Statistical Inference for Heavy-tailed and Partially Nonstationary Vector ARMA Models

Feifei Guo and Shiqing Ling

Beijing Institute of Technology

Hong Kong University of Science and Technology

Supplementary Material

S1 List of Notations

Notations in Theorem 1:

$\mathbf{C}_0 = \mathbf{A}_0\mathbf{B}_0$ $--$ true values of unknown parameters of $\mathbf{C} = \mathbf{AB}$

$\hat{\mathbf{C}}$ $--$ FLSE of $\mathbf{C}$.

$\mathbf{\Phi}^*_{0,j}$ $--$ true values of unknown parameters of $\mathbf{\Phi}^*_j$

$\mathbf{\Theta}_{0,j}$ $--$ true values of unknown parameters of $\mathbf{\Theta}_j$

$\hat{\mathbf{\Phi}}^*_j$ $--$ FLSE of $\mathbf{\Phi}^*_j$

$\hat{\mathbf{\Theta}}_j$ $--$ FLSE of $\mathbf{\Theta}_j$

$\mathbf{B}_\perp$ -- $d \times m$ orthogonal matrix of $\mathbf{B}_0$

$\mathbf{Q}' = [\mathbf{Q}_1, \mathbf{Q}_2] \equiv [\mathbf{B}_\perp', \mathbf{B}_0']$

$\bar{\mathbf{B}}_\perp = (\mathbf{B}_\perp \mathbf{B}_\perp')^{-1}\mathbf{B}_\perp$

$\bar{\mathbf{B}} = (\mathbf{B}_0 \mathbf{B}_0')^{-1}\mathbf{B}_0$

$\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2] \equiv [\bar{\mathbf{B}}_\perp', \bar{\mathbf{B}}']$

$\hat{\boldsymbol{\theta}} = [\{\mathrm{vec}(\hat{\mathbf{C}}\mathbf{P}_2)\}', (\mathrm{vec}[\hat{\boldsymbol{\Phi}}_1^*, \ldots, \hat{\boldsymbol{\Phi}}_{p-1}^*, \hat{\boldsymbol{\Theta}}_1, \ldots, \hat{\boldsymbol{\Theta}}_q])']'$

$\boldsymbol{\theta}_0 = [\{\mathrm{vec}(\mathbf{C}_0\mathbf{P}_2)\}', (\mathrm{vec}[\boldsymbol{\Phi}_{0,1}^*, \ldots, \boldsymbol{\Phi}_{0,p-1}^*, \boldsymbol{\Theta}_{0,1}, \ldots, \boldsymbol{\Theta}_{0,q}])']'$

$\boldsymbol{\Theta}_0 = \mathbf{I}_m - \sum_{j=1}^q \boldsymbol{\Theta}_{0,j}$

$\mathbf{W}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1} = \sum_{j=0}^\infty \mathbf{B}_j \boldsymbol{\varepsilon}_{t-j}$

$\mathbf{Z}_{1,t} = \mathbf{Q}_1'\mathbf{Y}_t = [\mathbf{I}_d, \mathbf{0}] \sum_{i=1}^t \sum_{j=0}^\infty \boldsymbol{\phi}_j \boldsymbol{\varepsilon}_{i-j}$

$\mathbf{Z}_{2,t} = \mathbf{Q}_2'\mathbf{Y}_t = \sum_{j=0}^\infty \mathbf{C}_j \boldsymbol{\varepsilon}_{t-j}$

$\mathbf{U}_t = [\mathbf{Z}_{2,t}', \mathbf{W}_t', \ldots, \mathbf{W}_{t-p+2}', -\boldsymbol{\varepsilon}_t', \ldots, -\boldsymbol{\varepsilon}_{t-q+1}']' = \sum_{i=0}^\infty \mathbf{A}_i \boldsymbol{\varepsilon}_{t-i}$

$(\mathbf{I}_m - \sum_{j=1}^q \boldsymbol{\Theta}_{0,j} L^j)^{-1} \equiv \sum_{k=0}^\infty \boldsymbol{\gamma}_{0,k} L^k$

$\mathbf{R}_1 = [\int_0^1 \mathbf{P}(r) d\mathbf{P}'(r)]' \boldsymbol{\phi}' [\mathbf{I}_d, \mathbf{0}]'$, $\boldsymbol{\phi} = \sum\limits_{i=0}^{\infty} \boldsymbol{\phi}_i$, $\mathbf{P}(r)$ is a stable process.

$\mathbf{S}_{11} = [\mathbf{I}_d, \mathbf{0}] \boldsymbol{\phi} [\int_0^1 \mathbf{P}(r)\mathbf{P}'(r) dr] \boldsymbol{\phi}' [\mathbf{I}_d, \mathbf{0}]'$

$\mathbf{F}_0 = \mathbf{R}_1 \mathbf{S}_{11}^{-1}$

$\mathbf{S}_1 = \sum_{i=1}^\infty \mathbf{P}_i^{(1)} \mathbf{P}_i^{(1)'}$, $\mathbf{P}_i^{(1)} = \mathbf{P}_i$

$\mathbf{S}_j = \sum_{i=1}^\infty \mathbf{P}_i^{(j)}$ for all $j > 1$

$\mathbf{R}_{2l} = \sum_{i=0}^\infty \mathbf{S}_{i+2+l} \mathbf{A}_i'$

$\mathbf{S}_{22kj} = \sum_{i=0}^\infty \mathbf{A}_i \mathbf{S}_1 \mathbf{A}_{i+k-j}'$

$\mathbf{S}_{22k'il} = \sum_{j=0}^{\infty} \mathbf{B}_j \mathbf{S}_1 \mathbf{A}'_{j+k'+i-l}$

$\mathbf{S}_{12l} = \{\mathbf{R}'_1 \sum_{i=0}^{\infty} \mathbf{A}'_{i+l} + [\mathbf{I}_d, \mathbf{0}] \sum_{i=0}^{\infty} \sum_{j=0}^{i} \boldsymbol{\phi}_j \mathbf{S}_1 \mathbf{A}'_{i+l}\}$

$\mathbf{F}_{1l} = \sum_{j'=1}^{q} \sum_{i=0}^{j'-1} \sum_{k'=0}^{\infty} \boldsymbol{\Theta}_0^{-1} \boldsymbol{\Theta}_{0,j'} \boldsymbol{\gamma}_{0,k'} \boldsymbol{\Theta}_0 \mathbf{R}_1 \mathbf{S}_{11}^{-1} \mathbf{B}_{\perp} \mathbf{S}_{22k'il}$

$\boldsymbol{\Gamma}_{22} = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \mathbf{S}_{22kj} \otimes \boldsymbol{\gamma}'_{0,k} \boldsymbol{\gamma}_{0,j}$

Notations in Corollary 1:

$\mathbf{B} = [\mathbf{I}_r, \mathbf{B}^*]$, where $\mathbf{B}^*$ is an $r \times d$ matrix of unknown parameters

$\mathbf{B}_0 = [\mathbf{I}_r, \mathbf{B}_0^*]$ $--$ true value of $\mathbf{B}$

$\hat{\mathbf{C}} = [\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2]$, where $\hat{\mathbf{C}}_1$ is an $m \times r$ matrix and $\hat{\mathbf{C}}_2$ is an $m \times d$ matrix.

$\hat{\mathbf{A}} \equiv \hat{\mathbf{C}}_1$

$\hat{\boldsymbol{\Theta}} = \mathbf{I}_m - \sum_{j=1}^{q} \hat{\boldsymbol{\Theta}}_j$

$\hat{\mathbf{B}}^* \equiv [\hat{\mathbf{A}}' \hat{\boldsymbol{\Theta}}'^{-1} \hat{\boldsymbol{\Theta}}^{-1} \hat{\mathbf{A}}]^{-1} [\hat{\mathbf{A}}' \hat{\boldsymbol{\Theta}}'^{-1} \hat{\boldsymbol{\Theta}}^{-1} \hat{\mathbf{C}}_2]$

$\hat{\boldsymbol{\delta}} = \text{vec}[\hat{\mathbf{A}}, \hat{\boldsymbol{\Phi}}_1^*, \ldots, \hat{\boldsymbol{\Phi}}_{p-1}^*, \hat{\boldsymbol{\Theta}}_1, \ldots, \hat{\boldsymbol{\Theta}}_q]$

$\boldsymbol{\delta}_0 = \text{vec}[\mathbf{A}_0, \boldsymbol{\Phi}_{0,1}^*, \ldots, \boldsymbol{\Phi}_{0,p-1}^*, \boldsymbol{\Theta}_{0,1}, \ldots, \boldsymbol{\Theta}_{0,q}]$

$\bar{\mathbf{B}}'_{\perp} = [\bar{\mathbf{B}}'_{\perp,1}, \bar{\mathbf{B}}'_{\perp,2}]'$, where $\bar{\mathbf{B}}_{\perp,2}$ is the last d rows of $\bar{\mathbf{B}}'_{\perp}$

$\bar{\mathbf{B}}' = [\bar{\mathbf{B}}'_1, \bar{\mathbf{B}}'_2]'$, where $\bar{\mathbf{B}}_2$ is the last d rows of $\bar{\mathbf{B}}'$

$\mathbf{M} = (\mathbf{A}'_0 \boldsymbol{\Theta}_0'^{-1} \boldsymbol{\Theta}_0^{-1} \mathbf{A}_0)^{-1} \mathbf{A}'_0 \boldsymbol{\Theta}_0'^{-1}$.

Notations in Theorem 2:

$\tilde{\mathbf{B}}^*$, $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\Phi}}_i^*$ and $\tilde{\boldsymbol{\Theta}}_i$ $--$ RLSE of $\mathbf{B}^*$, $\mathbf{A}$, $\boldsymbol{\Phi}_i^*$ and $\boldsymbol{\Theta}_i$

$\tilde{\boldsymbol{\delta}} = \text{vec}[\tilde{\mathbf{A}}, \tilde{\boldsymbol{\Phi}}_1^*, \ldots, \tilde{\boldsymbol{\Phi}}_{p-1}^*, \tilde{\boldsymbol{\Theta}}_1, \ldots, \tilde{\boldsymbol{\Theta}}_q]$

$\mathbf{H}' -- \ d \times m$ matrix $[\mathbf{0}, \mathbf{I}_d]$

$\mathbf{S}_{22l} = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{S}_1 \mathbf{A}_{j+l}'$

$\mathbf{I}_{m.q} -- \ m \times mq$ matrix by arranging q $m \times m$ identity matrices one by one.

Notations in Lemma 1:

$\boldsymbol{\beta} = \text{vec}[\mathbf{C}, \boldsymbol{\Phi}_1^*, \ldots, \boldsymbol{\Phi}_{p-1}^*, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q]$

$\boldsymbol{\varepsilon}_t(\boldsymbol{\beta}) = \mathbf{W}_t - ([\mathbf{Y}_{t-1}', \mathbf{W}_{t-1}', \ldots, \mathbf{W}_{t-p+1}', -\boldsymbol{\varepsilon}_{t-1}'(\boldsymbol{\beta}), \ldots, -\boldsymbol{\varepsilon}_{t-q}'(\boldsymbol{\beta})] \otimes \mathbf{I}_m)\boldsymbol{\beta}$

$\mathbf{c} = \text{vec}(\mathbf{C})$

$\boldsymbol{\Theta} = \mathbf{I}_m - \sum_{j=1}^{q} \boldsymbol{\Theta}_j$

$(\mathbf{I}_m - \sum_{j=1}^{q} \boldsymbol{\Theta}_j L^j)^{-1} = \sum_{k=0}^{\infty} \boldsymbol{\gamma}_k L^k$

Notations in Lemma 3:

$\mathbf{b} = \text{vec}(\mathbf{B}^*)$

$\boldsymbol{\delta} = \text{vec}[\mathbf{A}, \boldsymbol{\Phi}_1^*, \ldots, \boldsymbol{\Phi}_{p-1}^*, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q]$

$\boldsymbol{\eta} = (\mathbf{b}', \boldsymbol{\delta}')'$

$\boldsymbol{\varepsilon}_t(\boldsymbol{\eta}) = (\mathbf{I}_m - \sum_{j=1}^{q} \boldsymbol{\Theta}_j L^j)^{-1}(\mathbf{W}_t - \mathbf{A}[\mathbf{I}_r, \mathbf{B}^*]\mathbf{Y}_{t-1} - \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j^* \mathbf{W}_{t-j})$.

S2 Inclusion of Constant Term in Model

The results in Sections 2 and 3 can be readily extended to models with a constant term included, such that the purely nonstationary components $\mathbf{Z}_{1,t} = \mathbf{Q}_1'\mathbf{Y}_t$ have zero drift. Thus, the model considered is

$$\mathbf{W}_t = \boldsymbol{\mu}^* + \mathbf{C}\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j^* \mathbf{W}_{t-j} + \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{\varepsilon}_{t-j}, \tag{S2.1}$$

where $\boldsymbol{\mu}^*$ is an $m \times 1$ vector of unknown constants, and its true value is $\boldsymbol{\mu}_0^*$ such that $\mathbf{Q}_1'\boldsymbol{\mu}_0^* = \mathbf{0}$, which implies that $\mathbf{Z}_{1,t} - \mathbf{Z}_{1,t-1}$ has zero mean. The FLSE and RLSE can be obtained as before, by replacing $\mathbf{Y}_{t-1}$ by $\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(1)}$, and $\mathbf{W}_{t-j}$ by $\mathbf{W}_{t-j} - \bar{\mathbf{W}}_{(j)}$, $0 \le j \le (p-1)$, in estimation procedures, where $\bar{\mathbf{Y}}_{(1)} = (n-p)^{-1} \sum_{t=p+1}^{n} \mathbf{Y}_{t-1}$ and $\bar{\mathbf{W}}_{(j)} = (n-p)^{-1} \sum_{t=p+1}^{n} \mathbf{W}_{t-j}$. Denote

$$\mathbf{R}_{1n}^* = \sum_{t=1}^{n} (\boldsymbol{\varepsilon}_t - \bar{\boldsymbol{\varepsilon}})(\mathbf{Z}_{1,t-1} - \bar{\mathbf{Z}}_1)', \ \mathbf{R}_{2n}^* = \sum_{t=1}^{n} (\boldsymbol{\varepsilon}_t - \bar{\boldsymbol{\varepsilon}})(\mathbf{U}_{t-1-l} - \bar{\mathbf{U}})',$$
$$\mathbf{S}_{11n}^* = \sum_{t=1}^{n} (\mathbf{Z}_{1,t-1} - \bar{\mathbf{Z}}_1)(\mathbf{Z}_{1,t-1} - \bar{\mathbf{Z}}_1)',$$
$$\mathbf{S}_{12n}^* = \sum_{t=1}^{n} (\mathbf{Z}_{1,t-1} - \bar{\mathbf{Z}}_1)(\mathbf{U}_{t-1-l} - \bar{\mathbf{U}})' \text{ and } \mathbf{S}_{22n}^* = \sum_{t=1}^{n} (\mathbf{U}_{t-1-k} - \bar{\mathbf{U}})(\mathbf{U}_{t-1-j} - \bar{\mathbf{U}})',$$

where $\bar{\boldsymbol{\varepsilon}} = n^{-1} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t$, $\bar{\mathbf{Z}}_1 = n^{-1} \sum_{t=1}^{n} \mathbf{Z}_{1,t-1}$ and $\bar{\mathbf{U}} = n^{-1} \sum_{t=1}^{n} \mathbf{U}_{t-1}$.

To obtain the asymptotic distribution of the estimators, we note from

Resnick (1986) that

$$a_n^{-1}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t \xrightarrow{d} \mathbf{P}(1), \quad a_n^{-1}\sum_{t=1}^{n}\mathbf{U}_{t-1} \xrightarrow{d} \mathbf{S}_0 \ \text{ and } \ a_n^{-1}\bar{\mathbf{Z}}_1 \xrightarrow{d} \boldsymbol{\xi},$$

where $\mathbf{S}_0 = \sum_{i=0}^{\infty}\mathbf{A}_i\mathbf{P}(1)$ and $\boldsymbol{\xi} = [\mathbf{I}_d, \mathbf{0}]\boldsymbol{\phi}[\int_0^1 \mathbf{P}(r)dr]$. Then, by Theorem A.2 in She and Ling (2020), Lemma 1 and Lemma 2, we can show when $\alpha \in (0,1)$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$, as $n \to \infty$

$$\left(\frac{1}{na_n^2}\mathbf{S}_{11n}^*, \frac{1}{a_n^2}\mathbf{S}_{22n}^*, \frac{1}{a_n^2}\mathbf{S}_{12n}^*, \frac{1}{a_n^2}\mathbf{R}_{1n}^*, \frac{n}{a_n^2}\mathbf{R}_{2n}^*\right) \xrightarrow{d} (\boldsymbol{\Gamma}_{11}^*, \mathbf{S}_{22kj}, \boldsymbol{\Gamma}_{12}^*, \mathbf{R}_1^*, \mathbf{R}_2^*).$$

When $\alpha \in (1,2)$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$, as $n \to \infty$, we have

$$\left(\frac{1}{na_n^2}\mathbf{S}_{11n}^*, \frac{1}{a_n^2}\mathbf{S}_{22n}^*, \frac{1}{a_n^2}\mathbf{S}_{12n}^*, \frac{1}{a_n^2}\mathbf{R}_{1n}^*, \frac{1}{\tilde{a}_n}\mathbf{R}_{2n}^*\right) \xrightarrow{d} (\boldsymbol{\Gamma}_{11}^*, \mathbf{S}_{22kj}, \boldsymbol{\Gamma}_{12}^*, \mathbf{R}_1^*, \mathbf{R}_{2l}),$$

where $\boldsymbol{\Gamma}_{11}^* = \mathbf{S}_{11} - \boldsymbol{\xi}\boldsymbol{\xi}'$, $\boldsymbol{\Gamma}_{12}^* = \mathbf{S}_{12l} - \boldsymbol{\xi}\mathbf{S}_0'$, $\mathbf{R}_1^* = \mathbf{R}_1 - \mathbf{P}(1)\boldsymbol{\xi}'$ and $\mathbf{R}_2^* = -\mathbf{P}(1)\mathbf{S}_0'$. Hence the limiting distributions of the estimators in this case are similar to those in Theorem 1 and Theorem 2. Using similar arguments as for Theorem 1, we can get the following result.

Theorem S1. *Suppose that the condition of Theorem 1 hold, and* $[\mathbf{I}_d, \mathbf{0}]\boldsymbol{\phi}\boldsymbol{\mu}_0^* = 0$. *Then, as* $n \to \infty$, *it follows that*

$$(a). \quad n(\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}_\perp' \xrightarrow{d} \boldsymbol{\Theta}_0\mathbf{R}_1^*\boldsymbol{\Gamma}_{11}^{*-1},$$

$$(b).\quad n^{1/\alpha}\tilde{L}(n)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \xrightarrow{d} \boldsymbol{\Gamma}_{22}^{-1} vec[\sum_{l=0}^{\infty}\boldsymbol{\gamma}_{0,l}^{'}\mathbf{R}_{2l}],$$

$$when\ \alpha \in (1,2)\ or\ \alpha = 1\ and\ \tilde{L}(n) \to 0,$$

$$(c).\quad n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \xrightarrow{d} -\boldsymbol{\Gamma}_{22}^{-1} vec[\sum_{l=0}^{\infty}\boldsymbol{\gamma}_{0,l}^{'}(\mathbf{F}_{0l}^{*}+\mathbf{F}_{1}^{*}+\mathbf{R}_{2l}^{*})],$$

$$when\ \alpha \in (0,1)\ or\ \alpha = 1\ and\ \tilde{L}(n) \to \infty,$$

$$(d).\quad n^{1-1/\alpha}L^{-1}(n)(\hat{\boldsymbol{\mu}}^{*}-\boldsymbol{\mu}_0^{*}) \xrightarrow{d} \boldsymbol{\Theta}_0\mathbf{P}(1)-\boldsymbol{\Theta}_0\mathbf{R}_1^{*}\boldsymbol{\Gamma}_{11}^{*-1}\xi,$$

where $\mathbf{F}_1^{*} = \sum_{j'=1}^{q}\sum_{i=0}^{j'-1}\sum_{k'=0}^{\infty}\boldsymbol{\Theta}_0^{-1}\boldsymbol{\Theta}_{0,j'}\boldsymbol{\gamma}_{k'}\boldsymbol{\Theta}\mathbf{R}_1^{*}\boldsymbol{\Gamma}_{11}^{*-1}\mathbf{B}_{\perp}\mathbf{S}_{22k'il}$ *and* $\mathbf{F}_{0l}^{*} = \mathbf{R}_1^{*}\boldsymbol{\Gamma}_{11}^{*-1}\boldsymbol{\Gamma}_{12l}^{*'}$.

Similarly, we can obtain the limiting distribution of RLSE of model (S2.1). From (c), we can obtain that the rate of convergence of $\hat{\boldsymbol{\mu}}^{*}$ is slowdown when the tail index α decrease from 2 to 1. However, the estimator of $\boldsymbol{\mu}^{*}$ is not consistent when $\alpha \in (0,1)$.

S3 More Simulation Results

First, we consider the model (4.20) with same $\boldsymbol{\Phi}_1$ and $\boldsymbol{\varepsilon}_t$, while the parameter $\boldsymbol{\Theta}_{\mathbf{1}}$ is generated by the following method:

1. Initialize $\boldsymbol{\Theta}_{\mathbf{1}} = 0$.

2. Randomly generate 2×2 *real numbers from the uniform distribution on the interval* $[-0.9,-0.1] \cup [0.1,0.9]$.

3. Assign the 2×2 *real numbers of Step 2 as* $\boldsymbol{\Theta}_1$*, and detect whether all* *the eigenvalues of* $\mathbf{\Theta}_1$ *lie inside the unit circle. If not, we repeat Step 2.*

Using this method, five different $\mathbf{\Theta}_1$ are generated as showed in Example 1-5 below. In each example, we use 1000 replications and samples $n = 1000$. The tail index $\alpha = 0.5$ and 1.5 are considered. Table S3.1- S3.5 summarizes the sample mean (Mean) and the sample standard deviation (SD) of FLSE and RLSE corresponding to Example 1-5, respectively.

Example 1:

$$\mathbf{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} 0.6865950 & 0.4269934 \\ -0.3325087 & -0.8097146 \end{bmatrix},$$

Example 2:

$$\mathbf{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} -0.1900979 & 0.7447099 \\ -0.1754733 & 0.8343637 \end{bmatrix},$$

Example 3:

$$\mathbf{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} 0.3414708 & 0.4819453 \\ -0.1623931 & -0.1794810 \end{bmatrix},$$

Example 4:

$$\mathbf{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} -0.5958684 & -0.8676538 \\ 0.1798034 & 0.6164981 \end{bmatrix},$$

Example 5:

$$\mathbf{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} 0.5194453 & -0.5188983 \\ 0.3351903 & -0.7247159 \end{bmatrix},$$

Table S3.1: Means and SDs of the FLSE and RLSE of Example 1 when $n = 1000$

			a_1	a_2	b^*	θ_{11}	θ_{21}	θ_{12}	θ_{22}
$\alpha = 0.5$	FLSE	Bias	-0.00017	-0.00034	0.00177	-0.00288	-0.00429	0.00934	0.00589
		SD	0.01197	0.00755	0.01581	0.17569	0.20104	0.22651	0.18958
	RLSE	Bias	-0.00195	-0.00051	0.00241	-0.00477	-0.00053	0.00652	0.00132
		SD	0.01239	0.01193	0.01714	0.18923	0.35106	0.22694	0.20559
$\alpha = 1.5$	FLSE	Bias	0.00012	-0.00057	0.00336	0.00153	0.00004	-0.00008	0.00164
		SD	0.00861	0.00946	0.00721	0.05633	0.05114	0.09447	0.04239
	RLSE	Bias	-0.00317	-0.00090	0.00398	0.00156	0.00092	0.00019	0.00080
		SD	0.01063	0.01009	0.00891	0.05785	0.05508	0.09254	0.04010

Next, we consider model (4.20) with same $\mathbf{\Phi}_1$ and $\mathbf{\Theta}_1$, and $x_t \sim$i.i.d t_2 distribution. The results are given in Table S3.6.

Table S3.2: Means and SDs of the FLSE and RLSE of Example 2 when $n = 1000$

			a_1	a_2	b^*	θ_{11}	θ_{21}	θ_{12}	θ_{22}
$\alpha = 0.5$	FLSE	Bias	0.00030	-0.00062	0.00216	0.01158	0.01301	-0.01148	-0.00499
		SD	0.03049	0.02634	0.02828	0.15437	0.30648	0.17424	0.14532
	RLSE	Bias	-0.00415	0.00176	0.00165	0.00691	0.00812	-0.01221	-0.01127
		SD	0.03841	0.06737	0.06723	0.19254	0.32116	0.27287	0.31860
$\alpha = 1.5$	FLSE	Bias	-0.00257	0.00281	0.00970	0.00492	0.00426	-0.00154	-0.00408
		SD	0.06390	0.03726	0.03130	0.20436	0.05242	0.21264	0.08394
	RLSE	Bias	-0.01249	0.01096	0.00877	0.01234	0.00442	-0.01695	-0.00067
		SD	0.06731	0.05734	0.04587	0.20531	0.06232	0.22304	0.09037

Table S3.3: Means and SDs of the FLSE and RLSE of Example 3 when $n = 1000$

			a_1	a_2	b^*	θ_{11}	θ_{21}	θ_{12}	θ_{22}
$\alpha = 0.5$	FLSE	Bias	0.00043	-0.00074	0.01906	-0.00727	-0.01044	0.00639	0.00926
		SD	0.02703	0.01458	0.07569	0.24714	0.24438	0.29887	0.24070
	RLSE	Bias	-0.01857	-0.00115	0.01939	-0.00624	-0.00113	0.00912	0.00411
		SD	0.07888	0.01607	0.07708	0.25291	0.25381	0.31316	0.31473
$\alpha = 1.5$	FLSE	Bias	0.00016	-0.00040	0.03391	0.00019	-0.00060	0.00092	0.00091
		SD	0.01798	0.01481	0.09231	0.07569	0.06072	0.11869	0.06946
	RLSE	Bias	-0.03386	-0.00110	0.03415	0.00169	0.00396	0.00247	-0.00071
		SD	0.09392	0.01581	0.09303	0.08117	0.07007	0.11969	0.06931

Table S3.4: Means and SDs of the FLSE and RLSE of Example 4 when $n = 1000$

			a_1	a_2	b^*	θ_{11}	θ_{21}	θ_{12}	θ_{22}
$\alpha = 0.5$	FLSE	Bias	-0.00106	0.00036	0.01523	0.00159	-0.00195	0.00946	-0.00136
		SD	0.01595	0.01173	0.11533	0.12154	0.08069	0.24965	0.19051
	RLSE	Bias	-0.01572	0.00061	0.01491	0.00178	0.00135	0.01182	0.00003
		SD	0.08153	0.02268	0.11647	0.12374	0.10068	0.28323	0.24154
$\alpha = 1.5$	FLSE	Bias	-0.00314	-0.00022	0.02594	0.00284	0.00078	0.00629	-0.00414
		SD	0.01849	0.01442	0.06167	0.04035	0.04243	0.07191	0.04810
	RLSE	Bias	-0.02936	0.00058	0.02675	-0.00176	0.00373	0.00658	-0.00061
		SD	0.05985	0.02189	0.06351	0.03521	0.04213	0.07012	0.04485

Table S3.5: Means and SDs of the FLSE and RLSE of Example 5 when $n = 1000$

			a_1	a_2	b^*	θ_{11}	θ_{21}	θ_{12}	θ_{22}
$\alpha = 0.5$	FLSE	Bias	0.00016	-0.00029	-0.00139	0.00147	0.00020	-0.00193	0.00240
		SD	0.01024	0.01065	0.00922	0.03659	0.04575	0.10033	0.14322
	RLSE	Bias	0.00157	-0.00003	-0.00093	-0.00382	-0.00842	-0.00205	-0.00299
		SD	0.01084	0.01021	0.00945	0.07245	0.12054	0.19696	0.20448
$\alpha = 1.5$	FLSE	Bias	-0.00009	-0.00010	-0.00323	0.00208	0.00073	0.00102	0.00061
		SD	0.01358	0.01567	0.00992	0.03344	0.04636	0.06308	0.04869
	RLSE	Bias	0.00272	0.00028	-0.00201	0.00219	0.00131	0.00058	-0.00107
		SD	0.01465	0.01531	0.00855	0.03341	0.04664	0.06516	0.04832

Table S3.6: Means and SDs of the FLSE and RLSE when $n = 1000$

			a_1	a_2	b^*	θ_{11}	θ_{21}	θ_{12}	θ_{22}
$\alpha = 0.5$	FLSE	Bias	0.00127	0.00031	-0.00801	0.00142	-0.00082	0.00251	-0.00126
		SD	0.03030	0.02749	0.02378	0.06291	0.08421	0.10016	0.08918
	RLSE	Bias	0.00920	-0.00443	-0.00531	0.00054	-0.00291	0.01299	-0.00878
		SD	0.03743	0.03289	0.03055	0.07481	0.09195	0.09932	0.09354
$\alpha = 1.5$	FLSE	Bias	0.00111	-0.00023	-0.01426	0.00529	-0.00014	0.00332	-0.00336
		SD	0.03035	0.02989	0.02616	0.05148	0.04742	0.08666	0.08441
	RLSE	Bias	0.01536	-0.00513	-0.00818	0.00086	0.00388	0.01278	-0.01581
		SD	0.03956	0.03226	0.02487	0.05141	0.04866	0.08712	0.08786

Finally, we consider the model

$$\begin{pmatrix} \mathbf{Y}_{1,t} \\ \mathbf{Y}_{2,t} \\ \mathbf{Y}_{3,t} \end{pmatrix} = \mathbf{\Phi}_1 \begin{pmatrix} \mathbf{Y}_{1,t-1} \\ \mathbf{Y}_{2,t-1} \\ \mathbf{Y}_{3,t-1} \end{pmatrix} + \boldsymbol{\varepsilon}_t - \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1}, \tag{S3.2}$$

with $\boldsymbol{\varepsilon}_t = |x_t|^{1/\alpha}(\sin\zeta_t \cos\varphi_t, \sin\zeta_t \sin\varphi_t, \cos\zeta_t)$, where $\varphi_t, \zeta_t \sim i.i.d.\ U[0, 2\pi]$ and are independent of each other. In addition, $\mathbf{\Phi}_1$ and $\mathbf{\Theta}_1$is specified as follows:

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{21} & \phi_{22} & \phi_{33} \end{bmatrix} = \begin{bmatrix} 0.5 & -0.25 & 0.5 \\ 0.1 & 1.05 & -0.1 \\ 0.2 & 0.1 & 0.8 \end{bmatrix} \text{ and}$$

$$\boldsymbol{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{bmatrix} = \begin{bmatrix} 0.5 & -0.2 & 0.1 \\ 0.1 & -0.3 & 0.4 \\ 0.2 & 0.4 & 0.2 \end{bmatrix},$$

Hence, we have

$$\mathbf{C} = \begin{bmatrix} -0.5 & -0.25 & 0.5 \\ 0.1 & 0.05 & -0.1 \\ 0.2 & 0.1 & -0.2 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.1 \\ 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & -1 \end{bmatrix} = \mathbf{AB},$$

with $\mathbf{A} = [a_{11}, a_{21}, a_{31}]' = [-0.5, 0.1, 0.2]'$, and $\mathbf{B} = [1, b_{11}, b_{12}] = [1, 0.5, -1]$. For each of 1000 replications, samples of series length $n = 1000$ are generated. The tail index $\alpha = 0.5$ and 1.5 are considered. Table S3.7 and Table S3.8 summarize the sample mean (Mean) and the sample standard deviation (SD) of FLSE and RLSE. Overall, our method perform better when $\alpha = 1.5$ compared with the case when $\alpha = 0.5$.

Table S3.7: Finite-sample properties of **A** and **B** when $n = 1000$

		$\alpha = 0.5$					$\alpha = 1.5$				
		a_{11}	a_{21}	a_{31}	b_{11}	b_{12}	a_{11}	a_{21}	a_{31}	b_{11}	b_{12}
FLSE	Bias	-0.02519	-0.02624	0.03244	0.03083	0.08267	-0.00199	0.00128	-0.00068	0.00126	0.01235
	SD	0.33611	0.22068	0.31186	0.40620	0.47338	0.02159	0.02891	0.02746	0.03341	0.06885
RLSE	Bias	-0.02391	-0.02542	0.02739	0.04175	0.09019	-0.00191	-0.01026	-0.00075	0.00189	0.01206
	SD	0.39324	0.39500	0.30531	0.46255	0.48667	0.03019	0.03828	0.02840	0.03549	0.06843

Table S3.8: Finite-sample properties of $\boldsymbol{\Theta}_1$ when $n = 1000$

			θ_{11}	θ_{21}	θ_{31}	θ_{12}	θ_{22}	θ_{32}	θ_{13}	θ_{23}	θ_{33}
$\alpha = 0.5$	FLSE	Bias	-0.00080	-0.00407	0.01728	0.00439	0.00493	0.02827	-0.01897	-0.02378	0.01432
		SD	0.18944	0.19683	0.25883	0.20054	0.22991	0.27907	0.19304	0.23873	0.29695
	RLSE	Bias	0.03116	0.01473	0.08085	-0.00743	-0.04107	-0.01037	-0.02002	-0.00709	-0.00562
		SD	0.27779	0.33934	0.36223	0.25917	0.32814	0.37253	0.22436	0.30370	0.32854
$\alpha = 1.5$	FLSE	Bias	-0.00243	0.00331	0.00226	0.00157	0.00417	0.00412	-0.00183	0.00208	0.00273
		SD	0.04863	0.06898	0.09168	0.05998	0.06079	0.08208	0.03817	0.05085	0.05903
	RLSE	Bias	-0.00179	0.00328	0.00670	0.00193	0.00273	-0.00144	-0.00031	0.00255	0.00216
		SD	0.06280	0.07447	0.09890	0.06092	0.07104	0.09874	0.04036	0.05381	0.06435

S4 Proofs of Lemma 1 and Lemma 3

Proof of Lemma 1. $\partial\boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})/\partial\mathbf{c}$ satisfies

$$\partial\boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})/\partial\mathbf{c} = -(\mathbf{Y}_{t-1}\otimes\mathbf{I}_m) + \sum_{j=1}^{q}(\partial\boldsymbol{\varepsilon}_{t-j}'(\boldsymbol{\beta})/\partial\mathbf{c})\boldsymbol{\Theta}_j'.$$

Premultiplying the equation by $\mathbf{Q}_1'\otimes\mathbf{I}_m$ gives

$$(\mathbf{Q}_1'\otimes\mathbf{I}_m)\frac{\partial\boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})}{\partial\mathbf{c}} = -(\mathbf{Z}_{1,t-1}\otimes\mathbf{I}_m) + (\mathbf{Q}_1'\otimes\mathbf{I}_m)\sum_{j=1}^{q}(\frac{\partial\boldsymbol{\varepsilon}_{t-j}'(\boldsymbol{\beta})}{\partial\mathbf{c}})\boldsymbol{\Theta}_j'.$$

Let $\mathbf{V}_t = \partial\boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})/\partial\mathbf{c}$, so that

$$(\mathbf{Q}_1' \otimes \mathbf{I}_m)\mathbf{V}_t = -(\mathbf{Z}_{1,t-1} \otimes \mathbf{I}_m) + (\mathbf{Q}_1' \otimes \mathbf{I}_m)\sum_{j=1}^{q}\mathbf{V}_{t-j}\boldsymbol{\Theta}_j'. \tag{S4.1}$$

Set $\mathbf{V}_i = 0$ for $i = 0, -1, \ldots, -q+1$. Then, $\mathbf{Z}_{1,0} = 0$ implies that $\mathbf{V}_1 = 0$. Also, let $\dot{\mathbf{V}}_t = \mathbf{V}_t - \mathbf{V}_{t-1}$, so that $\mathbf{V}_{t-j} = \mathbf{V}_t - \sum_{l=0}^{j-1}\dot{\mathbf{V}}_{t-l}$ for $j = 1, \ldots, q$. Thus, on substitution in (S4.1), it follows that

$$(\mathbf{Q}_1' \otimes \mathbf{I}_m)\mathbf{V}_t(\mathbf{I}_m - \sum_{j=1}^{q}\boldsymbol{\Theta}_j') = -(\mathbf{Z}_{1,t-1} \otimes \mathbf{I}_m) - (\mathbf{Q}_1' \otimes \mathbf{I}_m)\sum_{j=1}^{q}(\sum_{l=0}^{j-1}\dot{\mathbf{V}}_{t-l})\boldsymbol{\Theta}_j'.$$

So we obtain

$$(\mathbf{Q}_1' \otimes \mathbf{I}_m)\frac{\partial\boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})}{\partial\mathbf{c}} = -(\mathbf{Z}_{1,t-1} \otimes \boldsymbol{\Theta}'^{-1}) + \mathbf{R}_t,$$

where $\mathbf{R}_t = -(\mathbf{Q}_1' \otimes \mathbf{I}_m)(\sum_{j=1}^{q}\sum_{l=0}^{j-1}\dot{\mathbf{V}}_{t-l}\boldsymbol{\Theta}_j')\boldsymbol{\Theta}'^{-1}$. From (S4.1), we have

$$(\mathbf{Q}_1' \otimes \mathbf{I}_m)\dot{\mathbf{V}}_t = -(\mathbf{Q}_1'\mathbf{W}_{t-1} \otimes \mathbf{I}_m) + (\mathbf{Q}_1' \otimes \mathbf{I}_m)\sum_{j=1}^{q}\dot{\mathbf{V}}_{t-j}\boldsymbol{\Theta}_j'. \tag{S4.2}$$

Thus, $\{\mathrm{vec}(\dot{\mathbf{V}}_t)\}$ is generated by a stationary process. From (S4.2), we have

$$(\mathbf{I}_m - \sum_{j=1}^{q}\boldsymbol{\Theta}_j L^j)\dot{\mathbf{V}}_t'(\mathbf{Q}_1 \otimes \mathbf{I}_m) = -(\mathbf{W}_{t-1}'\mathbf{Q}_1 \otimes \mathbf{I}_m),$$

and together with (2.8) it follows that

$$\dot{\mathbf{V}}_t'(\mathbf{Q}_1 \otimes \mathbf{I}_m) = -\sum_{k=0}^{\infty} \boldsymbol{\gamma}_k(\mathbf{W}_{t-1-k}' \otimes \mathbf{I}_m)(\mathbf{Q}_1 \otimes \mathbf{I}_m).$$

Thus, for any $j = 1, \ldots, q$,

$$\mathbf{R}_t = \sum_{j=1}^{q} \sum_{l=0}^{j-1} \sum_{k=0}^{\infty} (\mathbf{Q}_1^{'} \mathbf{W}_{t-1-k-l} \otimes \boldsymbol{\gamma}_k^{'} \boldsymbol{\Theta}_j^{'}) \boldsymbol{\Theta}^{'-1}.$$

This completes the proof. □

Proof of Lemma 3. $\partial \boldsymbol{\varepsilon}_t^{'}(\boldsymbol{\eta})/\partial \mathbf{b}$ satisfies

$$\frac{\partial \boldsymbol{\varepsilon}_t^{'}(\boldsymbol{\eta})}{\partial \mathbf{b}} = -(\mathbf{H}^{'} \mathbf{Y}_{t-1} \otimes \mathbf{A}^{'}) + \sum_{j=1}^{q} (\frac{\partial \boldsymbol{\varepsilon}_{t-j}^{'}(\boldsymbol{\eta})}{\partial \mathbf{b}}) \boldsymbol{\Theta}_j^{'}.$$

Let $\mathbf{V}_t^* = \partial \boldsymbol{\varepsilon}_t^{'}(\boldsymbol{\eta})/\partial \mathbf{b}$, so that

$$\mathbf{V}_t^* = -(\mathbf{H}^{'} \mathbf{Y}_{t-1} \otimes \mathbf{A}^{'}) + \sum_{j=1}^{q} \mathbf{V}_{t-j}^* \boldsymbol{\Theta}_j^{'}. \tag{S4.3}$$

Set $\mathbf{V}_i^* = 0$ for $i = 0, -1, \ldots, -q+1$. Denote $\dot{\mathbf{V}}_t^* = \mathbf{V}_t^* - \mathbf{V}_{t-1}^*$, so that $\mathbf{V}_{t-j}^* = \mathbf{V}_t^* - \sum_{l=0}^{j-1} \dot{\mathbf{V}}_{t-l}^*$, for $j = 1, \ldots, q$. Thus, on substitution in (S4.3),

$$\mathbf{V}_t^*(\mathbf{I}_m - \sum_{j=1}^{q} \boldsymbol{\Theta}_j^{'}) = -(\mathbf{H}^{'} \mathbf{Y}_{t-1} \otimes \mathbf{A}^{'}) - \sum_{j=1}^{q} (\sum_{l=0}^{j-1} \dot{\mathbf{V}}_{t-l}^*) \boldsymbol{\Theta}_j^{'}.$$

Hence

$$\frac{\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\eta})}{\partial \mathbf{b}} = -(\mathbf{H}'\mathbf{Y}_{t-1} \otimes \mathbf{A}'\boldsymbol{\Theta}'^{-1}) + \mathbf{R}_t^*,$$

where $\mathbf{R}_t^* = -(\sum_{j=1}^{q}\sum_{l=0}^{j-1}\dot{\mathbf{V}}_{t-l}^*\boldsymbol{\Theta}_j')\boldsymbol{\Theta}'^{-1}$. From (S4.3), we have

$$\dot{\mathbf{V}}_t^* = -(\mathbf{H}'\mathbf{W}_{t-1} \otimes \mathbf{A}') + \sum_{j=1}^{q}\dot{\mathbf{V}}_{t-j}^*\boldsymbol{\Theta}_j',$$

which together with (2.8) implies $\dot{\mathbf{V}}_t^{*\prime} = -\sum_{k=0}^{\infty}\boldsymbol{\gamma}_k L^k[(\mathbf{H}'\mathbf{W}_{t-1})' \otimes \mathbf{A}]$. So

$$\dot{\mathbf{V}}_{t-l}^*\boldsymbol{\Theta}_j' = -\sum_{k=0}^{\infty}(\mathbf{H}'\mathbf{W}_{t-1-k-l} \otimes \mathbf{A}'\boldsymbol{\gamma}_k'\boldsymbol{\Theta}_j').$$

Therefore, we have

$$\frac{\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\eta})}{\partial \mathbf{b}} = -\left[(\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1} + \bar{\mathbf{B}}_2\mathbf{Z}_{2,t-1}) \otimes \mathbf{A}'\boldsymbol{\Theta}'^{-1}\right] + \mathbf{R}_t^*,$$

where $\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t}$ is the nonstationary component of $\mathbf{H}'\mathbf{Y}_t$, and

$$\mathbf{R}_t^* = \sum_{j=1}^{q}\sum_{l=0}^{j-1}\sum_{k=0}^{\infty}(\mathbf{H}'\mathbf{W}_{t-1-k-l} \otimes \mathbf{A}'\boldsymbol{\gamma}_k'\boldsymbol{\Theta}_j')\boldsymbol{\Theta}'^{-1}.$$

This completes the proof. □

S5 Proofs of (6.30) in the Proof of Theorem 1

Proof. Note that

$$[\mathbf{U}_{t-1}(\bar{\boldsymbol{\beta}}) - \mathbf{U}_{t-1}]\mathbf{U}'_{t-1} = [\mathbf{0}, \boldsymbol{\varepsilon}'_{t-1} - \boldsymbol{\varepsilon}'_{t-1}(\bar{\boldsymbol{\beta}}), \ldots, \boldsymbol{\varepsilon}'_{t-q} - \boldsymbol{\varepsilon}'_{t-q}(\bar{\boldsymbol{\beta}})]'\mathbf{U}'_{t-1}.$$

Since

$$\begin{aligned}(\mathbf{I}_m - \sum_{j=1}^{q} \bar{\boldsymbol{\Theta}}_j L^j)[\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\beta}})] =&(\bar{\mathbf{C}} - \mathbf{C}_0)\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1}(\bar{\boldsymbol{\Phi}}^*_j - \boldsymbol{\Phi}^*_{0,j})\mathbf{W}_{t-j} \\ &+ \sum_{j=1}^{q}(\boldsymbol{\Theta}_{0,j} - \bar{\boldsymbol{\Theta}}_j)\boldsymbol{\varepsilon}_{t-j},\end{aligned}$$

we have

$$\begin{aligned}\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\beta}}) =&(\mathbf{I}_m - \sum_{j=1}^{q} \bar{\boldsymbol{\Theta}}_j L^j)^{-1}[(\bar{\mathbf{C}} - \mathbf{C}_0)\mathbf{P}_1\mathbf{Z}_{1,t-1} + (\bar{\mathbf{C}} - \mathbf{C}_0)\mathbf{P}_2\mathbf{Z}_{2,t-1} \\ &+ \sum_{j=1}^{p-1}(\bar{\boldsymbol{\Phi}}^*_j - \boldsymbol{\Phi}^*_{0,j})\mathbf{W}_{t-j} + \sum_{j=1}^{q}(\boldsymbol{\Theta}_{0,j} - \bar{\boldsymbol{\Theta}}_j)\boldsymbol{\varepsilon}_{t-j}]. \qquad \text{(S5.4)}\end{aligned}$$

By (2.10), (S5.4), and Lemma 2, we have

$$\frac{1}{a_n^2}\sum_{t=1}^{n}[\mathbf{U}_{t-1}(\bar{\boldsymbol{\beta}}) - \mathbf{U}_{t-1}]\mathbf{U}'_{t-1} = o_p(1). \qquad \text{(S5.5)}$$

Similarly, we have

$$\frac{1}{a_n^2}\sum_{t=1}^{n}[\mathbf{U}_{t-1}(\bar{\boldsymbol{\beta}})-\mathbf{U}_{t-1}][\mathbf{U}_{t-1}(\bar{\boldsymbol{\beta}})-\mathbf{U}_{t-1}]'=o_p(1),$$

which together with (S5.5) implies that

$$\frac{1}{a_n^2}\sum_{t=1}^{n}\mathbf{U}_{t-1-k}(\bar{\boldsymbol{\beta}})\mathbf{U}_{t-1-j}(\bar{\boldsymbol{\beta}})'=\frac{1}{a_n^2}\sum_{t=1}^{n}\mathbf{U}_{t-1-k}\mathbf{U}'_{t-1-j}+o_p(1).$$

□