Statistica Sinica Preprint No: SS-2022-0121	
Title	Statistical Inference for Heavy-tailed and Partially Nonstationary Vector ARMA Models
Manuscript ID	SS-2022-0121
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0121
Complete List of Authors	Feifei Guo and Shiqing Ling
Corresponding Authors	Shiqing Ling
E-mails	maling@ust.hk
Notice: Accepted version subject to English editing.	

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

Abstract: This paper studies the full rank least squares estimator (FLSE) and reduced rank least squares estimator (RLSE) of the heavy-tailed and partially nonstationary ARMA model with the tail index $\alpha \in (0, 2)$. It is shown that the rate of convergence of the FLSE related to the long-run parameters is n (sample size) and that related to the short-term parameters are $n^{1/\alpha}\tilde{L}(n)$ and n, respectively, when $\alpha \in (1, 2)$ and $\in (0, 1)$. Its limiting distribution is a stochastic integral in terms of two stable random processes when $\alpha \in (0, 2)$ for the long-run parameters and is a functional of some stable processes when $\alpha \in (1, 2)$ for the short-run parameters. Based on FLSE, we derive the asymptotic properties of the RLSE. The finite-sample properties of the estimation are examined through a simulation study and an application to three U.S. interest rate series is given.

Key words: ARMA models, Cointegration, Heavy-tailed time series, Full rank LSE, Reduced rank LSE.

1. Introduction

Consider the m-dimensional time series $\{Y_t\}$ generated by the ARMA model:

$$\mathbf{\Phi}(L)\mathbf{Y}_t = \mathbf{\Theta}(L)\boldsymbol{\varepsilon}_t, \tag{1.1}$$

where $\mathbf{\Phi}(z) = \mathbf{I}_m - \sum_{i=1}^{p} \mathbf{\Phi}_i z^i$ and $\mathbf{\Theta}(z) = \mathbf{I}_m - \sum_{i=1}^{q} \mathbf{\Theta}_i z^i$ are matrix polynomials in z of degrees p and q, L is the backshift operator, $\mathbf{\Phi}_i(1 \le i \le p)$ and $\mathbf{\Theta}_i(1 \le i \le q)$ are $m \times m$ matrices, $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, and $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of independent and identically distributed (i.i.d.) m-dimensional vector noises. When all the roots of $\det\{\mathbf{\Phi}(z)\} = 0$ lie outside the unit circle, model (1.1) is stationary, where det{} means the determinant of a matrix. In this case, model (1.1) has been extensively applied in many areas such as economics and finance and its modeling procedure was fully established in Tsay (2014).

When $\det\{\mathbf{\Phi}(z)\} = 0$ has $d < m$ unit roots and the remaining roots outside the unit circle, model (1.1) is called the partially nonstationary ARMA model. Denote $\mathbf{W}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$. Model (1.1) can be rewritten as

$$\mathbf{W}_t = \mathbf{C}\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \mathbf{\Phi}_j^* \mathbf{W}_{t-j} + \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \mathbf{\Theta}_j \boldsymbol{\varepsilon}_{t-j}, \tag{1.2}$$

where $\mathbf{\Phi}_j^* = -\sum_{k=j+1}^{p} \mathbf{\Phi}_k$, $\mathbf{C} = -\mathbf{\Phi}(1) = -(\mathbf{I}_m - \sum_{j=1}^{p} \mathbf{\Phi}_j)$ and $\mathbf{C} = \mathbf{A}\mathbf{B}$ has reduced rank $r = m - d$, where $\mathbf{A}$ and $\mathbf{B}$ are full-rank matrices of dimensions $m \times r$ and $r \times m$. The true values of unknown parameters are $\mathbf{C}_0 = \mathbf{A}_0\mathbf{B}_0$, $\mathbf{\Phi}_{0,j}^*$ and $\mathbf{\Theta}_{0,j}$. Let $\mathbf{A}_\perp$ be an $m \times d$ orthogonal matrix of $\mathbf{A}_0$ i.e. $\mathbf{A}_0^{'}\mathbf{A}_\perp = \mathbf{0}$ and $\mathbf{B}_\perp$ be a $d \times m$ orthogonal matrix of $\mathbf{B}_0$ i.e. $\mathbf{B}_\perp \mathbf{B}_0^{'} = \mathbf{0}$. In model (1.2), each component of $\mathbf{Y}_t$ is nonstationary, but $\mathbf{B}'\mathbf{Y}_t$ is stationary. Each column of $\mathbf{B}$ is called the cointegrating vector of $\mathbf{Y}_t$ and r is called its cointegrating rank. The parameters in model (1.2) can be estimated by the least squares estimation (LSE) under two scenarios: $\mathbf{C}$ is estimated as a whole matrix and as a reduced form $\mathbf{A}\mathbf{B}$, respectively, called the full rank LSE (FLSE) and reduced rank LSE (RLSE). When $E\|\varepsilon_t\|^2 < \infty$, they were studied by Yap and Reinsel (1995), see Tsay and Tiao (1990) for the vector unit root ARMA model.

When $q = 0$, model (1.2) reduces to the vector error correction (VEC) model. The VEC model was introduced by Granger (1983) and Engle and Granger (1987). The early research can be found in Phillips and Durlauf (1986) and Johansen (1995), among many others. Recently, Wang (2014) established a martingale limit theorem for a nonlinear cointegrating regression model. Cai, Gao, and Tjøstheim (2017) introduced a new class of bivariate threshold VAR cointegration models. Zhang, Robinson, and Yao (2019) studied a factor model dealing with high-dimensional nonstationary time series. Almost all research on cointegration systems required the noises to have a second or even high finite moment. Only very few results are available for the VEC model with the heavy-tailed noises.

The heavy-tailed time series do not have a finite second moment. They have been well observed in financial market, engineering, network system and other areas, see in Resnick (1997). Davis and Resnick (1985, 1986) showed the limiting distribution of the LSE of the parameters in a heavy-tailed AR model is the functional of two stable random variables with the rate of convergence much faster than $\sqrt{n}$, see Davis, Knight, and Liu (1992). Mikosch et al. (1995) studied the Whittle estimators for the heavy-tailed ARMA model and this result was extended by She, Mi and Ling (2021) for model (1.1). Davis, Mikosch and Pfaffel (2016) studied the sample covariance matrix of a heavy-tailed multivariate time series. Caner (1998) developed two tests for a special heavy-tailed VEC model. She and Ling (2020) derived the limiting distributions of the FLSE and RLSE of the heavy-tailed VEC model. Guo, Ling, and Mi (2021) proposed a Lasso approach to determine its cointegrating rank and estimate parameters simultaneously.

However, when ε_t is a heavy-tailed noise, no result is available yet for model (1.2). This paper is to fill in this gap. We assume that ε_t satisfies the following condition:

$$nP(\frac{\boldsymbol{\varepsilon}_1}{a_n} \in \cdot) \xrightarrow{v} \mu(\cdot), \tag{1.3}$$

as $n \to \infty$, where μ is a Radon measure on $(R^m, \mathcal{B}^m)$, a_n is an increasing sequence diverging to ∞, and $\xrightarrow{v}$ denotes the vague convergence. (1.3) is also known as a regular variation condition, under which exists $\alpha > 0$ called tail index of $\boldsymbol{\varepsilon}_t$ such that, for any $y > 0$ as $n \to \infty$,

$$nP(\frac{\|\boldsymbol{\varepsilon}_1\|}{a_n} > y) \longrightarrow c_0 y^{-\alpha},$$

where the notation $\|\cdot\|$ denotes the Frobenius norm and c_0 is some constant. When $\alpha \in (0, 2)$, $\boldsymbol{\varepsilon}_1$ does not have a finite covariance matrix and is called a heavy-tailed random vector. For example, when ε_t is defined as

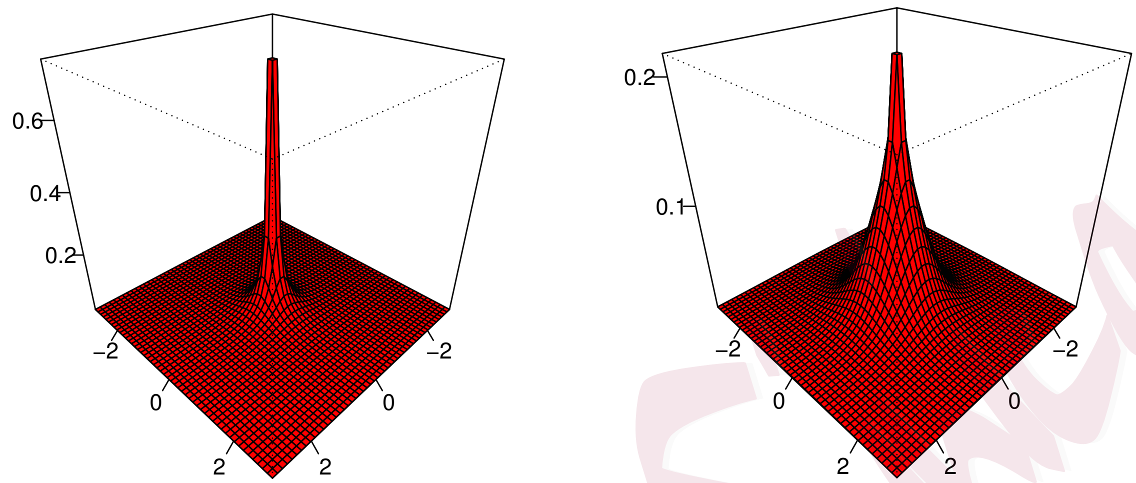
$$\boldsymbol{\varepsilon}_t = |x_t|^{1/\alpha}(\cos\varphi_t, \sin\varphi_t), \tag{1.4}$$

where x_t ∼i.i.d Cauchy distribution and $\varphi_t \sim$ i.i.d $U[0, 2\pi]$, and they are independent, we can show that it satisfies (1.3) and the measure μ has the following density function:

$$f(x, y) = \frac{\alpha(x^2 + y^2)^{\frac{\alpha}{2}-1}}{2\pi^2[1 + (x^2 + y^2)^\alpha]},$$

see Figure 1 (a) and (b) for its plots when $\alpha = 0.8$ and 1.6.

This paper will study the FLSE and the RLSE of model (1.2) under the condition (1.3). It is shown that the rate of convergence of the FLSE related to the long-run parameters is n (sample size) and that related to the short-term parameters are $n^{1/\alpha}\tilde{L}(n)$ and n, respectively,



(a) $\alpha = 0.8$ (b) $\alpha = 1.6$

Figure 1: Density $f(x, y)$

when $\alpha \in (1, 2)$ and $\in (0, 1)$. Its limiting distribution is a stochastic integral in terms of two stable random processes when $\alpha \in (0, 2)$ for the long-run parameters and is a functional of some stable processes when $\alpha \in (1, 2)$ for the short-run parameters. Based on FLSE, we derive the asymptotic properties of the RLSE. Unlike those in Yap and Reinsel (1995), the FLSE and RLSE related to the short-run parameters $\mathbf{A}$ have a different limiting distribution.

The rest of this paper is arranged as follows. Section 2 presents the FLSE and derives its limiting distribution. Section 3 shows the results of the RLSE. Simulation results are reported in section 4 and a real example is given in section 5. All proofs and the results of model (1.2) with a constant term are provided in the appendix and Supplementary Material.

2. Full Rank Estimation

This section studies the FLSE of model (1.2). We first make the following assumption.

Assumption 1. *(i).* $|\mathbf{C}z + \sum_{j=1}^{p-1} \mathbf{\Phi}_j^*(1-z)z^i - \mathbf{I}_m(1-z)| \neq 0$ *if* $|z| < 1$, *(ii). the matrix* $\mathbf{A}_\perp'(\mathbf{I}_m - \sum_{j=1}^{p-1}\mathbf{\Phi}_j^*)\mathbf{B}_\perp'$ *is nonsingular,*

(iii). $det\{\mathbf{\Theta}(z)\} = 0$ *has all its roots lying outside the unit circle and the term* $\mathbf{\Phi}(z)$ *and* $\mathbf{\Theta}(z)$ *are left coprime (i.e. if* $\mathbf{\Phi}(z) = \mathbf{U}(z)\mathbf{\Phi}_1(z)$ *and* $\mathbf{\Theta}(z) = \mathbf{U}(z)\mathbf{\Theta}_1(z)$*, then* $\mathbf{U}(z)$ *is unimodular with constant determinant,where* $\mathbf{\Phi}_1(z)$ *and* $\mathbf{\Theta}_1(z)$ *are matrix polynomials in* z*).*

The condition (i) and (ii) are the same as Ahn and Reinsel (1990) and Johansen (1995), and (iii) is the usual invertible and identification of the vector ARMA model, see Yap and Reinsel (1995). Let $\boldsymbol{\beta} = \text{vec}[\mathbf{C}, \mathbf{\Psi}]$, where $\mathbf{\Psi} = [\mathbf{\Phi}_1^*, \ldots, \mathbf{\Phi}_{p-1}^*, \mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q]$, denote the $(p+q)m^2$ vector of unkonwn parameters, and the vec operator transforms a matrix into a column vector by stacking the columns of matrix below each other. The true value of $\boldsymbol{\beta}$ is denoted by $\boldsymbol{\beta}_0 = \text{vec}[\mathbf{C}_0, \mathbf{\Psi}_0]$, where $\mathbf{\Psi}_0 = [\mathbf{\Phi}_{0,1}^*, \ldots, \mathbf{\Phi}_{0,p-1}^*, \mathbf{\Theta}_{0,1}, \ldots, \mathbf{\Theta}_{0,q}]$. The FLSE of $\boldsymbol{\beta}_0$, denoted by $\hat{\boldsymbol{\beta}} = \text{vec}[\hat{\mathbf{C}}, \hat{\mathbf{\Phi}}_1^*, \ldots, \hat{\mathbf{\Phi}}_{p-1}^*, \hat{\mathbf{\Theta}}_1, \ldots, \hat{\mathbf{\Theta}}_q]$, is the minimizer of the objective function:

$$L(\boldsymbol{\beta}) = \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})\boldsymbol{\varepsilon}_t(\boldsymbol{\beta}),$$

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\beta}) = \mathbf{W}_t - ([\mathbf{Y}_{t-1}', \mathbf{W}_{t-1}', \ldots, \mathbf{W}_{t-p+1}', -\boldsymbol{\varepsilon}_{t-1}'(\boldsymbol{\beta}), \ldots, -\boldsymbol{\varepsilon}_{t-q}'(\boldsymbol{\beta})] \otimes \mathbf{I}_m)\boldsymbol{\beta}$, and $\otimes$ denotes the Kronecker product. The partial derivatives of $L(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are given by

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{t=1}^{n} \frac{\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}_t(\boldsymbol{\beta}) = -\sum_{t=1}^{n} \mathbf{X}_{t-1}(\boldsymbol{\beta})\boldsymbol{\varepsilon}_t(\boldsymbol{\beta}),$$

where $\mathbf{X}_{t-1}(\boldsymbol{\beta}) = -\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ is a $(p+q)m^2 \times m$ matrix. Because the equations $\partial L(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = 0$ are nonlinear in $\boldsymbol{\beta}$ (except when $q = 0$), iterative numerical procedures are needed to obtain the solutions to these equations. Following Yap and Reinsel (1995), we obtain the FLSE of

$\boldsymbol{\beta}$ by the following one-step iteration

$$\hat{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}} + \left[\sum_{t=1}^{n} \mathbf{X}_{t-1}(\bar{\boldsymbol{\beta}})\mathbf{X}'_{t-1}(\bar{\boldsymbol{\beta}})\right]^{-1} \left[\sum_{t=1}^{n} \mathbf{X}_{t-1}(\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\beta}})\right],$$

where $\bar{\boldsymbol{\beta}}$ is an initial estimator.

We next consider the asymptotic properties of the FLSE. Let $\mathbf{P}$ and $\mathbf{Q} = \mathbf{P}^{-1}$ be $m \times m$ matrices. Partition $\mathbf{Q}' = [\mathbf{Q}_1, \mathbf{Q}_2]$ and $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2]$ such that $\mathbf{Q}_1$ and $\mathbf{P}_1$ are $m \times d$ matrices and $\mathbf{Q}_2$ and $\mathbf{P}_2$ are $m \times r$ matrices. Denote $\mathbf{Q}_1 = \mathbf{B}'_\perp$, $\mathbf{Q}_2 = \mathbf{B}'_0$, $\mathbf{P}_1 = \bar{\mathbf{B}}'_\perp$ and $\mathbf{P}_2 = \bar{\mathbf{B}}'$, where $\bar{\mathbf{B}}_\perp = (\mathbf{B}_\perp \mathbf{B}'_\perp)^{-1}\mathbf{B}_\perp$ and $\bar{\mathbf{B}} = (\mathbf{B}_0\mathbf{B}'_0)^{-1}\mathbf{B}_0$. Define $\mathbf{Z}_t = \mathbf{Q}\mathbf{Y}_t$ so that $\mathbf{Z}_{1,t} = \mathbf{Q}'_1\mathbf{Y}_t$ and $\mathbf{Z}_{2,t} = \mathbf{Q}'_2\mathbf{Y}_t$. To derive the asymptotic properties of the FLSE, we rewrite model (1.2) as

$$\mathbf{W}_t = \mathbf{C}\mathbf{P}_1\mathbf{Z}_{1,t-1} + \mathbf{C}\mathbf{P}_2\mathbf{Z}_{2,t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j^*\mathbf{W}_{t-j} + \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \boldsymbol{\Theta}_j\boldsymbol{\varepsilon}_{t-j},$$

where $\{\mathbf{Z}_{2,t}\}$ and $\{\mathbf{W}_t\}$ are stationary and have the following representations:

$$\mathbf{W}_t = \sum_{j=0}^{\infty} \mathbf{B}_j\boldsymbol{\varepsilon}_{t-j} \text{ and } \mathbf{Z}_{2,t} = \sum_{j=0}^{\infty} \mathbf{C}_j\boldsymbol{\varepsilon}_{t-j}, \tag{2.5}$$

with $\mathbf{B}_j = O(\rho^j)$ and $\mathbf{C}_j = O(\rho^j)$ and $\rho \in (0,1)$, but $\{\mathbf{Z}_{1,t}\}$ is purely nonstationary; see Yap and Reinsel (1995) and Johansen (1995)

Let $\mathbf{c}^* = \text{vec}(\mathbf{C}\mathbf{P}_1)$, which is an md-dimensional zero vector. Also, let $\mathbf{Q}^* = \text{Diag}[\mathbf{Q} \otimes \mathbf{I}_m, \mathbf{I}_{(p-1+q)m^2}]$, $\mathbf{Q}_2^* = \text{Diag}[\mathbf{Q}'_2 \otimes \mathbf{I}_m, \mathbf{I}_{(p-1+q)m^2}]$, $\mathbf{c} = \text{vec}(\mathbf{C})$, and $\boldsymbol{\theta} = [\{\text{vec}(\mathbf{C}\mathbf{P}_2)\}', \boldsymbol{\psi}']'$, where $\boldsymbol{\psi} = \text{vec}[\boldsymbol{\Psi}]$. Then, partition $\mathbf{Q}^*\mathbf{X}_{t-1}(\boldsymbol{\beta})$ into two parts corresponding to $\mathbf{c}^*$ and $\boldsymbol{\theta}$:

$$\mathbf{Q}^*\mathbf{X}_{t-1}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{Z}^*_{1,t-1}(\boldsymbol{\beta}) \\ \mathbf{U}^*_{t-1}(\boldsymbol{\beta}) \end{bmatrix}, \text{ where } \mathbf{Z}^*_{1,t-1}(\boldsymbol{\beta}) = -(\mathbf{Q}'_1 \otimes \mathbf{I}_m)\frac{\partial \boldsymbol{\varepsilon}'_t(\boldsymbol{\beta})}{\partial \mathbf{c}}, \tag{2.6}$$

$\mathbf{U}_{t-1}^*(\boldsymbol{\beta})$ is the $[rm + (p-1+q)m^2] \times m$ matrix,

$$\mathbf{U}_{t-1}^*(\boldsymbol{\beta}) = \mathbf{Q}_2^* \mathbf{X}_{t-1}(\boldsymbol{\beta}) = -\mathbf{Q}_2^* \frac{\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{bmatrix} -(\mathbf{Q}_2' \otimes \mathbf{I}_m)\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})/\partial \mathbf{c} \\ -\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\beta})/\partial \boldsymbol{\psi} \end{bmatrix}.$$

The matrices $\mathbf{U}_{t-1}^*(\boldsymbol{\beta})$ satisfy the recursive equations

$$(\mathbf{I}_m - \sum_{j=1}^{q} \boldsymbol{\Theta}_j L^j)\mathbf{U}_{t-1}^{*'}(\boldsymbol{\beta}) = \mathbf{U}_{t-1}'(\boldsymbol{\beta}) \otimes \mathbf{I}_m, \tag{2.7}$$

where $\mathbf{U}_{t-1}(\boldsymbol{\beta}) = [\mathbf{Z}_{2,t-1}', \mathbf{W}_{t-1}', \ldots, \mathbf{W}_{t-p+1}', -\boldsymbol{\varepsilon}_{t-1}'(\boldsymbol{\beta}), \ldots, -\boldsymbol{\varepsilon}_{t-q}'(\boldsymbol{\beta})]'$. Let

$$\boldsymbol{\Theta}(L)^{-1} = (\mathbf{I}_m - \sum_{j=1}^{q} \boldsymbol{\Theta}_j L^j)^{-1} \equiv \boldsymbol{\gamma}(L) = \sum_{k=0}^{\infty} \boldsymbol{\gamma}_k L^k, \tag{2.8}$$

where $\boldsymbol{\gamma}_k = O(\rho^k)$ for some $\rho \in (0,1)$, and

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

From (2.7) and (2.8), we have

$$\mathbf{U}_{t-1}^*(\boldsymbol{\beta}) = \sum_{k=0}^{\infty} (\mathbf{U}_{t-1-k}(\boldsymbol{\beta}) \otimes \mathbf{I}_m)\boldsymbol{\gamma}_k'. \tag{2.9}$$

Denote $\mathbf{U}_{t-1} = \mathbf{U}_{t-1}(\boldsymbol{\beta}_0) = [\mathbf{Z}_{2,t-1}', \mathbf{W}_{t-1}', \ldots, \mathbf{W}_{t-p+1}', -\boldsymbol{\varepsilon}_{t-1}', \ldots, -\boldsymbol{\varepsilon}_{t-q}']'$. By Lemma 1 in the appendix, $\mathbf{Z}_{1,t-1}^*(\boldsymbol{\beta}) = (\mathbf{Z}_{1,t-1} \otimes \boldsymbol{\Theta}'^{-1}) - \mathbf{R}_t$. By Theorem 4.2 in Johansen (1995) and

Assumption 1, we have the following expansions

$$\mathbf{U}_t = \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-i} \text{ and } \mathbf{Z}_{1,t} = [\mathbf{I}_d, \mathbf{0}] \sum_{i=1}^{t} \sum_{j=0}^{\infty} \boldsymbol{\phi}_j \boldsymbol{\varepsilon}_{i-j},$$

where $\mathbf{A}_i = O(\rho^i)$ and $\boldsymbol{\phi}_i = O(\rho^i)$ with some $\rho \in (0,1)$ are $[m(p-1+q)+r] \times m$ matrix and $m \times m$ matrix, respectively.

The estimator of $\boldsymbol{\beta}$ obtained using the iterative relations will be consistent if the initial estimator $\bar{\boldsymbol{\beta}}$ satisfies

$$\mathbf{D}_n \mathbf{Q}^{*\prime -1}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1), \tag{2.10}$$

where

$$\mathbf{D}_n = \begin{cases} n\mathbf{I}_{(p+q)m^2} & \text{if } \alpha \in (0,1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty, \\ \operatorname{diag}(n\mathbf{I}_{dm}, n^{1/\alpha}\tilde{L}(n)\mathbf{I}_{rm+(p-1+q)m^2}) & \text{if } \alpha \in (1,2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0. \end{cases}$$

In the last part of this section we will discuss the instrumental variable approach and Whittle parameter estimation for obtaining $\bar{\boldsymbol{\beta}}$. With these consistent initial estimators, we obtain the asymptotic representation

$$\begin{aligned} \mathbf{D}_n \mathbf{Q}^{*\prime -1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) =& [\mathbf{Q}^* \sum_{t=1}^{n} \mathbf{X}_{t-1}(\bar{\boldsymbol{\beta}}) \mathbf{X}_{t-1}'(\bar{\boldsymbol{\beta}}) \mathbf{Q}^{*\prime} \mathbf{D}_n^{-1}]^{-1} \\ & \{\mathbf{Q}^* \sum_{t=1}^{n} \mathbf{X}_{t-1}(\bar{\boldsymbol{\beta}})[\mathbf{X}_{t-1}'(\bar{\boldsymbol{\beta}})(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\beta}})]\}. \end{aligned} \tag{2.11}$$

Denote $\hat{\mathbf{c}}^* = \operatorname{vec}(\hat{\mathbf{C}}\mathbf{P}_1)$ and $\hat{\boldsymbol{\theta}} = [\{\operatorname{vec}(\hat{\mathbf{C}}\mathbf{P}_2)\}', \hat{\boldsymbol{\psi}}']'$. Then $\mathbf{Q}^{*\prime -1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = [(\hat{\mathbf{c}}^* - \mathbf{c}^*)', (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})']'$. $\mathbf{c}_0^*$ and $\boldsymbol{\theta}_0$ are the true values of $\mathbf{c}^*$ and $\boldsymbol{\theta}$, respectively. We now give the rate of convergence and the limiting distribution of FLSE as follows.

Theorem 1. *Suppose that (1.3) and Assumptions 1 hold, $\boldsymbol{\varepsilon}_t$ has a symmetric distribution and $E\|\boldsymbol{\varepsilon}_1\|^\alpha = \infty$ with $\alpha \in (0,2)$. If the initial estimator $\bar{\boldsymbol{\beta}}$ satisfy (2.10), then, as $n \to \infty$,*

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

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

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

$$(c).\quad n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \stackrel{d}{\longrightarrow} -\boldsymbol{\Gamma}_{22}^{-1} vec[\sum_{l=0}^{\infty} \boldsymbol{\gamma}_{0,l}'(\mathbf{F}_0\mathbf{S}_{12l} + \mathbf{F}_{1l})],$$

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

where $\boldsymbol{\Theta}_0 = \mathbf{I}_m - \sum_{j=1}^q \boldsymbol{\Theta}_{0,j}$, $\mathbf{F}_{1l} = \sum_{j'=1}^{q}\sum_{i=0}^{j'-1}\sum_{k'=0}^{\infty} \boldsymbol{\Theta}_0^{-1}\boldsymbol{\Theta}_{0,j'}\gamma_{0,k'}\boldsymbol{\Theta}_0\mathbf{R}_1\mathbf{S}_{11}^{-1}\mathbf{B}_\perp\mathbf{S}_{22k'il}$, $\mathbf{F}_0 = \mathbf{R}_1\mathbf{S}_{11}^{-1}$, $\boldsymbol{\Gamma}_{22} = \sum_{k=0}^\infty\sum_{j=0}^\infty \mathbf{S}_{22kj}\otimes\boldsymbol{\gamma}_{0,k}'\gamma_{0,j}$, $\tilde{L}(n)$ is a slowly variation function, and $\mathbf{R}_1$, $\mathbf{S}_{11}$, $\mathbf{R}_{2l}$, $\mathbf{S}_{22kj}$, $\mathbf{S}_{12l}$ and $\mathbf{S}_{22k'il}$ are defined as Lemma 2 in the appendix.

From Theorem 1, we see that the rate of convergence of the FLSE related to the long-term parameters $\mathbf{B}$ is always n, while the one related to the short-term parameters $(\mathbf{A}, \boldsymbol{\Psi})$ depends on the tail index α. When $q = 0$, the result is reduced to Theorem 2.1 in She and Ling (2020). The symmetry of $\boldsymbol{\varepsilon}_1$ is to use Theorem B.1 in She and Ling (2020) and Theorem 2 in She, Mi and Ling (2021). When $\boldsymbol{\varepsilon}_1$ is asymmetric, some shift terms need to be added into those limits in Lemma 2 in Appendix and the corresponding results of Theorem 1 can be derived. But in this case, they will be more complicated, see Theorem 3.3 of Davis and Resnick (1986) when $m = 1$.

We now discuss the initial values of our estimation. The instrumental variable approach was discussed by Hall (1989) and Pantula and Hall (1991) for a consistent estimator of

AR parameters in the univariate ARMA(p,q) model. It was applied for model (1.1) to obtain a consistent estimator for the AR parameters in Yap and Reinsel (1995). For illustration, we consider the ARMA(1,1) model, $\mathbf{W}_t = \mathbf{C}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$. Using the instrumental variable $\mathbf{Y}_{t-2}$ for $\mathbf{Y}_{t-1}$, we can show that the resulting estimator of $\mathbf{C}$, given by $\bar{\mathbf{C}} = (\sum_{t=3}^n \mathbf{W}_t\mathbf{Y}'_{t-2})(\sum_{t=3}^n \mathbf{Y}_{t-1}\mathbf{Y}'_{t-2})^{-1}$, is consistent and

$$n(\bar{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}'_\perp \xrightarrow{d} \boldsymbol{\Theta}_0\mathbf{R}_1\mathbf{S}_{11}^{-1}, \tag{2.12}$$

which is the same as the limiting distribution given in Theorem 1 of $n(\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}'_\perp$ for ARMA(1,1) model. Also,

$$n(\bar{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}' \xrightarrow{d} \mathbf{F}_1 \quad when\ \alpha \in (0,1)\ or\ \alpha = 1\ and\ \tilde{L}(n) \to \infty,$$
$$n^{1/\alpha}\tilde{L}(n)(\bar{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}' \xrightarrow{d} \mathbf{F}_2, \quad when\ \alpha \in (1,2)\ or\ \alpha = 1\ and\ \tilde{L}(n) \to 0,$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are functionals of some stable processes, which can be shown to be different from that in Theorem 1 (b)-(c). Since the limiting distributions of the initial estimator on the direction $\bar{\mathbf{B}}$ and that in Theorem 1(b)-(c) are not normal, it hard to claim which one is more efficient. Furthermore, we do not know the rank r in practice. The FLSE will be used together with the RLSE to determine r as for LRT in (3.19). Based on the consistent estimator $\bar{\mathbf{C}}$, $n^{1/\alpha}L(n)$-consistent estimators of the MA parameters can be obtained by Whittle estimation, see She, Mi and Ling (2021), of the MA(1) process $\mathbf{N}_t = \mathbf{W}_t - \mathbf{C}\mathbf{Y}_{t-1} = \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$, using $\bar{\mathbf{N}}_t = \mathbf{W}_t - \bar{\mathbf{C}}\mathbf{Y}_{t-1}$. This method can be readily extended to the higher-order ARMA(p,q) case to obtain consistent initial estimator $\bar{\boldsymbol{\beta}}$ satisfies (2.10).

3. Reduced-Rank Estimation

This section considers the RLSE of model (1.2), in which $\mathbf{C}$ has the reduced form $\mathbf{C} = \mathbf{AB}$. Let $\mathbf{H}'$ denote the $d \times m$ matrix $[\mathbf{0}, \mathbf{I}_d]$. Since $\mathbf{Y}_t$ has cointegrating rank r, $\mathbf{Y}_t$ can always be re-arranged so that the d-dimensional series $\mathbf{Y}_{2,t} = \mathbf{H}'\mathbf{Y}_t$ is not cointegrated. For unique parametrization, we assume $\mathbf{B} = [\mathbf{I}_r, \mathbf{B}^*]$, where $\mathbf{B}^*$ is an $r \times d$ matrix of unknown parameters. The model (1.2) becomes

$$\mathbf{W}_t = \mathbf{A}[\mathbf{I}_r, \mathbf{B}^*]\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \mathbf{\Phi}_j^* \mathbf{W}_{t-j} + \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \mathbf{\Theta}_j \boldsymbol{\varepsilon}_{t-j}. \tag{3.13}$$

Let $\mathbf{b} = \mathrm{vec}(\mathbf{B}^*)$ and $\boldsymbol{\delta} = \mathrm{vec}[\mathbf{A}, \mathbf{\Phi}_1^*, \ldots, \mathbf{\Phi}_{p-1}^*, \mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q]$. Then $\boldsymbol{\eta} = (\mathbf{b}', \boldsymbol{\delta}')'$ is the vector of unknown parameters with dimension $f = rd + mr + (p-1+q)m^2$ and its true value is denoted by $\boldsymbol{\eta}_0 = (\mathbf{b}_0', \boldsymbol{\delta}_0')'$. The RLSE of $\boldsymbol{\eta}$ is the minimizer of the objective function $L(\boldsymbol{\eta}) = \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t'(\boldsymbol{\eta})\boldsymbol{\varepsilon}_t(\boldsymbol{\eta})$, where

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

Denote the RLSE of $\boldsymbol{\eta}$ by $\tilde{\boldsymbol{\eta}} = (\tilde{\mathbf{b}}', \tilde{\boldsymbol{\delta}}')'$ and the estimator of $\mathbf{B}^*$, $\mathbf{A}$, $\mathbf{\Phi}_i^*$ and $\mathbf{\Theta}_i$ by $\tilde{\mathbf{B}}^*$, $\tilde{\mathbf{A}}$, $\tilde{\mathbf{\Phi}}_i^*$ and $\tilde{\mathbf{\Theta}}_i$. Then, the $f \times m$ matrices $\mathbf{X}_{t-1}^*(\boldsymbol{\eta}) = -\partial \boldsymbol{\varepsilon}_t'(\boldsymbol{\eta})/\partial \boldsymbol{\eta}$ satisfy the recursive equations

$$(\mathbf{I}_m - \sum_{j=1}^{q} \mathbf{\Theta}_j L^j)\mathbf{X}_{t-1}^{*'}(\boldsymbol{\eta}) = [(\mathbf{H}'\mathbf{Y}_{t-1})' \otimes \mathbf{A}, \tilde{\mathbf{U}}_{t-1}'(\boldsymbol{\eta}) \otimes \mathbf{I}_m],$$

where $\tilde{\mathbf{U}}_{t-1}(\boldsymbol{\eta}) = [(\mathbf{B}\mathbf{Y}_{t-1})', \mathbf{W}_{t-1}', \ldots, \mathbf{W}_{t-p+1}', -\boldsymbol{\varepsilon}_{t-1}'(\boldsymbol{\eta}), \ldots, -\boldsymbol{\varepsilon}_{t-q}'(\boldsymbol{\eta})]'$ and $\tilde{\mathbf{U}}_{t-1}(\boldsymbol{\eta}_0) = \mathbf{U}_{t-1}$. Since $L(\boldsymbol{\eta})$ is a nonlinear function, we need an initial estimator to obtain the RLSE.

The consistent initial estimators of $\mathbf{A}$ and $\mathbf{B}^*$ can be constructed using the full-rank estimators $\hat{\mathbf{C}}$ and $\hat{\mathbf{\Theta}}_i$, $i = 1, \ldots, q$, so that together with $\hat{\mathbf{\Phi}}_j^*$, $j = 1, \ldots, p-1$ and $\hat{\mathbf{\Theta}}_i$ they can be used to produce the consistent estimators for the parameters in model (3.13). Partition $\hat{\mathbf{C}}$ as $\hat{\mathbf{C}} = [\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2]$, where $\hat{\mathbf{C}}_1$ is an $m \times r$ matrix and $\hat{\mathbf{C}}_2$ is an $m \times d$ matrix. Then, $\hat{\mathbf{A}} \equiv \hat{\mathbf{C}}_1$ is an initial estimator of $\mathbf{A}$, and $\hat{\mathbf{B}}^* \equiv [\hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1}\hat{\mathbf{\Theta}}^{-1}\hat{\mathbf{A}}]^{-1}[\hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1}\hat{\mathbf{\Theta}}^{-1}\hat{\mathbf{C}}_2]$, with $\hat{\mathbf{\Theta}} = \mathbf{I}_m - \sum_{j=1}^q \hat{\mathbf{\Theta}}_j$, is an initial estimator of $\mathbf{B}^*$. Using $[\bar{\mathbf{B}}_\perp', \bar{\mathbf{B}}'][\mathbf{B}_\perp', \mathbf{B}_0']' = \mathbf{I}_m$ and $\mathbf{B}_0 = [\mathbf{I}_r, \mathbf{B}_0^*]$, we have

$$\hat{\mathbf{A}} - \mathbf{A}_0 = (\hat{\mathbf{C}} - \mathbf{C}_0)[\bar{\mathbf{B}}_\perp', \bar{\mathbf{B}}'][\mathbf{B}_{\perp,1}', \mathbf{I}_r]' = (\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}' + (\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}_\perp'\mathbf{B}_{\perp,1},$$

where $\mathbf{B}_{\perp,1}$ is the first r columns of $\mathbf{B}_\perp$. Let $\bar{\mathbf{B}}_\perp' = [\bar{\mathbf{B}}_{\perp,1}', \bar{\mathbf{B}}_{\perp,2}']'$ and $\bar{\mathbf{B}}' = [\bar{\mathbf{B}}_1', \bar{\mathbf{B}}_2']'$, where $\bar{\mathbf{B}}_{\perp,2}$ and $\bar{\mathbf{B}}_2$ is the last d rows of $\bar{\mathbf{B}}_\perp'$ and $\bar{\mathbf{B}}'$, respectively. Since $[\bar{\mathbf{B}}_\perp', \bar{\mathbf{B}}'][\mathbf{B}_\perp', \mathbf{B}']' = \mathbf{I}_m$, we can see that $\mathbf{B}_{\perp,1} + \bar{\mathbf{B}}_{\perp,2}^{-1}\bar{\mathbf{B}}_2 = 0$. Then, it follows that

$$\hat{\mathbf{A}} - \mathbf{A}_0 = (\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}' - (\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}_\perp'\bar{\mathbf{B}}_{\perp,2}^{-1}\bar{\mathbf{B}}_2. \tag{3.14}$$

Note that $\mathbf{B}_0\bar{\mathbf{B}}_\perp' = 0$. It follows that $\mathbf{B}_0^* = -\bar{\mathbf{B}}_{\perp,1}\bar{\mathbf{B}}_{\perp,2}^{-1}$ and $\mathbf{C}_0\bar{\mathbf{B}}_\perp' = 0$. Then,

$$\hat{\mathbf{B}}^* - \mathbf{B}_0^* = [\hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1}\hat{\mathbf{\Theta}}^{-1}\hat{\mathbf{A}}]^{-1}\hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1}\hat{\mathbf{\Theta}}^{-1}[(\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}_\perp'\bar{\mathbf{B}}_{\perp,2}^{-1}]. \tag{3.15}$$

Denote $\mathbf{M} = (\mathbf{A}_0'\mathbf{\Theta}_0'^{-1}\mathbf{\Theta}_0^{-1}\mathbf{A}_0)^{-1}\mathbf{A}_0'\mathbf{\Theta}_0'^{-1}$. By (3.14), (3.15) and Theorem 1, it is straightforward to get the following corollary.

Corollary 1. *If the conditions of Theorem 1 hold, then, as $n \to \infty$,*

(a). $n(\hat{\mathbf{B}}^* - \mathbf{B}_0^*) \xrightarrow{d} \mathbf{M}\mathbf{R}_1\mathbf{S}_{11}^{-1}\bar{\mathbf{B}}_{\perp,2}^{-1}$,

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

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

$$(c). \quad n(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} -\boldsymbol{\Gamma}_{22}^{-1} vec[\sum_{l=0}^{\infty} \boldsymbol{\gamma}_{0,l}^{'}(\mathbf{F}_0\mathbf{S}_{12l} + \mathbf{F}_{1l})] - vec\{\mathbf{F}_0\bar{\mathbf{B}}_{\perp,2}^{-1}\bar{\mathbf{B}}_2[\mathbf{I}_r, \mathbf{0}]\},$$

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

Denote $\hat{\boldsymbol{\eta}} = ([\text{vec}(\hat{\mathbf{B}}^*)]', [\text{vec}(\hat{\mathbf{A}}, \hat{\boldsymbol{\Psi}})]')'$. Then, using these consistent initial estimators, we can readily obtain $\tilde{\boldsymbol{\eta}}$ by one-step iteration

$$\tilde{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}} + \left[\sum_{t=1}^{n} \mathbf{X}_{t-1}^*(\hat{\boldsymbol{\eta}})\mathbf{X}_{t-1}^{*'}(\hat{\boldsymbol{\eta}})\right]^{-1} \left[\sum_{t=1}^{n} \mathbf{X}_{t-1}^*(\hat{\boldsymbol{\eta}})\boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\eta}})\right].$$

Let

$$\mathbf{D}_n^* = \begin{cases} n\mathbf{I}_{rd+mr+(p-1+q)m^2} & \text{if } \alpha \in (0,1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty, \\ \text{diag}(n\mathbf{I}_{rd}, n^{1/\alpha}\tilde{L}(n)\mathbf{I}_{mr+(p-1+q)m^2}) & \text{if } \alpha \in (1,2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0. \end{cases}$$

Thus, we have the following asymptotic representation

$$\begin{aligned} \mathbf{D}_n^*(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) =& [\sum_{t=1}^{n} \mathbf{X}_{t-1}^*(\hat{\boldsymbol{\eta}})\mathbf{X}_{t-1}^{*'}(\hat{\boldsymbol{\eta}})\mathbf{D}_n^{*-1}]^{-1} \\ & \times \{\sum_{t=1}^{n} \mathbf{X}_{t-1}^*(\hat{\boldsymbol{\eta}})[\mathbf{X}_{t-1}^{*'}(\hat{\boldsymbol{\eta}})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + \boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\eta}})]\}. \end{aligned} \tag{3.16}$$

The partial derivatives $\partial\boldsymbol{\varepsilon}_t^{'}(\boldsymbol{\eta})/\partial\mathbf{b}$ satisfy the recursive equations

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

where $\mathbf{H}^{'}\mathbf{Y}_{t-1} = [\mathbf{0}, \mathbf{I}_d]\mathbf{P}\mathbf{Z}_{t-1} = \bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1} + \bar{\mathbf{B}}_2\mathbf{Z}_{2,t-1}$. We note that $\mathbf{H}^{'}\mathbf{Y}_{t-1}$ and $\mathbf{Z}_{1,t}$ are purely nonstationary, $\mathbf{Z}_{2,t}$ is stationary and $\bar{\mathbf{B}}_{\perp,2}$ is nonsingular. By Lemma 3 in the appendix, we obtain

$$\mathbf{X}^*_{t-1}(\hat{\boldsymbol{\eta}}) = \begin{bmatrix} \tilde{\mathbf{Z}}^*_{1,t-1}(\hat{\boldsymbol{\eta}}) \\ \tilde{\mathbf{U}}^*_{t-1}(\hat{\boldsymbol{\eta}}) \end{bmatrix}, \tag{3.17}$$

where $\tilde{\mathbf{Z}}^*_{1,t-1}(\hat{\boldsymbol{\eta}}) = [(\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1} + \bar{\mathbf{B}}_2\mathbf{Z}_{2,t-1}) \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\Theta}}'^{-1}] + \hat{\mathbf{R}}^*_t$, with $\hat{\mathbf{R}}^*_t$ defined as in Lemma 3 in the appendix, and the matrices $\tilde{\mathbf{U}}^*_{t-1}(\hat{\boldsymbol{\eta}})$ satisfy the recursive equations

$$\hat{\boldsymbol{\Theta}}(L)\tilde{\mathbf{U}}^{*'}_{t-1}(\hat{\boldsymbol{\eta}}) = \tilde{\mathbf{U}}'_{t-1}(\hat{\boldsymbol{\eta}}) \otimes \mathbf{I}_m, \tag{3.18}$$

with $\tilde{\mathbf{U}}_{t-1}(\hat{\boldsymbol{\eta}}) = [(\hat{\mathbf{B}}\mathbf{Y}_{t-1})', \mathbf{W}'_{t-1}, \ldots, \mathbf{W}'_{t-p+1}, -\boldsymbol{\varepsilon}'_{t-1}(\hat{\boldsymbol{\eta}}), \ldots, -\boldsymbol{\varepsilon}'_{t-q}(\hat{\boldsymbol{\eta}})]'$. The limiting distribution of RLSE can be derived utilizing the similar arguments that led to Theorem 1. We now state the result of RLSE as follows.

Theorem 2. *Suppose that the condition of Theorem 1 hold. Then, as $n \to \infty$, we have*

$$(a). \quad n(\tilde{\mathbf{B}}^* - \mathbf{B}^*_0) \xrightarrow{d} \mathbf{M}\mathbf{R}_1\mathbf{S}_{11}^{-1}\bar{\mathbf{B}}_{\perp,2}^{-1},$$

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

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

$$(c). \quad n(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} \boldsymbol{\Gamma}_{22}^{-1} vec[\sum_{l=0}^{\infty} \boldsymbol{\gamma}'_{0,l}(\mathbf{F}_{2l} + \mathbf{F}_{3l} + \mathbf{F}_{4l})],$$

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

where $\mathbf{F}_{2l} = [\mathbf{I}_m - \boldsymbol{\Theta}_0^{-1}\mathbf{A}_0\mathbf{M}]\mathbf{F}_0\mathbf{R}'_1[\mathbf{M}', \mathbf{0}, \boldsymbol{\Theta}'_0\mathbf{I}_{\mathbf{m}.\mathbf{q}}]$, $\mathbf{F}_{3l} = -\boldsymbol{\Theta}_0^{-1}\mathbf{A}_0\mathbf{M}\mathbf{F}_0(\mathbf{S}_{12l} + \bar{\mathbf{B}}_{\perp,2}^{-1}\bar{\mathbf{B}}_2\mathbf{S}_{22l})$, $\mathbf{F}_{4l} = -\boldsymbol{\Theta}_0^{-1}\sum_{j'=1}^{q}\sum_{i=0}^{j'-1}\sum_{k'=0}^{\infty}\boldsymbol{\Theta}_{0,j'}\boldsymbol{\gamma}_{0,k'}\mathbf{A}_0\mathbf{M}\mathbf{F}_0\bar{\mathbf{B}}_{\perp,2}^{-1}\mathbf{H}\mathbf{S}_{22k'il}$.

When $q = 0$, we can show that Theorem 2 reduces to Theorem 3.1 in She and Ling (2020).

Corollary 1 and Theorem 2 show that $\tilde{\mathbf{B}}^*$ and $\hat{\mathbf{B}}^*$ have the same asymptotic distribution, which is similar to those as in Johansen (1995), Ahn and Reinsel (1988, 1990) and She and Ling (2020). When $\alpha \in (1,2)$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$, the FLSE and RLSE of $(\mathbf{A}, \mathbf{\Psi})$ have the same distribution. However, when $\alpha \in (0,1)$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$, their distributions are different from each other.

Using Theorems 1-2, we now consider the likelihood ratio test (LRT) for testing the rank r in model (1.2). Under the null hypothesis H_0: rank$(\mathbf{C}) = r$ against the alternative H_A: rank$(\mathbf{C}) = m$, the LRT is

$$\Lambda_n = -n \log \left(|\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_t'(\hat{\boldsymbol{\beta}})| / |\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\eta}})\boldsymbol{\varepsilon}_t'(\tilde{\boldsymbol{\eta}})| \right).$$

Using Theorem 1 and Theorem 2, it is straightforward to show that

$$\Lambda_n \xrightarrow{d} tr\{\mathbf{S}_1^{-1}\mathbf{D}\mathbf{R}_1\mathbf{S}_{11}^{-1}\mathbf{R}_1'\mathbf{D}\}, \tag{3.19}$$

where $\mathbf{D} = \mathbf{I}_m - \mathbf{\Theta}_0^{-1}\mathbf{A}_0(\mathbf{A}_0'\mathbf{\Theta}_0'^{-1}\mathbf{\Theta}_0^{-1}\mathbf{A}_0)^{-1}\mathbf{A}_0'\mathbf{\Theta}_0'^{-1}$. Based on the residuals $\{\boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\beta}})\}$, we can estimate α using the Hill's estimator as follows

$$H_n(k) = \left\{ \frac{1}{k} \sum_{t=1}^{k} \log \left(\frac{|\hat{\varepsilon}|_{(t)}}{|\hat{\varepsilon}|_{(k+1)}} \right) \right\}^{-1},$$

where $\{|\hat{\varepsilon}|_{(t)}\}$ is the decreasing order statistics of $\{|\boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\beta}})|\}$ and k is the number of chosen order statistics, see Resnick (1997).

The RLSE involves less parameters and directly estimates the cointegrating vector $\mathbf{B}_0^*$ and hence it is commonly used if the rank r is known. However, r is unknown in practice and this is a key issue. One usually needs to test the rank r by using the limiting distribution of the RLSE and FLSE. Our RLSE and FLSE are not new approaches, but Theorems 1-2 show that their limiting distributions are fully different from those when $E\|\varepsilon_t\|^2 < \infty$. Thus, the limiting distribution of the LRT in (3.19) is different from that when $E\|\varepsilon_t\|^2 < \infty$ and hence the critical values in Johansen (1995) and Yap and Reinsel (1995) cannot be used for testing the rank r under our setting.

4. Simulation Study

This section studies the finite performance of FLSE and RLSE in Sections 2 and 3. We consider the model

$$\mathbf{Y}_t = \mathbf{\Phi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t - \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} \tag{4.20}$$

where $\boldsymbol{\varepsilon}_t$ is defined by (1.4) and

$$\mathbf{\Phi}_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} = \begin{bmatrix} 0.6 & 1.0 \\ 0.12 & 0.7 \end{bmatrix} \text{ and } \mathbf{\Theta}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.4 \end{bmatrix},$$

Hence, corresponding to model (1.2), we have

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} -0.4 & 1.0 \\ 0.12 & -0.3 \end{bmatrix} = \mathbf{AB},$$

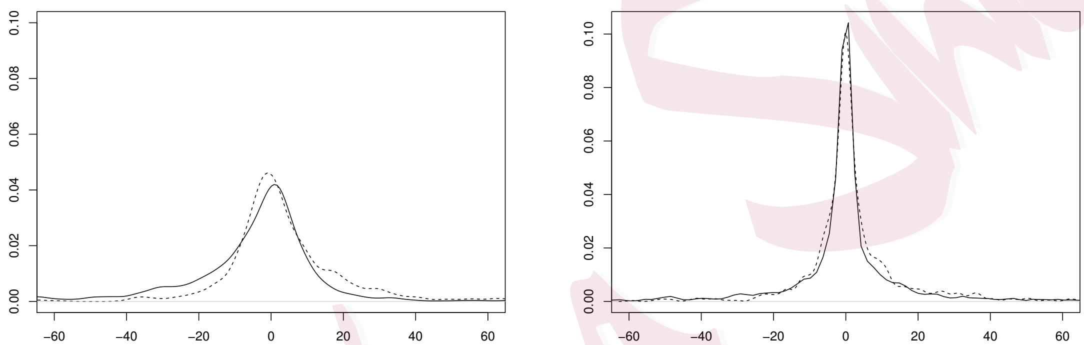
with $\mathbf{A} = (a_1, a_2)' = (-0.4, 0.12)'$ and $\mathbf{B} = (1, b^*) = (1, -2.5)$. For each of 1000 replications, samples of series length $n = 500$, 1000 and 1500 are generated. The tail index $\alpha = 0.5$ and 1.5 are considered.

We computed the estimate $\bar{\mathbf{C}}$ of $\mathbf{C}$ using the instrumental variable approach and com-

Table 1: Means, SDs and MSEs of the FLSE and RLSE

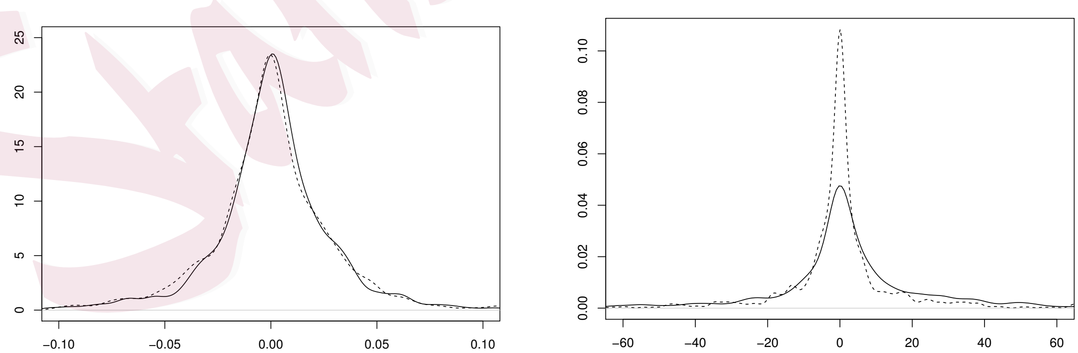
α	n			$a_1 = -0.4$	$a_2 = 0.12$	$b^* = -2.5$	$\theta_{11} = 0.3$	$\theta_{21} = 0.2$	$\theta_{12} = 0.2$	$\theta_{22} = 0.4$
$\alpha = 0.5$	500	FLSE	Mean	-0.3963	0.1193	-2.5084	0.2987	0.1886	0.1818	0.4145
			SD	0.0525	0.0511	0.0718	0.2423	0.4401	0.3127	0.3049
			MSE	0.0526	0.0511	0.0723	0.2421	0.4400	0.3131	0.3052
		RLSE	Mean	-0.3920	0.1235	-2.5042	0.3049	0.1928	0.1861	0.4186
			SD	0.0817	0.0832	0.1013	0.2501	0.4451	0.3172	0.3099
			MSE	0.0821	0.0832	0.1013	0.2502	0.4449	0.3173	0.3104
	1000	FLSE	Mean	-0.4014	0.1196	-2.5022	0.3007	0.1956	0.2012	0.3971
			SD	0.0439	0.0655	0.0873	0.1350	0.1457	0.1849	0.1519
			MSE	0.0442	0.0655	0.0875	0.1350	0.1458	0.1848	0.1519
		RLSE	Mean	-0.3967	0.1234	-2.4980	0.3042	0.1983	0.2049	0.4018
			SD	0.0447	0.0666	0.0856	0.1357	0.1464	0.1870	0.1517
			MSE	0.0448	0.0666	0.0856	0.1358	0.1464	0.1870	0.1518
	1500	FLSE	Mean	-0.3990	0.1201	-2.5018	0.3006	0.2018	0.1975	0.4009
			SD	0.0195	0.0204	0.0225	0.0556	0.1217	0.1009	0.1077
			MSE	0.0195	0.0204	0.0225	0.0556	0.1216	0.1008	0.1077
		RLSE	Mean	-0.3967	0.1224	-2.4995	0.3029	0.2042	0.1999	0.4013
			SD	0.0213	0.0219	0.0212	0.0565	0.1221	0.1008	0.1074
			MSE	0.0215	0.0220	0.0212	0.0566	0.1221	0.1007	0.1074
$\alpha = 1.5$	500	FLSE	Mean	-0.3876	0.1236	-2.5082	0.3095	0.1913	0.1687	0.3870
			SD	0.0808	0.0703	0.0405	0.0995	0.1045	0.2110	0.1789
			MSE	0.0817	0.0704	0.0413	0.0999	0.1049	0.2132	0.1793
		RLSE	Mean	-0.3792	0.1320	-2.4992	0.3179	0.2097	0.1771	0.3954
			SD	0.0831	0.0785	0.0427	0.1019	0.1102	0.2163	0.1780
			MSE	0.0856	0.0793	0.0427	0.1034	0.1106	0.2174	0.1780
	1000	FLSE	Mean	-0.3926	0.1211	-2.5062	0.3079	0.1953	0.1823	0.3931
			SD	0.0434	0.0460	0.0225	0.0618	0.0696	0.1155	0.1146
			MSE	0.0439	0.0460	0.0233	0.0622	0.0697	0.1168	0.1148
		RLSE	Mean	-0.3864	0.1267	-2.5004	0.3141	0.2015	0.1886	0.3965
			SD	0.0451	0.0487	0.0183	0.0633	0.0715	0.1166	0.1140
			MSE	0.0471	0.0491	0.0183	0.0648	0.0715	0.1171	0.1142
	1500	FLSE	Mean	-0.3949	0.1208	-2.5003	0.3041	0.1962	0.1928	0.3941
			SD	0.0428	0.0409	0.0140	0.0565	0.0604	0.1085	0.1057
			MSE	0.0430	0.0409	0.0143	0.0566	0.0604	0.1085	0.1057
		RLSE	Mean	-0.3909	0.1235	-2.4997	0.3077	0.1999	0.1945	0.3977
			SD	0.0439	0.0422	0.0118	0.0571	0.0603	0.1087	0.1054
			MSE	0.0440	0.0423	0.0118	0.0570	0.0603	0.1087	0.1056

puted the FLSE $\hat{\mathbf{C}}$ and $\hat{\mathbf{\Theta}}_1$. The initial estimates $\hat{\mathbf{A}}$ and $\hat{b}^*$ were then calculated as starting estimates for the reduced-rank estimation to obtain $\tilde{\mathbf{A}}$, $\tilde{b}^*$ and $\tilde{\mathbf{\Theta}}_1$. Table 1 summarizes the sample mean (Mean), the sample standard deviation (SD) and root mean square error (MSE) of FLSE and RLSE. It is clear that all estimators are close to true values and their SDs become smaller when n is increasing. Furthermore, the SD and root MSE of FLSE and RLSE do not have a big different. This phenomenon has been observed in She and Ling (2020). Most likely, this is because the rate of convergence of estimated parameters is very fast in the heavy-tailed case. In practice, we do not know the co-integration rank r and need both FLSE and RLSE to determine r as for LRT in (3.19).



(a) $\alpha = 1.5$ (b) $\alpha = 0.5$

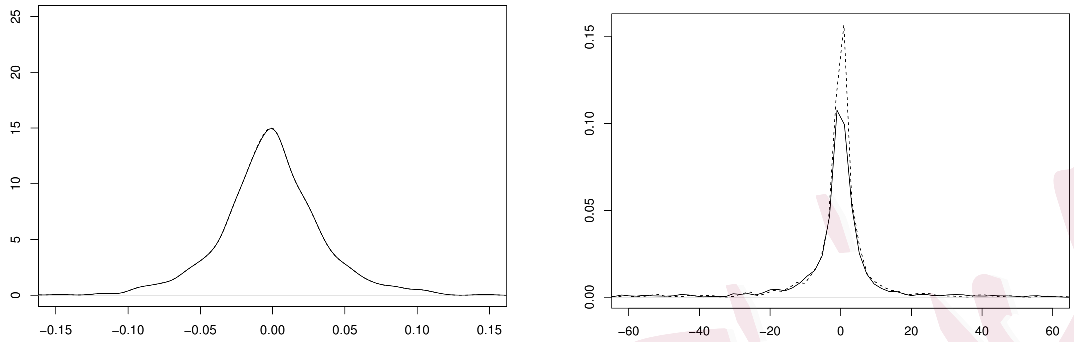
Figure 2: Empirical Densities of $n(\tilde{b}^* - b^*)$ and $n(\hat{b}^* - b^*)$



(a) $\alpha = 1.5$ and $c_n = n^{1/\alpha}\tilde{L}(n)$ (b) $\alpha = 0.5$ and $c_n = n$

Figure 3: Empirical Densities of $c_n(\tilde{a}_2 - a_2)$ and $c_n(\hat{a}_2 - a_2)$

To see the overall feature of limiting distribution, Figure 2-4 plot the empirical densities of



(a) $\alpha = 1.5$ and $c_n = n^{1/\alpha}\tilde{L}(n)$

(b) $\alpha = 0.5$ and $c_n = n$

Figure 4: Empirical Densities of $c_n(\tilde{\theta}_{11} - \theta_{11})$ and $c_n(\hat{\theta}_{11} - \theta_{11})$

FLSE (dashed) and RLSE (solid) when $n = 2500$. Figure 2-3 show that the density functions of FLSE and RLSE of (b^*, a_2) are close each other when $\alpha \in (1, 2)$. When $\alpha \in (0, 1)$, Figure 2 shows that the density functions of FLSE and RLSE of b^* are close each other. However, Figure 3 and Figure 4 clearly show that the density functions of FLSE and RLSE of (a_1, a_2) are different in this case. Figure 4 shows the density functions of of FLSE and RLSE of θ_{11} when $\alpha \in (1, 2)$ are almost identical. However, the density functions of of FLSE and RLSE of θ_{11} are clearly different when $\alpha \in (0, 1)$. More simulation results can be found in the supplementary materials. All simulation results show that the FLSE has a good performance even if we have a low-rank structure and $\alpha \in (0, 1)$.

Table 2: Critical values of Test Λ_n for Model (4.20)

H_0	$\alpha = 0.5$			$\alpha = 1.5$		
	10%	5%	1%	10%	5%	1%
$r = 1$	6.208	11.80	62.83	3.210	4.666	9.497
$r = 0$	.0303	.3442	26.49	5.804	9.665	37.48

Table 3: Sizes and Powers of Test Λ_n for Model (4.20)

			$\alpha = 0.5$			$\alpha = 1.5$		
			10%	5%	1%	10%	5%	1%
$r = 1$	Λ_n	$n = 500$	0.114	0.049	0.007	0.099	0.046	0.011
		$n = 1000$	0.105	0.057	0.010	0.092	0.046	0.010
(size)	YR	$n = 500$	0.186	0.130	0.076	0.479	0.441	0.313
		$n = 1000$	0.156	0.104	0.068	0.374	0.319	0.288
$r = 0$	Λ_n	$n = 500$	0.998	0.998	0.996	1.000	1.000	0.998
(power)		$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000

The limiting distribution of $tr\{\mathbf{S}_1^{-1}\mathbf{D}\mathbf{R}_1\mathbf{S}_{11}^{-1}\mathbf{R}_1'\mathbf{D}\}$ in (3.19) can be approximated by

$$tr\{n(\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')^{-1}\mathbf{D}\mathbf{R}_{1n}\mathbf{S}_{11n}^{-1}\mathbf{R}_{1n}'\mathbf{D}\},$$

see Lemma A.1 in She and Ling (2020), and Lemma 1 and Lemma 2 in the appendix. For model (4.20), we use sample size $n = 2500$ and 5000 replications to simulate the critical values of this limiting distribution under the hypotheses $r = 0$ and $r = 1$, respectively. The results are reported in Table 2. Using the critical values, we examine the performance of test Λ_n when the sample size $n = 500$ and $n = 1000$ via 1000 replications. The size and power of Λ_n are reported in Table 3. We can see that the sizes are all close to the nominal values under the null hypothesis $r = 1$, and the powers increase when the sample size n increases from 400 to 800 under alternative hypothesis $r = 0$. When using the critical values of Yap and Reinsel (1995) which is for model (1.2) with $E\|\varepsilon_t\|^2 < \infty$, the sizes of Λ_n are reported in the rows with YP in Table 3. It can be seen that the sizes are fully distorted in this case and hence these critical values cannot be used for testing the rank r when ε_t is a heavy-tailed noise.

5. Real Example

In this section, we consider three U.S. monthly interest rate series over the period 1970-01-01 to 2000-11-01 with 371 observations. The three series are the Federal Fund rate, 90-day Treasury Bill rate and 1-year Treasury Bill rate. Let $\mathbf{X}_t = (x_{1t}, x_{2t}, x_{3t})'$ denote the original data and $\mathbf{Y}_t = (y_{1t}, y_{2t}, y_{3t})'$ denote the log-rate, i.e. $y_{it} = \log(x_{it})$ for $i = 1, 2, 3$. $\{\mathbf{Y}_t\}$ is plotted in Figure 5. We use the first k largest data and the Hill's estimator to estimate the

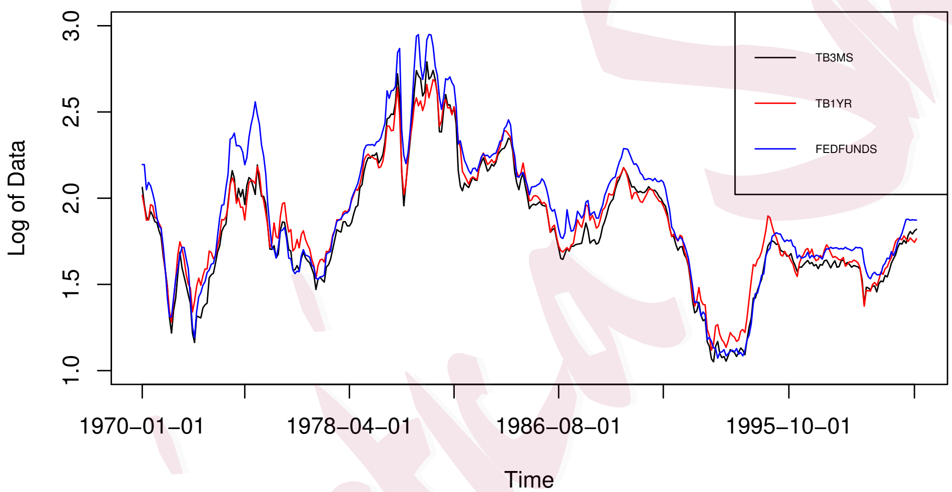


Figure 5: The Logarithms of $\{Y_t\}$.

tail index of log-returns (i.e. $r_{it} = y_{it} - y_{it-1}$) of each prices. Figure 6 is the plot of these estimated tail indices in term of k. It shows that the tail index of each log-return most likely is less than 2 but larger than 1. It seems to be reasonable to assume that these data are heavy-tailed time series. Similar to Section 7 in Yap and Reinsel (1995), we consider the following model

$$\mathbf{W}_t = \boldsymbol{\mu}^* + \mathbf{C}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1}. \tag{5.21}$$

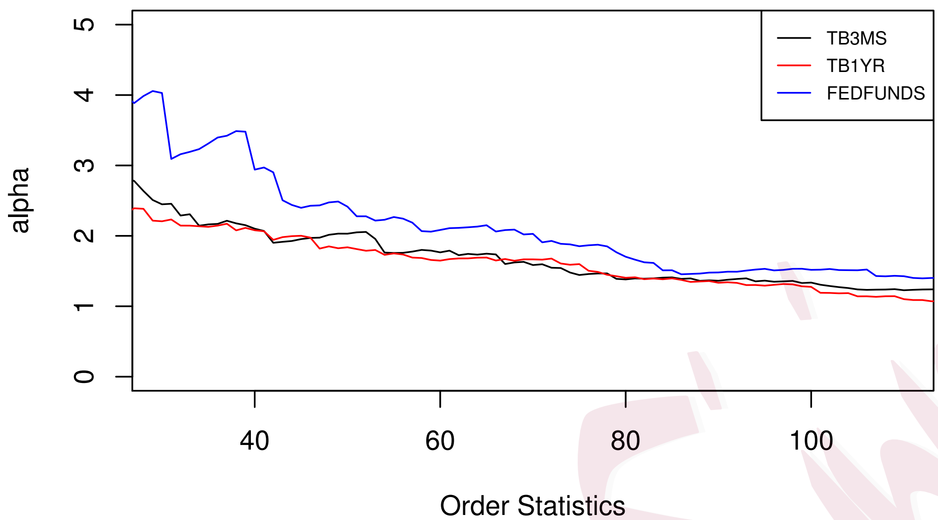


Figure 6: Hill estimator of tail index.

The full-rank LS estimates are given by

$$\hat{\mathbf{C}} = \begin{bmatrix} -0.3875 & 0.2384 & 0.1344 \\ -0.1288 & 0.0476 & 0.0538 \\ -0.0834 & 0.1965 & -0.1089 \end{bmatrix} \text{ and } \hat{\mathbf{\Theta}}_1 = \begin{bmatrix} -0.0667 & -0.2537 & -0.2583 \\ 0.0711 & -0.5475 & 0.0429 \\ -0.0752 & -0.2404 & -0.2715 \end{bmatrix},$$

and the estimated $\boldsymbol{\mu}^*$ is about $(0.0079, 0.0517, -0.0099)$.

Table 4: Tests for Cointegration Based on LRT Statistic

H_0	LRT	$\alpha = 1$			$\alpha = 1.5$		
		10%	5%	1%	10%	5%	1%
$r = 2$	4.18	6.93	10.36	27.69	6.61	8.74	16.36
$r = 1$	372.24	18.68	25.38	56.76	16.42	19.90	32.22
$r = 0$	118.97	36.08	47.52	98.33	30.41	34.99	40.71

We use the LRT in Caner (1998) to test the rank of cointegration in model (5.21). Table 4 reports the values of LRT and its critical values given in Caner (1998) at significant level 10%, 5% and 1% when $\alpha = 1$ and $\alpha = 1.5$. From this table, we can obtain that the rank

$r = 2$ cannot be rejected at significant level 5%.

A reduced-rank ARMA$(1, 1)$ model with rank$(\mathbf{C}) = 2$ imposed is fitted to the series using the procedure described in Section 3. The initial estimate of $\mathbf{B}^*$, calculated based on the full-rank LS estimation, is $\hat{\mathbf{B}}^* = [-0.9273, -0.9106]'$. The results of RLSE are as follows:

$$\tilde{\mathbf{C}} = \begin{bmatrix} -0.4313 & 0.3103 \\ -0.1578 & 0.0518 \\ -0.1435 & 0.2381 \end{bmatrix} \begin{bmatrix} 1 & 0 & -0.9270 \\ 0 & 1 & -0.9187 \end{bmatrix} = \begin{bmatrix} -0.4313 & 0.3103 & 0.1148 \\ -0.1578 & 0.0518 & 0.0987 \\ -0.1435 & 0.2381 & -0.0857 \end{bmatrix},$$

$$\text{and } \tilde{\mathbf{\Theta}}_1 = \begin{bmatrix} -0.0879 & -0.2378 & -0.2479 \\ 0.1566 & -0.5654 & -0.0261 \\ -0.0110 & -0.2662 & -0.3171 \end{bmatrix}.$$

This result shows that the 90-day Treasury Bill rate and the 1-year Treasury Bill rate do not have a cointegrating relationship, but each of them have a cointegrating relationship with the Federal Fund rate: $z_{1t} \equiv y_{1t} - 0.927 y_{2t}$ and $z_{2t} \equiv y_{1t} - 0.9187 y_{3t}$. $\{z_{1t}\}$ and $\{z_{2t}\}$ are plotted in Figure 7, from which we can see that two relationships are very stable.

6. Appendix

We need some preliminary results to prove the main theorems. We first choose a_n as follows

$$a_n = \inf\{x : P(\|\boldsymbol{\varepsilon}_1\| > x) < n^{-1}\}.$$

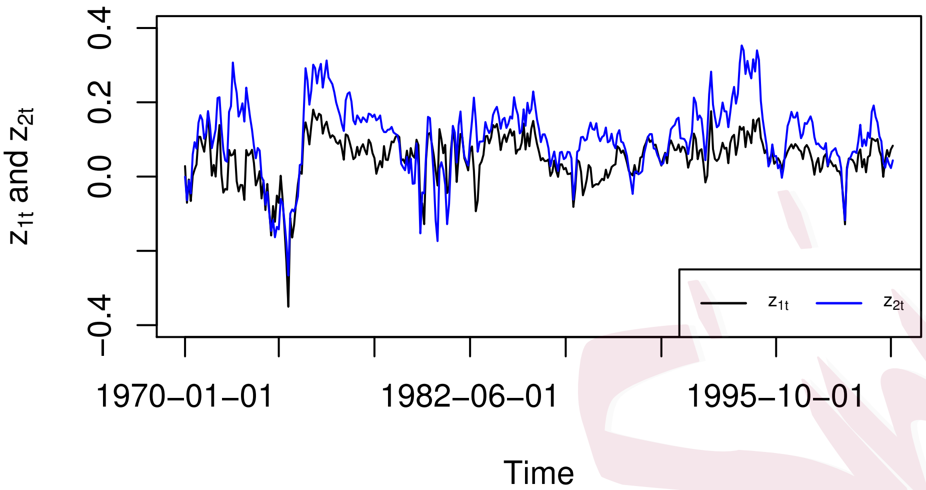


Figure 7: Time plot of z_{1t} and z_{2t}.

Then, (1.3) implies that $a_n = n^{1/\alpha}L(n)$, where $L(n)$ is a slowly variation function; see Bingham, Goldie and Teugels (1989). Similarly, we define $\tilde{a}_n$ as follows

$$\tilde{a}_n = \inf\{x : P(\|\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2^{'}\| > x) < n^{-1}\}.$$

We can show that $a_n^2/\tilde{a}_n = n^{1/\alpha}\tilde{L}(n)$, where $\tilde{L}(n)$ is also a slowly varying function, see Lemma 4 in She, Mi and Ling (2021) and Bingham, Goldie and Teugels (1989). From Proposition 3.1 in Resnick (1986), we can see that the condition (1.3) is equivalent to the following convergence

$$\sum_{t=1}^{n} \delta_{\frac{\boldsymbol{\varepsilon}_{t-1}}{a_n}} \xrightarrow{v} \sum_{i=1}^{\infty} \delta_{\mathbf{P}_i} = PRM(\mu),$$

as $n \to \infty$, where $PRM(\mu)$ is a Poisson random process with intensity measure μ and $\{\mathbf{P}_i\}$ is a sequence random vectors such that $\sum_{i=1}^{\infty} \delta_{\mathbf{P}_i}$ is the point representation of $PRM(\mu)$. From

Davis and Resnick (1986) we can see that

$$nP(\frac{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2'}{\tilde{a}_n} \in \cdot) \xrightarrow{v} \tilde{\mu}(\cdot) \text{ and } \tilde{a}_n/a_n \to \infty,$$

as $n \to \infty$ when $E\|\boldsymbol{\varepsilon}_1\|^\alpha = \infty$, where $\tilde{\mu}$ is a Radon measure on $(R^{m^2}, \mathcal{B}^{m^2})$. Let $\{\mathbf{P}_i^{(j)}\}$ be a sequence random vectors such that $\sum_{i=1}^{\infty} \delta_{\mathbf{P}_i^{(j)}}$ is the point representation of $PRM(\tilde{\mu})$ for $j = 2, 3, \ldots.$, and they are independent each other for different j. Using the same technique as that of Lemma 1 in Yap and Reinsel (1995), it is straightforward to show the following lemma.

Lemma 1. *Consider model (1.2), then we have*

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

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

Lemma 1 is fundamental to the proof of Theorem 1, and a proof can be found in the Supplementary Material. Let

$$\mathbf{P}_n(r) = a_n^{-1}\sum_{t=1}^{[nr]}\boldsymbol{\varepsilon}_t$$

with $r \in [0, 1]$. By Appendix 3.5 in Resnick (1986) and (1.3), and when $\boldsymbol{\varepsilon}_t$ has a symmetric distribution, we can show that there exists a stable process $\mathbf{P}(r)$ such that

$$\mathbf{P}_n(r) \xrightarrow{D} \mathbf{P}(r),$$

as $n \to \infty$, where $\xrightarrow{D}$ means weak convergence in the space D[0, 1] of càdlàg multivariate functions equipped with the Skorokhod J_1 topology, see Billingsley (1999). The process $\mathbf{P}(r)$ is a stable process with Itô representation

$$\mathbf{P}(r) = \sum_{t_k \le r} \mathbf{j}_k 1_{[\|\mathbf{j}_k\|>1]} + \lim_{\delta \downarrow 0} [\sum_{t_k \le r} \mathbf{j}_k 1_{[\|\mathbf{j}_k\| \in (\delta,1]]} - r \int_{\|s\| \in (\delta,1]} \mathbf{s} \boldsymbol{\nu}(d\mathbf{s})],$$

where the PRM has points $(t_k, \mathbf{j}_k) \in R_+ \times (R^0)^m$ with $R_+ = [0, \infty)$ and $R^0 = (\infty, 0) \cup (0, \infty)$, and $\boldsymbol{\nu}$ satisfies $\int (\|\mathbf{x}\|^2 \wedge 1) \boldsymbol{\nu}(d\mathbf{x}) < \infty$. The process $\mathbf{P}(r)$ does not have a closed distribution for each r, but it can approximated by $\mathbf{P}_n(r)$ in practice. By Theorem B.1 in She and Ling (2020) and Theorem 2 in She, Mi and Ling (2021), we have the following lemma.

Lemma 2. *Suppose that (1.3) and Assumptions 1 hold, $\boldsymbol{\varepsilon}_t$ has a symmetric distribution and $E\|\boldsymbol{\varepsilon}_1\|^\alpha = \infty$ with $\alpha \in (0, 2)$. Then, we have*

$$(a).\quad a_n^{-2} \sum_{t=1}^n \boldsymbol{\varepsilon}_t \mathbf{Z}_{1,t-1}' \xrightarrow{d} \mathbf{R}_1,$$

$$(b).\quad (na_n^2)^{-1} \sum_{t=1}^n \mathbf{Z}_{1,t-1} \mathbf{Z}_{1,t-1}' \xrightarrow{d} \mathbf{S}_{11},$$

$$(c).\quad \tilde{a}_n^{-1} \sum_{t=1}^n \boldsymbol{\varepsilon}_t \mathbf{U}_{t-1-l}' \xrightarrow{d} \mathbf{R}_{2l},$$

$$(d).\quad a_n^{-2} \sum_{t=1}^n \mathbf{U}_{t-1-k} \mathbf{U}_{t-1-j}' \xrightarrow{d} \mathbf{S}_{22kj},$$

$$(e).\quad a_n^{-2} \sum_{t=1}^n \mathbf{Z}_{1,t-1} \mathbf{U}_{t-1-l}' \xrightarrow{d} \mathbf{S}_{12l},$$

$$(f).\quad a_n^{-2} \sum_{t=1}^n \mathbf{Z}_{2,t-1} \mathbf{U}_{t-1-l}' \xrightarrow{d} \mathbf{S}_{22l},$$

$$(g).\quad a_n^{-2} \sum_{t=1}^n \mathbf{W}_{t-1-k'-i} \mathbf{U}_{t-1-l}' \xrightarrow{d} \mathbf{S}_{22k'il},$$

as $n \to \infty$, where $\mathbf{R}_1 = [\int_0^1 \mathbf{P}(r) d\mathbf{P}'(r)]' \boldsymbol{\phi}' [\mathbf{I}_d, \mathbf{0}]'$ with $\boldsymbol{\phi} = \sum_{i=0}^{\infty} \boldsymbol{\phi}_i$, $\mathbf{R}_{2l} = \sum_{i=0}^{\infty} \mathbf{S}_{i+2+l} \mathbf{A}_i'$,

$\mathbf{S}_{11} = [\mathbf{I}_d, \mathbf{0}]\boldsymbol{\phi}[\int_0^1 \mathbf{P}(r)\mathbf{P}'(r)dr]\boldsymbol{\phi}'[\mathbf{I}_d, \mathbf{0}]'$, $\mathbf{S}_{22kj} = \sum_{i=0}^{\infty} \mathbf{A}_i\mathbf{S}_1\mathbf{A}'_{i+k-j}$, $\mathbf{S}_{22l} = \sum_{j=0}^{\infty} \mathbf{C}_j\mathbf{S}_1\mathbf{A}'_{j+l}$, $\mathbf{S}_{22k'il} = \sum_{j=0}^{\infty} \mathbf{B}_j\mathbf{S}_1\mathbf{A}'_{j+k'+i-l}$, $\mathbf{S}_{12l} = \{\mathbf{R}'_1 \sum_{i=0}^{\infty} \mathbf{A}'_{i+l} + [\mathbf{I}_d, \mathbf{0}] \sum_{i=0}^{\infty}\sum_{j=0}^{i} \boldsymbol{\phi}_j\mathbf{S}_1\mathbf{A}'_{i+l}\}$, $\mathbf{P}(r)$ *is a stable process,* $\mathbf{S}_1 = \sum_{i=1}^{\infty} \mathbf{P}_i^{(1)}\mathbf{P}_i^{(1)'}$ *with* $\mathbf{P}_i^{(1)} = \mathbf{P}_i$ *and* $\mathbf{S}_j = \sum_{i=1}^{\infty} \mathbf{P}_i^{(j)}$ *for all* $j > 1$.

Proof of Theorem 1. Denote

$$\hat{\mathbf{R}}_{1n} = \sum_{t=1}^{n} \mathbf{Z}^*_{1,t-1}(\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_t + \mathbf{e}_{1n},\ \hat{\mathbf{R}}_{2n} = \sum_{t=1}^{n} \mathbf{U}^*_{t-1}(\bar{\boldsymbol{\beta}})\boldsymbol{\varepsilon}_t + \mathbf{e}_{2n},$$
$$\hat{\mathbf{S}}_{11n} = \sum_{t=1}^{n} \mathbf{Z}^*_{1,t-1}(\bar{\boldsymbol{\beta}})\mathbf{Z}^{*'}_{1,t-1}(\bar{\boldsymbol{\beta}}),\ \hat{\mathbf{S}}_{22n} = \sum_{t=1}^{n} \mathbf{U}^*_{t-1}(\bar{\boldsymbol{\beta}})\mathbf{U}^{*'}_{t-1}(\bar{\boldsymbol{\beta}}),$$
$$\hat{\mathbf{S}}_{12n} = \sum_{t=1}^{n} \mathbf{Z}^*_{1,t-1}(\bar{\boldsymbol{\beta}})\mathbf{U}^{*'}_{t-1}(\bar{\boldsymbol{\beta}}),$$

where $[\mathbf{e}_{1n}, \mathbf{e}_{2n}] = [\mathbf{Z}^{*'}_{1,t-1}(\bar{\boldsymbol{\beta}}), \mathbf{U}^{*'}_{t-1}(\bar{\boldsymbol{\beta}})]'[\mathbf{X}'_{t-1}(\bar{\boldsymbol{\beta}})(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\beta}}) - \boldsymbol{\varepsilon}_t]$. By (2.11), we have

$$[n(\hat{\mathbf{c}}^* - \mathbf{c}_0^*)', n^{1/\alpha}\tilde{L}(n)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)']' = \begin{bmatrix} \frac{1}{na_n^2}\hat{\mathbf{S}}_{11n} & \frac{\tilde{a}_n}{a_n^4}\hat{\mathbf{S}}_{12n} \\ \frac{1}{n\tilde{a}_n}\hat{\mathbf{S}}'_{12n} & \frac{1}{a_n^2}\hat{\mathbf{S}}_{22n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{a_n^2}\hat{\mathbf{R}}_{1n} \\ \frac{1}{\tilde{a}_n}\hat{\mathbf{R}}_{2n} \end{bmatrix}.$$

Solve the previous group equations, it follows that

$$n(\hat{\mathbf{c}}^* - \mathbf{c}_0^*) = \{\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n} - \frac{\tilde{a}_n}{a_n^4}\hat{\mathbf{S}}_{12n}(\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n})^{-1}\frac{1}{n\tilde{a}_n}\hat{\mathbf{S}}'_{12n}\}^{-1}$$
$$\times \{\frac{1}{a_n^2}\hat{\mathbf{R}}_{1n} - \frac{\tilde{a}_n}{a_n^4}\hat{\mathbf{S}}_{12n}(\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n})^{-1}\frac{1}{\tilde{a}_n}\hat{\mathbf{R}}_{2n}\}, \tag{6.22}$$
$$n^{1/\alpha}\tilde{L}(n)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \{\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n} - \frac{1}{n\tilde{a}_n}\hat{\mathbf{S}}'_{12n}(\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n})^{-1}\frac{\tilde{a}_n}{a_n^4}\hat{\mathbf{S}}_{12n}\}^{-1}$$
$$\times \{\frac{1}{\tilde{a}_n}\hat{\mathbf{R}}_{2n} - \frac{1}{n\tilde{a}_n}\hat{\mathbf{S}}'_{12n}(\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n})^{-1}\frac{1}{a_n^2}\hat{\mathbf{R}}_{1n}\}. \tag{6.23}$$

Let

$$\bar{\mathbf{\Theta}}(L)^{-1} = (\mathbf{I}_m - \sum_{j=1}^{q} \bar{\mathbf{\Theta}}_j L^j)^{-1} = \sum_{k=0}^{\infty} \bar{\gamma}_k L^k, \tag{6.24}$$

where $\bar{\boldsymbol{\gamma}}_k = O(\rho^k)$ for some $\rho \in (0,1)$. Under (2.6) and Lemma 1, we have

$$\mathbf{Z}^*_{1,t-1}(\bar{\boldsymbol{\beta}}) = (\mathbf{Z}_{1,t-1} \otimes \bar{\mathbf{\Theta}}'^{-1}) + \bar{\mathbf{R}}_t, \tag{6.25}$$

where $\bar{\mathbf{R}}_t = \sum_{j=1}^{q}\sum_{l=0}^{j-1}\sum_{k=0}^{\infty}(\mathbf{Q}_1'\mathbf{W}_{t-1-k-l} \otimes \bar{\gamma}_k'\bar{\mathbf{\Theta}}_j')\bar{\mathbf{\Theta}}'^{-1}$. By Lemma 2 (b), (d), (e), we have

$$\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n} = \frac{1}{na_n^2}\sum_{t=1}^{n}(\mathbf{Z}_{1,t-1}\mathbf{Z}'_{1,t-1}) \otimes \bar{\mathbf{\Theta}}'^{-1}\bar{\mathbf{\Theta}}^{-1} + o_p(1). \tag{6.26}$$

Note that

$$\begin{aligned}(\mathbf{I}_m - \sum_{j=1}^{q}\bar{\mathbf{\Theta}}_j L^j)[\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_t(\bar{\boldsymbol{\beta}})] =& (\bar{\mathbf{C}} - \mathbf{C}_0)\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1}(\bar{\mathbf{\Phi}}^*_j - \mathbf{\Phi}^*_{0,j})\mathbf{W}_{t-j} \\ &+ \sum_{j=1}^{q}(\mathbf{\Theta}_{0,j} - \bar{\mathbf{\Theta}}_j)\boldsymbol{\varepsilon}_{t-j}. \end{aligned}\tag{6.27}$$

By (2.9), (6.25) and (6.27), we have

$$\begin{aligned}\frac{1}{a_n^2}\hat{\mathbf{S}}_{12n} =& \frac{1}{a_n^2}\sum_{l=0}^{\infty}\sum_{t=1}^{n}(\mathbf{Z}_{1,t-1}\mathbf{U}'_{t-1-l}) \otimes \bar{\mathbf{\Theta}}'^{-1}\bar{\gamma}_l \\ &+ \frac{1}{a_n^2}\sum_{l=0}^{\infty}\sum_{t=1}^{n}(\mathbf{Z}_{1,t-1}\mathbf{Z}'_{1,t-1}\mathbf{P}_1'(\bar{\mathbf{C}} - \mathbf{C}_0)'\mathbf{J}\bar{\mathbf{\Theta}}(L)'^{-1}) \otimes \bar{\mathbf{\Theta}}'^{-1}\bar{\gamma}_l \\ &+ \frac{1}{a_n^2}\sum_{l=0}^{\infty}\sum_{t=1}^{n}\sum_{j=1}^{q}\sum_{i=0}^{j-1}\sum_{k=0}^{\infty}\mathbf{Q}_1'\mathbf{W}_{t-1-k-l}\mathbf{U}'_{t-1-l}(\bar{\boldsymbol{\beta}}) \otimes \bar{\gamma}_k'\bar{\mathbf{\Theta}}_j'\bar{\mathbf{\Theta}}'^{-1}\bar{\gamma}_l, \end{aligned}\tag{6.28}$$

where $\mathbf{J}$ denotes the $m \times [r + (p-1+q)m]$ matrix $[\mathbf{0}, \mathbf{J}_1, \ldots, \mathbf{J}_q]$ with $\mathbf{J}_i = \mathbf{I}_m$ for $i = 1, \ldots, q$.

By (6.27), (2.10), Lemma 1 and Lemma 2 (a), (b), (e), we obtain $a_n^{-2}\mathbf{e}_{1n} = o_p(1)$. Thus,

$$\frac{1}{a_n^2}\hat{\mathbf{R}}_{1n} = \frac{1}{a_n^2}\sum_{t=1}^{n}(\mathbf{Z}_{1,t-1}\otimes\bar{\mathbf{\Theta}}'^{-1})\boldsymbol{\varepsilon}_t + o_p(1). \tag{6.29}$$

By (6.27), (2.10) and Lemma 2, we can show that

$$\frac{1}{a_n^2}\sum_{t=1}^{n}\mathbf{U}_{t-1-k}(\bar{\boldsymbol{\beta}})\mathbf{U}_{t-1-j}(\bar{\boldsymbol{\beta}})' = \frac{1}{a_n^2}\sum_{t=1}^{n}\mathbf{U}_{t-1-k}\mathbf{U}'_{t-1-j} + o_p(1), \tag{6.30}$$

and its proof can be found in the Supplementary Material. Then, we have

$$\begin{aligned}\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n} =& \frac{1}{a_n^2}\sum_{t=1}^{n}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}[\mathbf{U}_{t-1-k}\mathbf{U}'_{t-1-j}]\otimes\bar{\boldsymbol{\gamma}}'_k\bar{\boldsymbol{\gamma}}_j + o_p(1)\\ \xrightarrow{d}& \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\mathbf{S}_{22kj}\otimes\boldsymbol{\gamma}'_{0,k}\boldsymbol{\gamma}_{0,j} \equiv \mathbf{\Gamma}_{22},\end{aligned} \tag{6.31}$$

as $n\to\infty$. When $1<\alpha<2$ or $\alpha=1$ and $\tilde{L}(n)\to 0$, we have $n^{\frac{1}{\alpha}-1}\tilde{L}(n)\to 0$. Thus, we have $\tilde{a}_n^{-1}\mathbf{e}_{2n} = o_p(1)$. Furthermore, we have

$$\frac{1}{\tilde{a}_n}\hat{\mathbf{R}}_{2n} = \frac{1}{\tilde{a}_n}\sum_{t=1}^{n}\sum_{k=0}^{\infty}[\mathbf{U}_{t-1-k}\otimes\mathbf{I}_m]\bar{\boldsymbol{\gamma}}'_k\boldsymbol{\varepsilon}_t + o_p(1). \tag{6.32}$$

When $0<\alpha<1$ or $\alpha=1$ and $\tilde{L}(n)\to\infty$, then $n^{1-\frac{1}{\alpha}}\tilde{L}^{-1}(n)\to 0$. Thus, we have $na_n^{-2}\mathbf{e}_{2n} = o_p(1)$. Furthermore, we have

$$\frac{n}{a_n^2}\hat{\mathbf{R}}_{2n} = \frac{n}{a_n^2}\sum_{t=1}^{n}\sum_{k=0}^{\infty}[(\mathbf{U}_{t-1-k}(\bar{\boldsymbol{\beta}})-\mathbf{U}_{t-1-k})\otimes\mathbf{I}_m]\bar{\boldsymbol{\gamma}}'_k\boldsymbol{\varepsilon}_t + o_p(1). \tag{6.33}$$

Note that $a_n^2/\tilde{a}_n = n^{\frac{1}{\alpha}}\tilde{L}(n)\to\infty$, as $n\to\infty$. By (2.10), (6.27)-(6.28), (6.31)-(6.32), Lemma

1 and Lemma 2, we can show that

$$\frac{1}{na_n^2}\hat{\mathbf{S}}_{12n}\hat{\mathbf{S}}_{22n}^{-1}\hat{\mathbf{S}}_{12n}' = o_p(1) \text{ and } \frac{\tilde{a}_n}{a_n^4}\hat{\mathbf{S}}_{12n}(\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n})^{-1}\frac{1}{\tilde{a}_n}\hat{\mathbf{R}}_{2n} = o_p(1). \tag{6.34}$$

Thus, by (6.22) and (6.34), we have

$$n(\hat{\mathbf{c}}^* - \mathbf{c}_0^*) = \{\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n}\}^{-1}\{\frac{1}{a_n^2}\hat{\mathbf{R}}_{1n}\} + o_p(1).$$

Therefore, under Lemma 1 and Lemma 2,

$$n(\hat{\mathbf{c}}^* - \mathbf{c}_0^*) = a_n^{-2}\sum_{t=1}^{n}\{[(na_n^2)^{-1}\sum_{t=1}^{n}\mathbf{Z}_{1,t-1}\mathbf{Z}_{1,t-1}']^{-1}\mathbf{Z}_{1,t-1}\otimes\bar{\mathbf{\Theta}}\}\boldsymbol{\varepsilon}_t + o_p(1),$$

or equivalently

$$\begin{aligned} n(\hat{\mathbf{C}} - \mathbf{C}_0)\bar{\mathbf{B}}_{\perp}' = & \ (a_n^{-2}\sum_{t=1}^{n}\bar{\mathbf{\Theta}}\boldsymbol{\varepsilon}_t\mathbf{Z}_{1,t-1}')[(na_n^2)^{-1}\sum_{t=1}^{n}\mathbf{Z}_{1,t-1}\mathbf{Z}_{1,t-1}']^{-1} + o_p(1) \\ & \xrightarrow{d} \mathbf{\Theta}_0\mathbf{R}_1\mathbf{S}_{11}^{-1}, \end{aligned}$$

as $n \to \infty$. That is, (a) holds. When $1 < \alpha < 2$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$, we have $n^{\frac{1}{\alpha}-1}\tilde{L}(n) \to 0$. By (2.12), (6.26)-(6.29), Lemma 1 and Lemma 2 (a), (b), (d), (e), (g), we have

$$\frac{1}{a_n^2}\hat{\mathbf{S}}_{12n}'\hat{\mathbf{S}}_{11n}^{-1}\hat{\mathbf{S}}_{12n} = o_p(1) \text{ and } \frac{1}{n\tilde{a}_n}\hat{\mathbf{S}}_{12n}'(\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n})^{-1}\frac{1}{a_n^2}\hat{\mathbf{R}}_{1n} = o_p(1),$$

which follows that

$$n^{1/\alpha}\tilde{L}(n)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \{\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n}\}^{-1}\frac{1}{\tilde{a}_n}\hat{\mathbf{R}}_{2n} + o_p(1).$$

Therefore, by (6.31)-(6.32) and Lemma 2

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

as $n\to\infty$. That is, (b) holds. When $0<\alpha<1$ or $\alpha=1$ and $\tilde{L}(n)\to\infty$, by (6.23), (6.26)-(6.31) and (6.33), we have

$$\begin{aligned}n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=&\{\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n}\}^{-1}\frac{n}{a_n^2}\hat{\mathbf{R}}_{2n}\\&-\{\frac{1}{a_n^2}\hat{\mathbf{S}}_{22n}\}^{-1}\{-\frac{1}{a_n^2}\hat{\mathbf{S}}_{12n}^{'}(\frac{1}{na_n^2}\hat{\mathbf{S}}_{11n})^{-1}\frac{1}{a_n^2}\hat{\mathbf{R}}_{1n}\}+o_p(1)\\ \xrightarrow{d}&-\boldsymbol{\Gamma}_{22}^{-1}\text{vec}[\sum_{l=0}^{\infty}\boldsymbol{\gamma}_{0,l}^{'}(\mathbf{F}_0\mathbf{S}_{12l}+\mathbf{F}_{1l})],\end{aligned}$$

as $n\to\infty$. That is, (c) holds, which completes the proof. □

Lemma 3. *Consider the model (3.13), then*

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

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

Lemma 3 is used to prove Theorem 2. Its proof is given in Supplemental material of this paper. Under Lemma 2 (c), (d) and Lemma 3, we obtain that $\tilde{a}_n^{-1}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\mathbf{R}_t^{*'}$ is of order $O_p(1)$, and $a_n^{-2}\sum_{t=1}^{n}\mathbf{R}_t^*\mathbf{R}_t^{*'}$ is of order $O_p(1)$.

Proof of Theorem 2: Denote

$$\tilde{\mathbf{R}}_{1n}=\sum_{t=1}^{n}\tilde{\mathbf{Z}}_{1,t-1}^{*}(\hat{\boldsymbol{\eta}})\boldsymbol{\varepsilon}_t+\tilde{\mathbf{e}}_{1n},\ \tilde{\mathbf{R}}_{2n}=\sum_{t=1}^{n}\tilde{\mathbf{U}}_{t-1}^{*}(\hat{\boldsymbol{\eta}})\boldsymbol{\varepsilon}_t+\tilde{\mathbf{e}}_{2n},$$

$$\tilde{\mathbf{S}}_{11n} = \sum_{t=1}^{n} \tilde{\mathbf{Z}}^*_{1,t-1}(\hat{\boldsymbol{\eta}})\tilde{\mathbf{Z}}^{*'}_{1,t-1}(\hat{\boldsymbol{\eta}}),\ \tilde{\mathbf{S}}_{22n} = \sum_{t=1}^{n} \tilde{\mathbf{U}}^*_{t-1}(\hat{\boldsymbol{\eta}})\tilde{\mathbf{U}}^{*'}_{t-1}(\hat{\boldsymbol{\eta}}),$$
$$\tilde{\mathbf{S}}_{12n} = \sum_{t=1}^{n} \tilde{\mathbf{Z}}^*_{1,t-1}(\hat{\boldsymbol{\eta}})\tilde{\mathbf{U}}^{*'}_{t-1}(\hat{\boldsymbol{\eta}}),$$

where $[\tilde{\mathbf{e}}_{1n}, \tilde{\mathbf{e}}_{2n}] = [\tilde{\mathbf{Z}}^*_{1,t-1}(\hat{\boldsymbol{\eta}}), \tilde{\mathbf{U}}^*_{t-1}(\hat{\boldsymbol{\eta}})][\mathbf{X}^{*'}_{t-1}(\hat{\boldsymbol{\eta}})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + \boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\eta}}) - \boldsymbol{\varepsilon}_t]$. Then, by (3.16), we have

$$[n(\tilde{\mathbf{b}} - \mathbf{b}_0)', n^{1/\alpha}\tilde{L}(n)(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)']' = \begin{bmatrix} \frac{1}{na_n^2}\tilde{\mathbf{S}}_{11n} & \frac{\tilde{a}_n}{a_n^4}\tilde{\mathbf{S}}_{12n} \\ \frac{1}{n\tilde{a}_n}\tilde{\mathbf{S}}'_{12n} & \frac{1}{a_n^2}\tilde{\mathbf{S}}_{22n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{a_n^2}\tilde{\mathbf{R}}_{1n} \\ \frac{1}{\tilde{a}_n}\tilde{\mathbf{R}}_{2n} \end{bmatrix}.$$

Solve the previous group equaiton, it follows that

$$\begin{aligned} n\{\mathrm{vec}(\tilde{\mathbf{B}}^* - \mathbf{B}^*_0)\} &= \{\frac{1}{na_n^2}\tilde{\mathbf{S}}_{11n} - \frac{\tilde{a}_n}{a_n^4}\tilde{\mathbf{S}}_{12n}(\frac{1}{a_n^2}\tilde{\mathbf{S}}_{22n})^{-1}\frac{1}{n\tilde{a}_n}\tilde{\mathbf{S}}'_{12n}\}^{-1} \\ &\quad \times \{\frac{1}{a_n^2}\tilde{\mathbf{R}}_{1n} - \frac{\tilde{a}_n}{a_n^4}\tilde{\mathbf{S}}_{12n}(\frac{1}{a_n^2}\tilde{\mathbf{S}}_{22n})^{-1}\frac{1}{\tilde{a}_n}\tilde{\mathbf{R}}_{2n}\}, \qquad (6.35) \\ n^{1/\alpha}\tilde{L}(n)(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) &= \{\frac{1}{a_n^2}\tilde{\mathbf{S}}_{22n} - \frac{1}{n\tilde{a}_n}\tilde{\mathbf{S}}'_{12n}(\frac{1}{na_n^2}\tilde{\mathbf{S}}_{11n})^{-1}\frac{\tilde{a}_n}{a_n^4}\tilde{\mathbf{S}}_{12n}\}^{-1} \\ &\quad \times \{\frac{1}{\tilde{a}_n}\tilde{\mathbf{R}}_{2n} - \frac{1}{n\tilde{a}_n}\tilde{\mathbf{S}}'_{12n}(\frac{1}{na_n^2}\tilde{\mathbf{S}}_{11n})^{-1}\frac{1}{a_n^2}\tilde{\mathbf{R}}_{1n}\}. \qquad (6.36) \end{aligned}$$

Let

$$\hat{\boldsymbol{\Theta}}(L)^{-1} = (\mathbf{I}_m - \sum_{j=1}^{q}\hat{\boldsymbol{\Theta}}_j L^j)^{-1} = \sum_{k=0}^{\infty}\hat{\boldsymbol{\gamma}}_k L^k, \qquad (6.37)$$

where $\hat{\boldsymbol{\gamma}}_k = O(\rho^k)$ for some $\rho \in (0, 1)$. By Lemma 3 and (3.17), we have

$$\tilde{\mathbf{Z}}^*_{1,t-1}(\hat{\boldsymbol{\eta}}) = [(\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1} + \bar{\mathbf{B}}_2\mathbf{Z}_{2,t-1}) \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\Theta}}'^{-1}] + \hat{\mathbf{R}}^*_t, \qquad (6.38)$$

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

$$\frac{1}{na_n^2}\tilde{\mathbf{S}}_{11n} = \frac{1}{na_n^2}\sum_{t=1}^{n}\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1}\mathbf{Z}_{1,t-1}'\bar{\mathbf{B}}_{\perp,2}' \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\Theta}}'^{-1}\hat{\boldsymbol{\Theta}}^{-1}\hat{\mathbf{A}} + o_p(1). \tag{6.39}$$

By (3.18) and (6.38), we have

$$\begin{aligned}\frac{1}{a_n^2}\tilde{\mathbf{S}}_{12n} =& \frac{1}{a_n^2}\sum_{t=1}^{n}\sum_{l=0}^{\infty}(\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1}\mathbf{U}_{t-1-l}') \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\Theta}}'^{-1}\hat{\boldsymbol{\gamma}}_l \qquad (6.40)\\ &+ \frac{1}{a_n^2}\sum_{t=1}^{n}\sum_{l=0}^{\infty}(\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1}[\tilde{\mathbf{U}}_{t-1-l}(\hat{\boldsymbol{\eta}}) - \mathbf{U}_{t-1-l}]') \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\Theta}}'^{-1}\hat{\boldsymbol{\gamma}}_l\\ &+ \frac{1}{a_n^2}\sum_{t=1}^{n}\sum_{l=0}^{\infty}(\bar{\mathbf{B}}_2\mathbf{Z}_{2,t-1}\mathbf{U}_{t-1-l}') \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\Theta}}'^{-1}\hat{\boldsymbol{\gamma}}_l\\ &+ \frac{1}{a_n^2}\sum_{t=1}^{n}\sum_{l=0}^{\infty}\sum_{j=1}^{q}\sum_{i=1}^{j-1}\sum_{k=0}^{\infty}(\mathbf{H}'\mathbf{W}_{t-1-k-i}\mathbf{U}_{t-1-l}') \otimes \hat{\mathbf{A}}'\hat{\boldsymbol{\gamma}}_k'\hat{\boldsymbol{\Theta}}_j'\hat{\boldsymbol{\Theta}}'^{-1}\hat{\boldsymbol{\gamma}}_l + o_p(1).\end{aligned}$$

Note that

$$\begin{aligned}(\mathbf{I}_m - \sum_{j=1}^{q}\hat{\boldsymbol{\Theta}}_jL^j)[\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\eta}})] =& (\hat{\mathbf{A}}\hat{\mathbf{B}} - \mathbf{A}_0\mathbf{B}_0)\mathbf{Y}_{t-1} + \sum_{j=1}^{p-1}(\hat{\boldsymbol{\Phi}}_j^* - \boldsymbol{\Phi}_{0,j}^*)\mathbf{W}_{t-j}\\ &+ \sum_{j=1}^{q}(\boldsymbol{\Theta}_{0,j} - \hat{\boldsymbol{\Theta}}_j)\boldsymbol{\varepsilon}_{t-j}. \qquad (6.41)\end{aligned}$$

By (6.41), Corollary 1 and Lemma 2, we can show that

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

Thus,

$$\frac{1}{a_n^2}\tilde{\mathbf{S}}_{22n} = \frac{1}{a_n^2}\sum_{t=1}^{n}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}[\mathbf{U}_{t-1-k}\mathbf{U}_{t-1-j}'] \otimes \hat{\boldsymbol{\gamma}}_k'\hat{\boldsymbol{\gamma}}_j + o_p(1). \tag{6.42}$$

By Corollary 3.1, (6.38), (6.41) and Lemma 2, we obtain $a_n^{-2}\tilde{\mathbf{e}}_{1n} = o_p(1)$. Thus,

$$\frac{1}{a_n^2}\tilde{\mathbf{R}}_{1n} = \frac{1}{a_n^2}\sum_{t=1}^{n}(\bar{\mathbf{B}}_{\perp,2}\mathbf{Z}_{1,t-1} \otimes \hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1})\boldsymbol{\varepsilon}_t + o_p(1). \tag{6.43}$$

When $1 < \alpha < 2$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$, we have $n^{\frac{1}{\alpha}-1}\tilde{L}(n) \to 0$. Thus, we have $\tilde{a}_n^{-1}\tilde{\mathbf{e}}_{2n} = o_p(1)$. Furthermore, we have

$$\frac{1}{\tilde{a}_n}\tilde{\mathbf{R}}_{2n} = \frac{1}{\tilde{a}_n}\sum_{t=1}^{n}\sum_{k=0}^{\infty}[\mathbf{U}_{t-1-k} \otimes \mathbf{I}_m]\hat{\boldsymbol{\gamma}}_k'\boldsymbol{\varepsilon}_t + o_p(1). \tag{6.44}$$

When $0 < \alpha < 1$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$, then $n^{1-\frac{1}{\alpha}}\tilde{L}^{-1}(n) \to 0$. Thus, we have $na_n^{-2}\tilde{\mathbf{e}}_{2n} = o_p(1)$. Furthermore, we have

$$\frac{n}{a_n^2}\tilde{\mathbf{R}}_{2n} = \frac{n}{a_n^2}\sum_{t=1}^{n}\sum_{k=0}^{\infty}[(\tilde{\mathbf{U}}_{t-1-k}(\hat{\boldsymbol{\eta}}) - \mathbf{U}_{t-1-k}) \otimes \mathbf{I}_m]\hat{\boldsymbol{\gamma}}_k'\boldsymbol{\varepsilon}_t + o_p(1). \tag{6.45}$$

By (6.40)-(6.42), (6.44)-(6.45) and Lemma 2, we have

$$\frac{\tilde{a}_n}{a_n^4}\tilde{\mathbf{S}}_{12n}(\frac{1}{a_n^2}\tilde{\mathbf{S}}_{22n})^{-1}\frac{1}{\tilde{a}_n}\tilde{\mathbf{R}}_{2n} = o_p(1) \quad \text{and} \quad \frac{1}{na_n^2}\tilde{\mathbf{S}}_{12n}\tilde{\mathbf{S}}_{22n}^{-1}\tilde{\mathbf{S}}_{12n}' = o_p(1). \tag{6.46}$$

Denote $\hat{\mathbf{M}} = (\hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1}\hat{\mathbf{\Theta}}^{-1}\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}'\hat{\mathbf{\Theta}}'^{-1}$. Then, by (6.35), (6.39), (6.43) and Lemma 2,

$$\begin{aligned} n(\tilde{\mathbf{B}}^* - \mathbf{B}_0^*) &= \hat{\mathbf{M}}(a_n^{-2}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\mathbf{Z}_{1,t-1}')[(na_n^2)^{-1}\sum_{t=1}^{n}\mathbf{Z}_{1,t-1}\mathbf{Z}_{1,t-1}']^{-1}\bar{\mathbf{B}}_{\perp,2}^{-1} + o_p(1) \\ &\xrightarrow{d} \mathbf{M}\mathbf{R}_1\mathbf{S}_{11}^{-1}\bar{\mathbf{B}}_{\perp,2}^{-1}, \end{aligned}$$

as $n \to \infty$. That is, (a) holds. When $1 < \alpha < 2$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$, we have

$n^{\frac{1}{\alpha}-1}\tilde{L}(n) \to 0$. By Lemma 2, (6.40), (6.39) and (6.43), we have

$$\frac{1}{n\tilde{a}_n}\tilde{\mathbf{S}}_{12n}'(\frac{1}{na_n^2}\tilde{\mathbf{S}}_{11n})^{-1}\frac{1}{a_n^2}\tilde{\mathbf{R}}_{1n} = o_p(1).$$

It follows that

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

as $n \to \infty$. That is, (b) holds. When $0 < \alpha < 1$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$, by (6.36), (6.39) (6.40)-(6.43) and (6.45), it follows that

$$n(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} \boldsymbol{\Gamma}_{22}^{-1}\text{vec}[\sum_{l=0}^{\infty}\boldsymbol{\gamma}_{0,l}'(\mathbf{F}_{2l} + \mathbf{F}_{3l} + \mathbf{F}_{4l})],$$

as $n \to \infty$. That is, (c) holds. This completes the proof. □

Supplementary Material

In the Supplementary Material, we provide a list of notation, additional simulation results, the results of the model with a constant term included, and proofs of Lemma 1, Lemma 3 and result (6.30).

Acknowledgments

Authors would like to thank Co-Editor Su-Yun Huang and two referees for their very helpful comments. Ling's research was supported by Hong Kong Research Grants Commission Grants (16303118, 16301620, 16300621, 16500522 and SRFS2223-6S02) and Australian Research Council.

References

Ahn, S. K. and Reinsel, G. C. (1988), Nested Reduced-Rank Autogressive Models for Multiple Time Series. *Journal of the American Statistical Association*, 83, 849–856.

Ahn, S. K. and Reinsel, G. C. (1990), Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85, 813–823.

Billingsley, P. (1999), *Convergence of Probability Measures, Second ed*. John Wiley & Sons, New York.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989), *Regular Variation*, Cambridge Univ. Press, Cambridge.

Cai, B.Q., Gao, J.T. and Tjøstheim, D. (2017), A New Class of Bivariate Threshold Cointegration Models. *Journal of Business and Economic Statistics* , 35, 288–305.

Caner, M. (1998), Tests for Cointegration With Infinite Variance Errors. *Journal of Econometrics*, 86, 155–175.

Davis, R. A., Knight, K. and Liu, J. (1992), M-estimation for Autoregressions with Infinite Variance. *Stochastic Processes and their Applications*, 40, 145–180.

Davis, R. A., Mikosch, T. and Pfaffel, O. (2016), Asymptotic Theory for the Sample Covariance Matrix of a Heavy-tailed Multivariate Time Series. *Stochastic Processes and their Applications*, 126, 767–799.

Davis, R. A. and Resnick, S. I. (1985), Limit Theory for Moving Averages of Random Variables With Regularly Varying Tail Probabilities. *The Annals of Probability*, 13, 179–195.

Davis, R. A. and Resnick, S. I. (1986), Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages. *The Annals of Statistics*, 14, 533–558.

Engle, R. F. and Granger, C. W. J. (1987), Cointegration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55, 251–276.

Granger, C. W. J. (1983), Co-integrated variables and error-correcting models. *UCSD discussion paper, Univ. California, San Diego*.

Guo, F., Ling, S. and Mi, Z. (2021), Automated estimation of heavy-tailed vector error correction models. *Statistica Sinica*, 32, 2171–2198.

Hall, A. (1989), Testing for a Unit Root in the Presence of Moving Average Errors. *Biometrika*, 76, 49–56.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press: Oxford.

Mikosch, T., Gadrich, T., Klüppelberg, C. and Adler, R. J. (1995), Parameter Estimation for ARMA Models With Infinite Variance Innovations. *The Annals of Statistics*, 23, 305–326.

Pantula, S. G. and Hall, A. (1991), Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach. *Journal of Econometrics*, 48, 325–353.

Phillips, P. C. B. and Durlauf, S. N. (1986), Multiple Time Series Regression With Integrated Processes. *Review of Economic Studies*, 53, 473–495.

Resnick, S. I. (1986), Point Processes, Regular Variation and Weak Convergence, *Advances in Applied Probability*, 18, 66–138.

Resnick, S. I. (1997), Heavy Tail Modeling and Teletraffic Data. *The Annals of Statistics*, 25, 1805–1869.

She, R. and Ling, S. (2020), Asymptotic Inference of Heavy-Tailed Vector Error Correction Models. *Journal of Econometrics*, 214, 433–450.

She, R., Mi, Z. and Ling, S. (2021), Whittle Parameter Estimation for Vector ARMA Models with Heavy-tailed Noises. *Journal of Statistical Planning and Inference*, 219, 216–230.

Tsay, R. S. (2014), *Multivariate Time Series Analysis with R and Financial Applications*. Wiley, New York: NY.

Tsay, R. S. and Tiao, G. C. (1990), Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *The Annals of Statistics*, 18, 220-250.

Wang, Q. (2014), Martingale Limit Theorem Revisited and Nonlinear Cointegrating Regression. *Econometric Theory*, 30, 509–535.

Yap, S. F. and Reinsel, G. C. (1995), Estimation and Testing for Unit Roots in a Partially Nonstationary Vector

Autoregressive Moving Average Model. *Journal of the American Statistical Association*, 90, 253–267.

Zhang, R., Robinson, P. and Yao Q. (2019), Identifying Cointegration by Eigenanalysis. *Journal of the American Statistical Association*, 114, 916–927.

School of Mathematics and Statistics

Beijing Institute of Technology, Beijing, China

E-mail: ffguo@bit.edu.cn

Department of Mathematics

Hong Kong University of Science and Technology, Hong Kong

E-mail: maling@ust.hk