda\mu}$, where $\lambda(\widetilde{\mathbf{D}}_2^x)$ consists of the r smallest eigenvalues of $\mathbf{A}'\widehat{\mathbf{W}}\mathbf{A} =: \widehat{\mathbf{W}}^x$. By Theorem 2.4 of Dopico, Moro and Molera (2000), we have

$$||\sin\Theta(\widehat{\mathbf{A}}_2, \mathbf{A}_2)||_F \leq ||(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2}||_F/\eta. \qquad \text{(S.61)}$$

Note that

$$(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2} = (\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2} - (\mathbf{W}^y)^{1/2}(\widehat{\mathbf{W}})^{-1/2}. \qquad \text{(S.62)}$$

Thus, by equations (S.60), (S.61) and (S.62), we have

$$D(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{A}_2)) \leq (||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F + ||(\mathbf{W}^y)^{1/2}(\widehat{\mathbf{W}})^{-1/2}||_F)/\eta.$$

Next, we show that$||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F = O_p(1)$, which is equivalent to

$$||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F = O_p(1). \qquad \text{(S.63)}$$

Note that

$$0 \leq \mathbf{\Sigma}_0^x \leq (\mathbf{W}^x)^{1/2} \leq \sum_{j=0}^{j_0}\{\mathbf{\Sigma}_j^x(\mathbf{\Sigma}_j^x)'\}^{1/2} \text{ and } 0 \leq \widehat{\mathbf{\Sigma}}_0^x \leq (\widehat{\mathbf{W}}^x)^{1/2} \leq \sum_{j=0}^{j_0}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}. \qquad \text{(S.64)}$$

It follows from (S.64) that

$$||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F \leq \sum_{j=0}^{j_0} ||(\mathbf{\Sigma}_0^x)^{-1}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}||_F.$$

Thus, for (S.63), it is enough to show the eigenvalues of $(\mathbf{\Sigma}_0^x)^{-1}\sum_{j=0}^{j_0}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}$ are $O_p(1)$, which is equivalent to

$$\text{the solutions } \lambda \text{ of } |\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2} - \lambda\mathbf{\Sigma}_0^x| = 0 \text{ are } O_p(1). \tag{S.65}$$

Since $\operatorname{diag}\left(\int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt,\ \operatorname{Var}(\mathbf{x}_{1,2})\right) > 0$, by Lemma 4 the solutions (λ) of equation

$$|\mathbf{D}_n^{-1}\mathbf{\Theta}_n\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}\mathbf{\Theta}_n'\mathbf{D}_n^{-1} - \lambda\mathbf{D}_n^{-1}\mathbf{\Theta}_n\mathbf{\Sigma}_0^x\mathbf{\Theta}_n'\mathbf{D}_n^{-1}| = 0 \tag{S.66}$$

are bounded in probability. Thus, we have (S.65) and (S.63) as desired.

Similarly, we can show

$$||(\mathbf{W}^y)^{1/2}(\widehat{\mathbf{W}})^{-1/2}||_F = ||(\mathbf{W}^x)^{1/2}(\widehat{\mathbf{W}}^x)^{-1/2}||_F = O_p(1). \tag{S.67}$$

Using equations (S.64) and (S.67), the remainder of the proof of Theorem 3.1 consists of showing that there exist two positive constants c_1, c_2 such that in probability $\eta \geq c_1 n^{2a_1-1}/\sqrt{j_0}$ provided $|I_0| \geq 2$ or $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$ and $\eta \geq c_2 n^{2a_1}/\sqrt{j_0}$ provided $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} \neq 0$.

Define $\lambda_i(\mathbf{A})$ to be the i-th eigenvalue of a matrix $\mathbf{A}$. Note that

$$\operatorname{diag}\left(\int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt,\ \operatorname{Var}(\mathbf{x}_{1,2})\right) > 0.$$

By Lemmas 2 and 4, it follows that when $|I_0| \geq 2$ or $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$, $\lambda_{p-r}(\mathbf{\Sigma}_j^x) = O_e(n^{2a_1-1})$ and $\lambda_{p-r+1}(\widehat{\mathbf{\Sigma}}_j^x) = O_e(1)$. Thus, there exist two positive constants c_3, c_4 such that in probability

$$\lambda_{p-r}(\mathbf{W}^x) \geq \lambda_{p-r}(\mathbf{\Sigma}_0^x(\mathbf{\Sigma}_0^x)') \geq c_3 n^{2(2a_1-1)} \tag{S.68}$$

and

$$c_3 \leq \lambda_{p-r+1}(\widehat{\mathbf{\Sigma}}_0^x(\widehat{\mathbf{\Sigma}}_0^x)') \leq \lambda_{p-r+1}(\widehat{\mathbf{W}}^x) \leq \left[\lambda_{p-r+1}\Big(\sum_{j=0}^{j_0}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}\Big)\right]^2 \leq c_4 j_0^2. \tag{S.69}$$

Hence, in probability

$$\eta \geq |c_3 n^{2(2a_1-1)} - c_4 j_0^2|/\sqrt{c_3 n^{2(2a_1-1)} c_4 j_0^2} \geq c' n^{2a_1-1}/j_0.$$

Similarly, we have $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} \neq 0$, then in probability,

$$\eta \geq c' n^{2a_1}/j_0. \tag{S.70}$$

Since j_0 is fixed, combining (S.63), (S.70) and (S.70), we complete the proof of (i) and (ii). Conclusion (iii) can be shown similarly by treating A_{1i} as the role of A_2, see also the proof of Theorem 1 of Chen and Hurvich (2006), we omit the details here. □

Let $\mathbf{A}_{1,0} = \mathbf{A}_2$ and $\widehat{\mathbf{B}}_{1i} = (\widehat{\gamma}_{\nu_i+1}, \cdots, \widehat{\gamma}_{\nu_i+r_i})$ for $i = 1, \cdots, q$ and $\widehat{\mathbf{B}}_{10} = (\widehat{\gamma}_{p-r+1}, \cdots, \widehat{\gamma}_p)$.

Lemma 5. *Under Condition 1, we have*

$$\|\mathbf{B}_{1,l}\mathbf{A}_{1,h}\|_F = O_p(n^{-2|a_h - a_l|}), \ \textit{for } l \neq h.$$

Proof. Let $\eta(\mathbf{B}_{1,l}, \mathbf{A}_{1,h})$ be defined as η above, i.e.,

$$\eta(\mathbf{B}_{1,l}, \mathbf{A}_{1,h}) = \min_{\lambda \in \{\widehat{\lambda}_{\nu_l+1}, \cdots, \widehat{\gamma}_{\nu_l+r_l}\}, \mu \in \{\lambda_{\nu_h+1}, \cdots, \lambda_{\nu_h+r_h}\}} |\lambda - \mu| / \sqrt{\lambda\mu}.$$

By Lemmas 2 and 4, using the same arguments as in Theorem 3.1, we have

$$\eta(\mathbf{B}_{1,l}, \mathbf{A}_{1,h}) \geq cn^{2|a_h - a_l|} \tag{S.71}$$

for some $c > 0$. It has been shown in Theorem 1 that $||(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2}||_F = O_p(1)$, thus by Theorem 2.4 of Dopico, Moro and Molera (2000) (see also Theorem 4.1 of Barlow and Slapničar (2000)), we have

$$\begin{aligned} \|\mathbf{B}_{1,l}\mathbf{A}_{1,h}\|_F &\leq ||(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2}||_F / \eta(\mathbf{B}_l, \mathbf{A}_h) \\ &= O_p(n^{-2|a_h - a_l|}). \end{aligned}$$

This completes the proof of Lemma 5. □

Proof of Theorem 3.2. First, we prove the consistency of $\widehat{r}$. For any $1 \leq i \leq p$,

$$\widehat{x}_t^i = \widehat{\gamma}_i'\mathbf{y}_t = (\widehat{\gamma}_i'\mathbf{A}_{1q}\mathbf{x}_{t1q}, \cdots, \widehat{\gamma}_i'\mathbf{A}_{11}\mathbf{x}_{t11}, \widehat{\gamma}_i'\mathbf{A}_2\mathbf{x}_{t2}). \tag{S.72}$$

Let ν_i be defined as that after Lemma 1 and $r_0 = r$. By Lemma 5, when $\nu_l + 1 \leq i \leq \nu_l + r_l$, $l \neq h$,

$$\widehat{\gamma}_i'\mathbf{A}_{1h} = O_p(n^{-2|a_h - a_l|}).$$

Thus, by $\sup_{1 \leq t \leq n} |\mathbf{x}_{t1h}| = O_p(n^{a_h - 1/2})$ for $h \geq 1$ (see Lemma 1), we have

$$\widehat{\gamma}_i'\mathbf{A}_{1h}\mathbf{x}_{t1h} = O_p(n^{-a_h + 2a_l - 1/2})I(h > l) + O_p(n^{-2a_l + 3a_h - 1/2})I(1 \leq h < l).$$

As a result, by (S.72), it follows that for any $\nu_l + 1 \leq i \leq \nu_l + r_l$,

$$\begin{aligned} \widehat{x}_t^i &= \widehat{\gamma}_i'\mathbf{A}_{1l}\mathbf{x}_{t1l} + O_p(\sum_{h=l+1}^{q} n^{-a_h + 2a_l - 1/2} + \sum_{h=1}^{l} n^{-2a_l + 3a_{l-1} - 1/2}) \\ &= \widehat{\gamma}_i'\mathbf{A}_{1l}\mathbf{x}_{t1l} + O_p(n^{-a_{l+1} + 2a_l - 1/2} + n^{-2a_l + 3a_{l-1} - 1/2}), \end{aligned}$$

where $\mathbf{x}_{t10} = \mathbf{x}_{t2}$. Thus, for any given m, we have

$$\begin{aligned} &\sum_{k=1}^{m} \Big(\frac{1}{n-k} \sum_{t=1}^{n-k} (\widehat{x}_{t+k}^i - \overline{\widehat{x}}^i)(\widehat{x}_{t,i} - \overline{\widehat{x}}^i) \Big) \\ = \ &\frac{\widehat{\gamma}_i'\mathbf{A}_{1l}}{n-k} \sum_{k=1}^{m} \sum_{t=1}^{n-k} (\mathbf{x}_{t+k,1l} - \overline{\mathbf{x}}_{1l})(\mathbf{x}_{t1l} - \overline{\mathbf{x}}_{1l})'\mathbf{A}_{1l}'\widehat{\gamma}_i(1 + o_p(1)). \end{aligned} \tag{S.73}$$

By (S.73), we have that for any $\nu_l+1\le i\le \nu_l+r_l$, $l=1,\cdots,q$

$$\begin{aligned}
&\sum_{k=1}^{m}\Big(\frac{1}{n-k}\sum_{t=1}^{n-k}(\widehat{x}^i_{t+k}-\overline{\widehat{x}}^i)(\widehat{x}_{t,i}-\overline{\widehat{x}}^i)\Big)\\
= \;& m\widehat{\gamma}_i'\mathbf{A}_{1l}\Big(\frac{1}{n}\sum_{t=1}^{n}(\mathbf{x}_{t1l}-\overline{\mathbf{x}}_{1l})(\mathbf{x}_{t1l}-\overline{\mathbf{x}}_{1l})'\Big)\mathbf{A}_{1l}'\widehat{\gamma}_i(1+o_p(1))=O_e(mn^{2a_l-1}).
\end{aligned}\tag{S.74}$$

On the other hand, by (S.73) and $\|\sum_{k=1}^m\frac{1}{n-k}\sum_{t=1}^{n-k}(\mathbf{x}_{t+k,2}-\overline{\mathbf{x}}_2)(\mathbf{x}_{t,2}'-\overline{\mathbf{x}}_2)\|\le C$ in probability, it follows that for $p-r+1\le i\le p$,

$$\sum_{k=1}^{m}\Big(\frac{1}{n-k}\sum_{t=1}^{n-k}(\widehat{x}^i_{t+k}-\overline{\widehat{x}}^i)(\widehat{x}_{t,i}-\overline{\widehat{x}}^i)\Big)=O_p(1).\tag{S.75}$$

Equation (S.74) together with (S.75) yields the conclusion of Theorem 3.2 as desired. □

S.2 Proofs for Section 3.2

Proof of Theorems 3.3 and 3.4. Theorem 3.3 can be shown similarly to Theorem 3.1 by using Lemma 3 instead of Lemma 2, except that when $p\to\infty$,

$$||(\mathbf{\Sigma}_0^x)^{-1}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}||_F=O_p\left(\left(\sum_{i=1}^{p}(\widetilde{\lambda}_i)^2\right)^{1/2}\right)=O_p(p^{1/2}),$$

where $\widetilde{\lambda}_i, 1\le i\le p$ are solutions of (S.65). As a result, (S.63) should be replaced by

$$||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F=O_p(p^{1/2})\text{ and }||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F=O_p(p^{1/2}).\tag{S.76}$$

Theorem 3.4 can be shown similarly to Theorem 3.2. We omit the details. □

S.3 Proofs for Section 4

To prove Theorems 4.1 and 4.2, we first introduce some notations. Let $k_{ni}=n^{d_i-1/2}I(d_i>1/2)+n^{d_i+1/2}I(d_i<1/2)$ and $\lambda_i(t-s)=(t-s)^{d_i-1}/\Gamma(d_i)I(d_i>1/2)+(t-s)^{d_i}/\Gamma(d_i+1)I(d_i<1/2)$. Define $\mathbf{K}_n=\mathrm{diag}(k_{n1},\cdots,k_{np})$, $\mathbf{\Lambda}(t,s)=\mathrm{diag}(\lambda_1(t-s),\cdots,\lambda_p(t-s))$ and

$$\mathbf{B}_0=0,\ \mathbf{B}_t=(B_t^1,\cdots,B_t^p)'=\int_0^t\mathbf{\Lambda}(t,s)\,d\mathbf{W}_s,\ \mathbf{U}_t=\mathbf{B}_t-\int_0^1\mathbf{B}_t\,dt,$$

where $\mathbf{W}_s$ is given in (ii) of Condition 3. Let $\nabla^{d_l}v_t^l=\mu_l$, $I_1^c=\{i:d_i>1/2\}$ and $\mathbf{x}_{t,I}=(x_t^i,\,i\in I)'$ and $\mathbf{v}_{t,I}=(v_t^i,\,i\in I)'$

Lemma 6. *Let $\mathbf{Z}_n(t)=((\mathbf{x}_{[nt],I_1^c}-\mathbf{v}_{[nt],I_1^c})',\sum_{j=1}^{[nt]}(\mathbf{x}_{j,I_1}-\mathbf{v}_{j,I_1})')'$. Under (ii) of Condition 3,*

$$\mathbf{K}_n^{-1}\mathbf{Z}_n(t)\overset{J_1}{\Longrightarrow}\mathbf{B}_t,\ \text{on } D[0,1]^p.\tag{S.77}$$

Proof. Let $d_{I_1} = \{d_i : i \in I_1\}$, then $\sum_{j=1}^{[nt]} \mathbf{x}_{j,I_1}$ is an integrated fractional process with order $d_{I_1}+1$, and each of its components has order larger than $1/2$. We can show this lemma similarly to Theorem 1 of Marinucci and Robinson (2000) by replacing their Lemma 2 with Condition 3(ii). $\square$

Let $\boldsymbol{\Theta}_n$ and $\mathbf{M}_i(t)$ be defined as that after Lemma 1 by using $H^d(t) = t^d/\Gamma(d_i+1) - 1/\Gamma(d_i+2)$ and $\mathbf{F}_i(t) = (U^{\nu_i+1}(t), \cdots, U^{\nu_i+r_i}(t))'$, where U_t^i be the i-th component of $\mathbf{U}_t$. Let $\mathbf{L}_n = \mathrm{diag}(l_{n1}, \cdots, l_{np})$, $l_{ni} = n^{d_i-1/2} I(d_i > 1/2) + I(d_i < 1/2)$. Similar to Lemma 2, by Lemma 6 and the continuous mapping theorem, we have the following lemma.

Lemma 7. *Let the conditions of Theorem 5 hold. Then the following assertions hold for any* $0 \le j \le j_0$.

(i) If $\delta > 1/2$, *then*

$$\mathbf{L}_n^{-1}\boldsymbol{\Theta}_n\widehat{\boldsymbol{\Sigma}}_j^x\boldsymbol{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \int_0^1 (\mathbf{M}_t', \mathbf{U}_{t2}')'(\mathbf{M}_t', \mathbf{U}_{t2}')\,dt, \quad \textit{and}$$

$$\mathbf{L}_n^{-1}\boldsymbol{\Theta}_n\boldsymbol{\Sigma}_j^x\boldsymbol{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \mathrm{diag}\Big(\int_0^1 \mathbf{M}_t\mathbf{M}_t'\,dt,\ \int_0^1 \mathbf{U}_{t,2}\mathbf{U}_{t,2}'\,dt\Big),$$

where $\mathbf{U}_{t,2}$ *is corresponding to the last* p *components of* $\mathbf{U}_t$.

(ii) If $\delta < 1/2$, *then*

$$\mathbf{L}_n^{-1}\boldsymbol{\Theta}_n\widehat{\boldsymbol{\Sigma}}_j^x\boldsymbol{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \mathrm{diag}\left(\int_0^1 \mathbf{M}_t\mathbf{M}_t'\,dt,\ \mathrm{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1})\right), \tag{S.78}$$

and

$$\mathbf{L}_n^{-1}\boldsymbol{\Theta}_n\boldsymbol{\Sigma}_j^x\boldsymbol{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \mathrm{diag}\left(\int_0^1 \mathbf{M}_t\mathbf{M}_t'\,dt,\ \mathrm{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1})\right). \tag{S.79}$$

Proof. We only give the proof for $\mu_j = 0$, $j = 1, \cdots, p$ in details, other cases can be proved similarly. By Lemma 6 and the continuous mapping theorem, it follows that (i) holds for $j = 0$. Thus, for the proof of (i), it suffices to show for any $1 \le i, h \le p$,

$$\sup_{1\le j\le j_0}\left|\frac{1}{n^{d_i+d_h}}\left(\sum_{t=1}^{n-j}(x_{t+j}^i - \bar{x}^i)(x_t^h - \bar{x}^h) - \sum_{t=1}^{n}(x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h)\right)\right| = o_p(1). \tag{S.80}$$

Observe that

$$\begin{aligned}
&\sum_{t=1}^{n-j}(x_{t+j}^i - \bar{x}^i)(x_t^h - \bar{x}^h) - \sum_{t=1}^{n}(x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h)\\
=\ &\sum_{t=1}^{n-j}(x_{t+j}^i - x_t^i)(x_t^h - \bar{x}^h) - \sum_{t=n-j+1}^{n}(x_t^i - \bar{x}^i)(x_t^h - \bar{x}^h) =: \alpha_{n1}(j,i,h) + \alpha_{n2}(j,i,h).
\end{aligned}$$

By Lemma 5, it follows that for any $1 \leq i, h \leq p$,

$$\Big((x^i_{[nt]} - \bar{x}^i)/n^{d_i - 1/2},\, (x^h_{[ns]} - \bar{x}^h)/n^{d_h - 1/2}\Big) \overset{J_1}{\Longrightarrow} (U^i(t), U^h(s)) \text{ on } D[0,1]^2. \tag{S.81}$$

This gives

$$\sup_{0 \leq j \leq j_0} |\alpha_{n2}(j,i,h)|/n^{d_i + d_h} = O_p(1/n). \tag{S.82}$$

Further, for any $\varepsilon > 0$, then

$$\lim_{n\to\infty} P\{\sup_{1\leq j\leq j_0} \sup_{1\leq t\leq n-j} |x^i_{t+j} - x^i_t|/n^{d_i - 1/2} > \varepsilon\} = 0. \tag{S.83}$$

Thus,

$$\sup_{0 \leq j \leq j_0} |\alpha_{n1}(j,i,h)|/n^{d_i + d_h} = o_p(1). \tag{S.84}$$

Combining (S.82) and (S.84) gives (S.80) as desired.

By (i) of Condition 3, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}(\mathbf{x}_{t+j,I_1} - \bar{\mathbf{x}}_{I_1})(\mathbf{x}_{t,I_1} - \bar{\mathbf{x}}_{I_1}) \xrightarrow{p} \mathrm{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1}).$$

Thus, by (i) of Lemma 6, we have (S.79).

As for (S.78), it is enough to show for any $i \in I_1^c$ and $h \in I_1$,

$$\begin{aligned}\frac{1}{n^{d_i+1/2}}\sum_{i=1}^{n-j}(x^i_{t+j} - \bar{x}^i)(x^h_t - \bar{x}^h) &= \frac{\alpha_{n1}(j,i,h)}{n^{d_i+1/2}} + \frac{\alpha_{n2}(j,i,h)}{n^{d_i+1/2}} + \frac{1}{n^{d_i+1/2}}\sum_{i=1}^{n}(x^i_t - \bar{x}^i)(x^h_t - \bar{x}^h)\\ &\xrightarrow{p} 0, \end{aligned} \tag{S.85}$$

holds for all $0 \leq j \leq j_0$. Similar to Lemma 1, we can show

$$\frac{1}{n^{d_i+1/2}}\sum_{i=1}^{n}(x^i_t - \bar{x}^i)(x^h_t - \bar{x}^h) \xrightarrow{p} 0. \tag{S.86}$$

By (S.81) and $n^{-1}\sum_{t=n-j_0}^{n} \mathrm{E}|x^h_t - \bar{x}^h| = O(j_0 n^{-1})$, we have

$$\sup_{j\leq j_0} \alpha_{n2}(j,i,h)/n^{d_i+1/2} = O_p(1/n). \tag{S.87}$$

By (S.83) and $\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}|x^h_t - \bar{x}^h| = O(1)$, we have

$$\sup_{j\leq j_0} \alpha_{n1}(j,i,h)/n^{d_i+1/2} = o_p(1). \tag{S.88}$$

(S.85) follows by equations (S.86)–(S.88). □

Proof of Theorems 4.1 and 4.2. By Lemma 7, Theorems 4.1 and 4.2 can be established in a similar manner as to Theorems 3.1 and 3.2. Therefore we omit the detailed proofs.

S.4 Proofs for Remarks 3.4 and 3.5

Proof of Remark 3.4. (i) By the martingale version of the Skorokhod representation theorem (Strassen 1967, Hall and Heyde 1980, and Wu 2007), we have for all i, on a richer probability space, there exists a standard Brownian motion $\{W(t)\}$ and a non-negative stopping times $\{\tau_j^i\}$ such that for $t \geq 1$,

$$S_t^i = W(\sum_{j=1}^t \tau_j^i) \quad \text{and} \quad \mathrm{E}[\tau_t^i|\mathcal{F}_{t-1}(i)] = \mathrm{E}[(\varepsilon_t^i)^2|\mathcal{F}_{t-1}(i)], \tag{S.89}$$

where $\mathcal{F}_t(i)$ is the σ-algebra generated by $\{\varepsilon_s^i,\, s \leq t\}$. This implies that

$$\begin{aligned}
&\mathrm{E}|S_t^i - W(\sigma_{ii}t)|^2 = \mathrm{E}|\sum_{j=1}^t \tau_j^i - \sigma_{ii}t| \\
\leq\ & \mathrm{E}\left|\sum_{j=1}^t (\tau_j^i - \mathrm{E}(\tau_j^i|\mathcal{F}_{t-1}(i))\right| + \mathrm{E}\left|\sum_{j=1}^t [(\varepsilon_j^i)^2 - \mathrm{E}((\varepsilon_j^i)^2|\mathcal{F}_{t-1}(i))]\right| + \mathrm{E}\left|\sum_{j=1}^t (\varepsilon_j^i)^2 - \sigma_{ii}t\right|.
\end{aligned}$$

Since both $\{\tau_j^i - \mathrm{E}(\tau_j^i|\mathcal{F}_{t-1}(i))\}$ and $\{(\varepsilon_j^i)^2 - \mathrm{E}((\varepsilon_j^i)^2|\mathcal{F}_{t-1}(i))\}$ are martingale difference and $\mathrm{E}|\varepsilon_t^i|^q < \infty$, it follows that

$$\sup_{1\leq t\leq n} \mathrm{E}\left|\sum_{j=1}^t (\tau_j^i - \mathrm{E}(\tau_j^i|\mathcal{F}_{t-1}(i))\right| = O\left(\mathrm{E}\left|\sum_{j=1}^n [\tau_j^i - \mathrm{E}(\tau_j^i|\mathcal{F}_{t-1}(i))]\right|\right) = O(n^{2/q^*}).$$

Similarly, $\sup_{1\leq t\leq n} \mathrm{E}\left|\sum_{j=1}^t [(\varepsilon_j^i)^2 - \mathrm{E}((\varepsilon_j^i)^2|\mathcal{F}_{t-1}(i))]\right| = O(n^{2/q^*})$. Further, condition $\mathrm{E}\Big|\sum_{k=1}^n [(\varepsilon_t^i)^2 -\sigma_{ii}^2]\Big| = O(n^{2/q^*})$ implies that $\sup_{1\leq t\leq n} \mathrm{E}\left|\sum_{k=1}^t [(\varepsilon_t^i)^2 - \sigma_{ii}^2]\right| = O(n^{2/q^*})$. Thus, Condition 2(i) holds for any $\tau > 1/q^*$. If $p = o(n^{1/2})$, Condition 2(ii) holds. Since the components of $\boldsymbol{\varepsilon}_t$ are independent, Condition 2(iii) follows with $\sup_j \sum_{s,t=1}^n \mathrm{E}|\varepsilon_s^j \varepsilon_t^j| = O(n)$.

(ii) By the proof of Theorem 9.3.1 of Lin and Lu (1996), we know that there exists a martingale difference sequence $\{m_t^i\}$ such that $R_t = S_t^i - M_t^i$ satisfying $\mathrm{E}|R_t|^q = O(1)$, where $M_t^i = \sum_{j=1}^t m_j^i$. Further,

$$\mathrm{E}\left|\sum_{j=1}^n [(m_j^i)^2 - \mathrm{E}(m_j^i)^2]\right|^{q/2} \leq Cn\log n. \tag{S.90}$$

As a result, Condition 2(i) holds for any $\tau > 1/q$. Similarly, Condition 2(iii) can be easily obtained by basic inequality for mixing processes, see Lemma 1.2.2 of Lin and Lu (1996). Note that for any given j,

$$\begin{aligned}
&\mathrm{E}\Big\|\frac{1}{n}\sum_{t=1}^{n-j} (\mathbf{x}_{t+j,2} - \bar{\mathbf{x}}_2)(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)' - \mathrm{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}_{1,2})\Big\|_2 \\
\leq\ & \sum_{i,j=p-r+1}^p \mathrm{E}\left(\frac{1}{n}\sum_{t=1}^n [(x_t^i - \bar{x}^i)(x_t^j - \bar{x}^j) - \mathrm{Cov}(x_t^i, x_t^j)]\right)^2 = O(p^2/n) \to 0
\end{aligned}$$

as $p = o(n^{1/2})$. Condition 2(ii) holds too.

(iii) By Beveridge-Nelson decomposition, ε_t^i can be represented as

$$\varepsilon_t^i = \Big(\sum_{j=0}^{\infty} c_{ij}\Big)\eta_t^i - (\epsilon_t - \epsilon_{t-1}),\ \epsilon_t = \sum_{j=0}^{\infty} \widetilde{c}_{ij}\eta_{t-j}^i,\ \widetilde{c}_{ij} = \sum_{h=j+1}^{\infty} c_h.$$

Let $R_t^i = S_t^i - \Big(\sum_{j=0}^{\infty} c_{ij}\Big)\sum_{j=1}^{t}\eta_j^i = \epsilon_t - \epsilon_0$. Then

$$\sup_{1\le t\le n}\left|S_t^i - \Big(\sum_{j=0}^{\infty} c_{ij}\Big)\sum_{j=1}^{t}\eta_j^i\right|^2 = O(1) \tag{S.91}$$

Since $\{\eta_t^i\}$ is an i.i.d sequence with $\mathrm{E}|\eta_j^i|^q < \infty$, by Theorem 4.3 of Strassen (1967), the stopping $\{\tau_t^i\}$ defined as in (S.89) is an independent sequence with $\mathrm{E}\tau_t^i = \mathrm{E}(\eta_j^i)^2$ and $\mathrm{E}|\tau_t^i|^{q/2} < \infty$. Thus, $\sup_{1\le t\le n}\mathrm{E}|\sum_{j=1}^{t}[\tau_j^i - \mathrm{E}(\eta_j^i)^2]|^{q/2} = O(n + n^{q/4})$. As a result, we have for $q_0 = \min\{q, 4\}$,

$$\sup_{1\le t\le n}\mathrm{E}|\sum_{j=1}^{t}[\tau_j^i - \mathrm{E}(\eta_j^i)^2]| = O(n^{2/q_0}).$$

Let $a_i = \mathrm{E}(\eta_j^i)^2$, then on a richer space there exist a standard Brownian motion $W(t)$ such that

$$\mathrm{E}(\sum_{j=1}^{t}\eta_j^i - W(a_i t))^2 = O(n^{2/q_0}). \tag{S.92}$$

Thus, by (S.91), (S.92), Condition 2(i) holds for $\tau = 1/q_0$. It is easy to show that $\sup_{p-r<j\le p}\sum_{s,t=1}^{n}|\mathrm{E}(\varepsilon_t^j\varepsilon_s^j)|$ $= O(n)$. Thus, Condition 2(iii) follows by the independence of the components. Condition 2(ii) can be shown similarly to Remark 3.4(ii). □

Proof of Remark 3.5. It is easy to get $\lambda^* = O_e(1)$ when $p - r$ is fixed. We only show the case when $m := p - r \to \infty$ as $n \to \infty$. Let $d = d_{\min}$, ξ_t^i be $I(1)$ process defined as in Lemma 3, $\bar{\boldsymbol{\xi}}^i = (\bar{\xi}^i, \cdots, \bar{\xi}^i)'$ and $\mathbf{e} = (1, \cdots, 1)'$ be two n dimensional vectors. Let $\mathbf{E}_n$ and $\mathbf{\Pi}_n$ be $n \times n$ matrices given by

$$\mathbf{E}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{e}\mathbf{e}' \ \text{ and } \ \mathbf{\Pi}_n = \begin{pmatrix} 1 & 0 & 0 & & 0\cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix},$$

then for any $1 \le i, j \le m$,

$$\begin{aligned} \boldsymbol{\xi}^i &= (\xi_1^i, \cdots, \xi_n^i)' = \sigma_{ii}\mathbf{\Pi}_n^{-d}(v_1^i, v_2^i, \cdots, v_n^i)' =: \sigma_{ii}\mathbf{\Pi}_n^{-d}\mathbf{V}^i, \text{ and} \\ (\boldsymbol{\xi}^i - \bar{\boldsymbol{\xi}}^i)'(\boldsymbol{\xi}^j - \bar{\boldsymbol{\xi}}^j) &= \sigma_{ii}\sigma_{jj}(\mathbf{V}^i)'(\mathbf{\Pi}_n^{-d})'\mathbf{E}_n\mathbf{E}_n'\mathbf{\Pi}_n^{-d}\mathbf{V}^j. \end{aligned}$$

Let $\delta_1 \leq \cdots \leq \delta_n$ and $\gamma_1 \geq \cdots \geq \gamma_n$ be the eigenvalues of $\mathbf{\Pi}_n\mathbf{\Pi}_n'$ and $(\mathbf{\Pi}_n^{-d})'\mathbf{E}_n\mathbf{E}_n'\mathbf{\Pi}_n^{-d}$ respectively. Since $\lambda_1(\mathbf{E}_n\mathbf{E}_n') = \cdots = \lambda_{n-1}(\mathbf{E}_n\mathbf{E}_n') = 1$, by Theorem 9 of Merikoski and Kumar (2004), it follows that

$$\delta_{i+1}^{-d} = \lambda_{i+1}((\mathbf{\Pi}_n^{-d})'\mathbf{\Pi}_n^{-d})\lambda_{n-1}(\mathbf{E}_n\mathbf{E}_n') \leq \gamma_i \leq \lambda_i((\mathbf{\Pi}_n^{-d})'\mathbf{\Pi}_n^{-d})\lambda_1(\mathbf{E}_n\mathbf{E}_n') = \delta_i^{-d}. \tag{S.93}$$

Further, $\delta_k = 2 - 2\cos(2k\pi/(2n+1))$, $k = 1, 2, \cdots, n$ (see Yueh (2005)), which implies

$$\delta_k \sim 4(k\pi/(2n+1))^2, \text{ as } k/n \to 0. \tag{S.94}$$

Let $\mathbf{U}$ be an orthogonal matrix with row vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$ such that $(\mathbf{\Pi}_n^{-d})'\mathbf{E}_n\mathbf{E}_n'\mathbf{\Pi}_n^{-d} = \mathbf{U}\text{diag}(\gamma_1, \cdots, \gamma_n)\mathbf{U}'$ and let $\mathbf{\Omega} = (\mathbf{V}^1, \cdots, \mathbf{V}^m)$. For $\mathbf{x} \in \mathcal{R}^m$ with $\mathbf{x}'\mathbf{x} = 1$, define $\mathbf{U\Omega x} = (b_{1\mathbf{x}}, \cdots, b_{n\mathbf{x}})' = \mathbf{b}_\mathbf{x} \in \mathcal{R}^n$. By (S.93), we have

$$\begin{aligned}
&\lambda_{\min}\left(\frac{1}{n^{2d}}\sum_{t=1}^n(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_t' - \bar{\boldsymbol{\xi}}')\right) \\
= \;& \lambda_{\min}\left(\frac{1}{n^{2d}}(\boldsymbol{\xi}^1 - \bar{\boldsymbol{\xi}}^1, \cdots, \boldsymbol{\xi}^m - \bar{\boldsymbol{\xi}}^m)'(\boldsymbol{\xi}^1 - \bar{\boldsymbol{\xi}}^1, \cdots, \boldsymbol{\xi}^m - \bar{\boldsymbol{\xi}}^m)\right) \\
\geq \;& \{\min_i(\sigma_{ii})\}^2 \min_\mathbf{x} \frac{1}{n^{2d}}\mathbf{x}'(\mathbf{U\Omega})'\text{diag}(\gamma_1, \cdots, \gamma_n)(\mathbf{U\Omega})\mathbf{x} \\
= \;& \{\min_i(\sigma_{ii})\}^2 \min_\mathbf{x} \frac{(\mathbf{b}_\mathbf{x}'\mathbf{b}_\mathbf{x})}{n^{2d}}\frac{\mathbf{b}_\mathbf{x}'(\gamma_1, \cdots, \gamma_n)\mathbf{b}_\mathbf{x}}{\mathbf{b}_\mathbf{x}'\mathbf{b}_\mathbf{x}} \\
\geq \;& \{\min_i(\sigma_{ii})\}^2 n^{1-2d}\lambda_{\min}(\mathbf{\Omega}'\mathbf{\Omega}/n)\min_\mathbf{x}\left(\sum_{l=1}^k \delta_{k+1}^{-d} b_{l\mathbf{x}}^2 / \sum_{l=1}^n b_{l\mathbf{x}}^2\right) \\
\geq \;& \{\min_i(\sigma_{ii})\}^2 (k/n^{2d})[\lambda_{\min}(\mathbf{\Omega}'\mathbf{\Omega}/n)/\lambda_{\max}(\mathbf{\Omega}'\mathbf{\Omega}/n)]\delta_{k+1}^{-d}\min_\mathbf{x}\frac{1}{k}\sum_{l=1}^k b_{l\mathbf{x}}^2 \\
= \;& O_e(k^{1-2d})\lambda_{\min}\{[\mathbf{\Omega}'(\mathbf{u}_1', \cdots, \mathbf{u}_k')][(\mathbf{u}_1', \cdots, \mathbf{u}_k')'\mathbf{\Omega}]\},
\end{aligned} \tag{S.95}$$

where $\lambda_{\min}$, $\lambda_{\max}$ denote the smallest and largest eigenvalues of a matrix respectively, the last equation follows by (S.94) and the fact that there exist two positive constants C_1, C_2 such that $C_1 < \lambda_{\min}(\mathbf{\Omega}'\mathbf{\Omega}/n) \leq \lambda_{\max}(\mathbf{\Omega}'\mathbf{\Omega}/n) < C_2$ in probability when $p/n^{1/2} \to 0$. Since U is orthogonal and the elements of $\mathbf{\Omega}$ are independent standard normal variables, it follows that the elements of $(\mathbf{u}_1', \cdots, \mathbf{u}_k')'\mathbf{\Omega}$ are independent and standard normal variables, thus by Theorem 2 of Bai and Yin (1993), we have if $m/k \in (0, 1)$,

$$\lambda_{\min}\{[\mathbf{\Omega}'(\mathbf{u}_1', \cdots, \mathbf{u}_k')][(\mathbf{u}_1', \cdots, \mathbf{u}_k')'\mathbf{\Omega}]\} = (1 - \sqrt{m/k})^2, \text{ a.s.} \tag{S.96}$$

Taking $k = 2m$, then by (S.95) and (S.96), we have

$$\lambda_{\min}\left(\frac{1}{n^{2d}}\sum_{t=1}^n(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_t' - \bar{\boldsymbol{\xi}}')\right) \geq Cm^{1-2d}. \tag{S.97}$$

Since $||\frac{1}{n^{2d}}\sum_{t=1}^n(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_t' - \bar{\boldsymbol{\xi}}') - \int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\,dt||_2 = o_p(1)$ (see Lemma 3), Remark 3.5 follows from (S.97). $\square$

S.5 Numerical studies for j_0 and m

To evaluate the impact of the choice of j_0 and m, we report some of our simulation results in this subsection.

We first present a numerical result for the choice j_0. We let $\mathbf{x}_{t2}$ in model (1) consist of r stationary AR(1) processes with coefficients generated independently from $U(-0.8, 0.8)$, $\mathbf{x}_{t1}$ be $p-r$ ARIMA(1,1,1) processes with coefficients generated from $U(0, 0.8)$ and $U(-0.8, 0.8)$. Let the elements of $\mathbf{A}$ be generated independently from $U(-3, 3)$. We estimate the cointegration rank r by the ACF unit-root test defined in (5). For each setting, we replicate the exercise 500 times with sample size $n = 300$, 500, 1000 and j_0 ranging from 5 to 90. The relative frequencies for the occurrence of events $\{\widehat{r} = r\}$ and the average distance between the true cointegrating space and its true space are reported in Table S.1. It is shown from Table S.1 that the performance is stable with respect to the choice of j_0, especially for small p and large n.

Table S.1: Relative frequencies (RF) of the occurrence of event $\{\widehat{r} = r\}$ and average distance D_1 with j_0 ranging over $(5, 90)$ and 500 replications.

(p,r)	n		j=5	10	20	30	40	50	60	70	80	85	90
(4, 2)	300	RF	.950	.928	.938	.960	.940	.924	.928	.948	.934	.926	.926
		D1	.062	.069	.057	.047	.065	.065	.070	.053	.064	.068	.068
	500	RF	.982	.982	.986	.978	.984	.974	.978	.980	.978	.976	.988
		D1	.029	.024	.023	.029	.027	.036	.028	.030	.029	.030	.024
	1000	RF	.992	.994	.996	.996	.994	.998	1.00	1.00	.994	.998	.998
		D1	.013	.012	.013	.012	.011	.014	.008	.009	.014	.012	.009
(6, 3)	300	RF	.822	.794	.834	.810	.802	.826	.828	.794	.834	.816	.812
		D1	.128	.137	.115	.131	.136	.121	.122	.136	.120	.124	.129
	500	RF	.934	.948	.946	.938	.960	.962	.964	.970	.952	.960	.958
		D1	.061	.053	.053	.062	.052	.049	.047	.045	.050	.045	.050
	1000	RF	.988	.990	.994	.994	.976	.984	.992	.994	.988	.994	.994
		D1	.024	.018	.018	.018	.026	.022	.017	.017	.021	.020	.020
(8, 4)	300	RF	.562	.564	.578	.628	.612	.648	.620	.646	.592	.598	.610
		D1	.230	.224	.223	.204	.211	.198	.209	.194	.217	.213	.213
	500	RF	.874	.886	.858	.908	.884	.920	.910	.934	.900	.898	.914
		D1	.093	.078	.101	.078	.085	.077	.077	.067	.081	.081	.078
	1000	RF	.966	.978	.986	.980	.984	.986	.986	.986	.988	.988	.990
		D1	.046	.031	.028	.032	.030	.030	.027	.029	.031	.028	.028

Next table is reported for the choice of m. In this simulation, $\mathbf{x}_{t1}$ and $\mathbf{x}_{t2}$ are generated from model (1) as the previous example. We also replicate the exercise 500 times in each setting with sample size $n = 300$, 500 and 1000 and the lags number m is taken from 5 to 90. The corresponding relative frequencies for the occurrence of events $\{\widehat{r} = r\}$ and the average distance between the true cointegrating space and its true space are reported in Table S.2. It is shown from Table S.2 that a relatively small m always works well for the estimation of the cointegrating space. On the contrary, if m is selected too large, the performance is relatively poor, especially when the sample size n is relatively small. This is reasonable, because from Remark 2.1, we know that only when $n/m \to \infty$, $\sum_{k=1}^{m} \widehat{\rho}(k)/m - 1 \to 0$, otherwise, it is difficult to distinguish the integrated process from the stationary process, which means that m could not be selected too large, especially when n is relatively small. This simulation also confirms that $m = 20$ is usually good enough for the procedure.

In our simulation, we also use a data driven procedure in selecting m, which is given by

$$\widehat{m} = \text{argmin}_m \{f(m) \vee f^{-1}(m)\}$$

where $T(m) = \left\{ \frac{n}{m(m+1)} \sum_{i=1}^{p} \sum_{k=1}^{m} (1 - \widehat{\rho}_i(k)) \right\}$ and $f(m) = T(m)/T(m+1)$. It also work reasonably, for example, when $(p, r) = (6, 3)$, the corresponding relative correct frequencies and average distance is $(0.776, 0.119)$ for $n = 300$, $(0.888, 0.062)$ for $n = 500$ and $(0.970, 0.0229)$ for $n = 1000$.

Table S.2: Relative frequencies (RF) of the occurrence of event $\{\widehat{r} = r\}$ and average distance D_1 with m ranging over $(5, 90)$ and 500 replications.

(p,r)	n		j=5	10	20	30	40	50	60	70	80	85	90
(3, 2)	300	RF	.872	.976	.980	.968	.936	.882	.816	.744	.644	.602	.586
		D1	.016	.021	.025	.031	.048	.078	.115	.156	.211	.235	.243
	500	RF	.864	.980	.998	.998	.996	.986	.968	.942	.914	.896	.866
		D1	.008	.009	.010	.010	.011	.017	.026	.041	.056	.067	.084
	1000	RF	.874	.980	.996	.996	.996	.994	.994	.992	.992	.992	.992
		D1	.006	.005	.006	.006	.006	.007	.007	.008	.008	.008	.008
(6, 4)	300	RF	.850	.942	.916	.818	.702	.560	.398	.250	.140	.102	.070
		D1	.043	.050	.064	.101	.149	.211	.283	.349	.403	.422	.441
	500	RF	.848	.966	.972	.958	.916	.854	.762	.660	.568	.518	.462
		D1	0.024	.026	.029	.034	.051	.076	.115	.160	.200	.223	.248
	1000	RF	.842	.974	.996	.996	.992	.984	.976	.968	.942	.928	.906
		D1	.011	.012	.013	.013	.014	.017	.020	.023	.034	.039	.049
(9, 6)	300	RF	.802	.848	.732	.510	.290	.118	.040	.004	0	0	0
		D1	.083	.095	.128	.203	.286	.363	.414	.452	.476	.483	.494
	500	RF	.800	.944	.930	.896	.792	.638	.438	.294	.186	.132	.104
		D1	.039	.045	.052	.063	.098	.154	.227	.286	.334	.359	.375
	1000	RF	.856	.984	.990	.980	.980	.964	.946	.914	.862	.834	.786
		D1	.018	.019	.019	.021	.021	.026	.032	.044	.062	.072	.089

References

[1] Bai, Z. D. and Yin, Y. Q. (1993). Limit of smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, **21**, 1275–1294.

[2] Barlow, J. L. and Slapničar, I. (2000). Optimal perturbation bounds for Hermitian eigenvalue problem. *Linear Algebra and its Applications*, **309**, 19–43.

[3] Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *The Annals of Statistics*, **16**, 367–401.

[4] Chang, J. Y., Guo, B. and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, to appear.

[5] Chen, W. W. and Hurvich, C. M. (2006). Semiparametric estimation of fractional cointegrating subspaces. *The Annals of Statistics*, **34**, 2939–2979.

[6] Dopico, F. M., Moro, J. and Molera, J. M. (2000). Weyl-type relative perturbation bounds for eigensystems of Hermitian matrices. *Linear Algebra and its Applications*, **309**, 3–18.

[7] Golub, G. and Van Loan, C. (1996). *Matrix Computations (3rd edition)*. John Hopkins University Press.

[8] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

[9] Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, **40**, 694–726.

[10] Lin, Z. and Lu, C. (1997). *Limit Theory on Mixing Dependent Random Variables*. Kluwer Academic Publishers, 1997.

[11] Ling, S. and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *The Annals of Statistics*, **26**, 84–125.

[12] Marnucci, D. and Robinson, P. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications*, **86**, 103–120.

[13] Strassen, V. (1967). Almost sure behaviour of sums of independent random variables and martingales. *Proceedings of the Fifth Berkeley Symposium of Mathematical Statistics and Probability*, **2**, 315-343. University of California Press, Berkeley.

[14] Stewart, G. W. and Sun, J. (1990). *Matrix perturbation Theory*. Academic Press.

[15] Vu, V. Q. and Lei, J. (2013). Minimax sparse principle subspace estimation in high dimensions. *The Annals of Statistics*, **41**, 2095–2947.

[16] Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability*, **35**, 2294–2320.

[17] Yueh, W. C. (2005). Eigenvalues of several tridiagonal matrices. *Applied Mathematics E-Notes*, **5**, 66–74.