

AUTOMATED ESTIMATION OF HEAVY-TAILED VECTOR ERROR CORRECTION MODELS

Feifei Guo, Shiqing Ling and Zichuan Mi

*Beijing Institute of Technology, Hong Kong University
of Science and Technology and Shanxi University of Finance and Economics*

Abstract: This paper proposes an automated approach that uses adaptive shrinkage techniques to determine the cointegrating rank and estimate the parameters simultaneously in a vector error correction model with unknown order p when its noise is represented by independent and identically distributed heavy-tailed random vectors with tail index $\alpha \in (0, 2)$. We show that the estimated cointegrating rank and order p are equal to the true rank and the true order p_0 , respectively, with probability trending to one as the sample size $n \rightarrow \infty$. The other estimated parameters achieve the oracle property. That is, they have the same rate of convergence and the same limiting distribution as those of the estimated parameters when the cointegrating rank and the true order p_0 are known. This paper also proposes a data-driven procedure for selecting the tuning parameters. Simulation studies are carried out to evaluate the performance of the proposed procedure for finite samples. Lastly, we use our techniques to explore the long-run and short-run behavior of the prices of wheat, corn, and wheat flour in the United States.

Key words and phrases: Cointegration, heavy-tailed, LASSO, reduced rank LSE.

1. Introduction

The vector error correction (VEC) model was introduced by Granger (1983) and Engle and Granger (1987). Estimating and testing cointegration is the most essential target for the VEC model, and various approaches have been proposed in the literature. Early research can be found in Phillips and Durlauf (1986), Ahn and Reinsel (1990), Reinsel and Ahn (1992), Stock and Watson (1993), and Johansen (1988, 1995), among many others. Recently, Wang and Phillips (2012) proposed a test for nonlinear nonstationary models. Kristensen and Rahbek (2013) develop tests and estimators for nonlinear cointegrating VEC models. Cavaliere, Nielsen and Rahbek (2015) consider a bootstrap test on the cointegration rank relation in vector autoregressive (AR) models. To determine the cointegrating relationship of vector time series, the classical method needs to im-

Corresponding author: Shiqing Ling, Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong. E-mail: maling@ust.hk.

plement a pre-testing procedure. Liao and Phillips (2015) proposed a method to estimate the cointegration vector and its rank simultaneously using the shrinkage technique–group Lasso approach. This approach does not need to estimate the long-variance of linear processes using a nonparametric approach, as in Phillips and Solo (1992), or predetermine the order of the VEC models, as in Johansen (1995).

The research on cointegration systems focuses mainly on time series with a finite second or even higher moment. Heavy-tailed time series do not have a finite second moment, and are often observed in financial markets, engineering, network systems and other areas; see Resnick (1997). Davis and Resnick (1985, 1986) show that the limiting distribution of the least square estimator (LSE) of the parameters in a heavy-tailed AR process is a function of two stable random variables, with a rate of convergence much faster than \sqrt{n} . Mikosch et al. (1995) studied the whittle estimators for the heavy-tailed ARMA model and give its asymptotic properties. Zhang and Ling (2015) established the asymptotic properties of the AR model with heavy-tailed G-GARCH noise. Caner (1998) developed the asymptotic theory for residual-based tests and quasi-likelihood ratio tests for cointegration under the assumption of infinite variance errors. She and Ling (2020) studied the heavy-tailed VEC model and established the asymptotic theory of the full-rank LSE (FLSE) and reduced-rank LSE (RLSE). However, their theory cannot be applied to test the cointegrating rank of a heavy-tailed VEC model. Thus, except for a very special case in Caner (1998), it remains an open problem to determine the cointegrating rank in a VEC model when its noise is a heavy-tailed random vector.

This study develops an automated approach that uses adaptive shrinkage techniques to determine the cointegrating rank and estimate the parameters simultaneously in a VEC model with unknown order p and noise represented by independent and identically distributed (i.i.d.) heavy-tailed random vectors with tail index $\alpha \in (0, 2)$. We show that the estimated cointegrating rank and order p are equal to the true rank and the true order p_o , respectively, with probability tending to one as the sample size $n \rightarrow \infty$. The other estimated parameters achieve the oracle property. That is, they have the same rate of convergence and the same limiting distribution as those of the estimated parameters when the cointegrating rank and the true order p_o are known. We also propose a data-driven procedure for selecting the tuning parameters.

The Lasso approach was developed by Tibshirani (1996) for selecting variables and estimating parameters. It has been studied extensively and many variants have been proposed; see, for example, Fan and Li (2001) for a non-concave

penalized likelihood, Fan and Li (2002) for Cox's proportional hazards model, Knight and Fu (2000) and Wang, Li and Tsai (2007) for Lasso-type estimators of regression models, Yuan and Lin (2006) for model selection with grouped variables, Zou (2006) for the adaptive Lasso, and Huang, Ma and Zhang (2008) for the adaptive Lasso in high-dimensional regression. Chen and Chan (2011) considered the adaptive Lasso for ARMA model selection, and obtained asymptotic normality for the estimated parameters. Song and Bickel (2011) studied the Lasso estimator for a large vector AR model. Kock (2016) investigated the adaptive Lasso for AR models. Chan, Ling and Yau (2020) studied Lasso-based variable selection for stationary and unit-root ARMA models. The results presented here may provide a new insight into the Lasso approach for both stationary and nonstationary heavy-tailed time series.

The remainder of the paper is organized as follows. Section 2 proposes the shrinkage LSE for VEC models and gives its consistency. Section 3 gives the oracle property of the shrinkage LSE. Section 4 presents the selection of the adaptive tuning parameters. Simulation results are reported in Section 5. Section 6 applies our method to an empirical example. All proofs of the main results are provided in the appendix and Supplementary Material.

2. Model and LS Shrinkage Estimation

We consider the following VEC representation of a cointegrated system:

$$\Delta \mathbf{Y}_t = \mathbf{\Pi}_o \mathbf{Y}_{t-1} + \sum_{j=1}^p \mathbf{B}_{o,j} \Delta \mathbf{Y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad (2.1)$$

where $\Delta \mathbf{Y}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$, \mathbf{Y}_t is an m -dimensional vector-valued time series, $\mathbf{\Pi}_o = \boldsymbol{\alpha}_o \boldsymbol{\beta}'_o$, with $\boldsymbol{\alpha}_o$ and $\boldsymbol{\beta}_o$ being $m \times r_o$ full-rank matrices, $\mathbf{B}_{o,j}$ ($j = 1, \dots, p$) are $m \times m$ coefficient matrices, where $p > \text{true order } p_o$ and $\mathbf{B}_{o,j} = 0$ if $j > p_o$, and $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of i.i.d. m -dimensional random vectors. Model (2.1) is a partially nonstationary vector AR($p+1$) model of $\{Y_t\}$; see, for example, Ahn and Reinsel (1990) and Johansen (1988, 1995). Here, $\{Y_t\}$ is not stationary, but $\{\boldsymbol{\beta}'_o Y_t\}$ is a stationary time series. Thus, $\boldsymbol{\beta}_o$ is called the cointegrating vector or long-run cointegrating relations of Y_t . The rank r_o of $\mathbf{\Pi}_o$ is called the cointegrating rank of Y_t , and it measures the number of cointegrating relations in the system. The set of nonzero matrices $\mathbf{B}_{o,j}$ ($j = 1, \dots, p$) is characterizes the transient dynamics in the systems.

We assume that $\boldsymbol{\varepsilon}_t$ satisfies the following condition:

$$nP\left(\frac{\varepsilon_1}{a_n} \in \cdot\right) \xrightarrow{v} \mu(\cdot), \quad (2.2)$$

as $n \rightarrow \infty$, where μ is a Radon measure on (R^m, \mathcal{B}^m) , a_n is an increasing sequence diverging to ∞ , and \xrightarrow{v} means the vague convergence in Durrett (2019, p.121). Here, (2.2) is called the regular variation function, and is equivalent to there existing a probability measure μ^* on the unit sphere \mathbb{S}^m in R^m , such that, for any $x > 0$,

$$\frac{P(\|\varepsilon_1\| > tx, \varepsilon_1/\|\varepsilon_1\| \in \cdot)}{P(\|\varepsilon_1\| > t)} \xrightarrow{v} x^{-\alpha} \mu^*(\cdot),$$

as $t \rightarrow \infty$, where $\alpha > 0$ is called the tail index and $\|\cdot\|$ denotes the Euclidean norm; see Resnick (1986). When $\alpha \in (0, 2)$, ε_1 does not have a finite covariance matrix and is called a heavy-tailed random vector. One class of heavy-tailed random vectors is ε_1 with its characteristic function as follows:

$$\phi(u) = E \exp^{iu' \varepsilon_1} = \exp^{-\int_{s \in \mathbb{S}^m} \{|u's|^\alpha + iv(u's, \alpha)\} \Lambda(ds) + iu'\delta}, \quad \forall u \in R^m,$$

where Λ is a finite measure on \mathbb{S}^m , δ is a shift vector in R^m , and for any $y \in R$,

$$v(y, \alpha) = \begin{cases} -\text{sign}(y) * \tan\left(\frac{\pi\alpha}{2}\right) |y|^\alpha, & \alpha \neq 1 \\ \left(\frac{2}{\pi}\right) y * \ln(|y|), & \alpha = 1. \end{cases}$$

In this case, $\mu^*(\cdot)$ is equal to $\Lambda(\cdot)/\Lambda(\mathbb{S}^m)$. Furthermore, (2.2) implies that, for any $y > 0$,

$$nP\left(\frac{\|\varepsilon_1\|}{a_n} > y\right) \rightarrow c_0 y^{-\alpha},$$

as $n \rightarrow \infty$, where c_0 is some constant; see Resnick (1986). We choose a_n as follows:

$$a_n = \inf\{x : P(\|\varepsilon_1\| > x) < n^{-1}\}.$$

Then, $a_n = n^{1/\alpha} L(n)$, where $L(n)$ is a slowly varying function; see Bingham, Goldie and Teugels (1989). For example, when ε_t is defined as in (5.1) in Section 5, it has the following density function:

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

Figure 1 (a) and (b) show plots of $f(x, y)$ when $\alpha = 0.8$ and 1.6, respectively.

We simultaneously determine the cointegrating rank r_o and the lag order p_o , in conjunction with an oracle-like efficient estimation of the cointegrating matrix

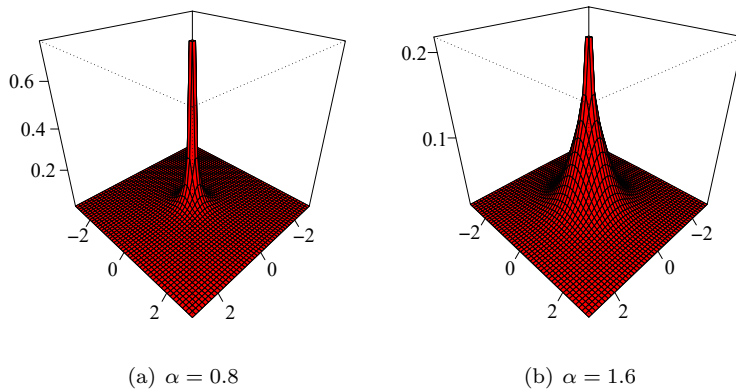


Figure 1. Density $f(x, y)$.

and transient dynamics. When $r_o = 0$, we simply take $\mathbf{\Pi}_o = \mathbf{0}$. Let $\alpha_{o,\perp}$ and $\beta_{o,\perp}$ be the matrices composed of normalized left and right eigenvectors, respectively, corresponding to the zero eigenvalues in $\mathbf{\Pi}_o$. Then, $\alpha_{o,\perp}$ and $\beta_{o,\perp}$ are $m \times (m - r_o)$ full-rank matrices, and are orthogonal complements of α_o and β_o , respectively. Denote $\mathbf{Q} = [\beta_o, \alpha_{o,\perp}]'$. Following the same arguments as those in Liao and Phillips (2015), we can show that $\mathbf{Q}^{-1} = [\alpha_o(\beta_o' \alpha_o)^{-1}, \beta_{o,\perp}(\alpha'_{o,\perp} \beta_{o,\perp})^{-1}]$,

$$\mathbf{Q}\mathbf{\Pi}_o = \begin{bmatrix} \beta_o' \alpha_o \beta_o' \\ \mathbf{0} \end{bmatrix}, \quad \text{and} \quad \mathbf{Q}\mathbf{\Pi}_o \mathbf{Q}^{-1} = \begin{bmatrix} \beta_o' \alpha_o \mathbf{0} \\ \mathbf{0} \quad \mathbf{0} \end{bmatrix}.$$

Thus, the cointegrating rank r_o is the nonzero row vector count of $\mathbf{Q}\mathbf{\Pi}_o$. It follows that a consistent selection of the cointegration rank r_o is equivalent to determining the number of zero rows in $\mathbf{Q}\mathbf{\Pi}_o$. Because of this, \mathbf{Q} plays an important role in our approach.

The row vectors of $\mathbf{Q}\mathbf{\Pi}_o$ are denoted by $\Phi'(\mathbf{\Pi}_o) = [\Phi'_1(\mathbf{\Pi}_o), \Phi'_2(\mathbf{\Pi}_o), \dots, \Phi'_m(\mathbf{\Pi}_o)]$. Let $\mathcal{S}_\phi = \{k : \Phi_k(\mathbf{\Pi}_o) \neq 0\}$ be the index set of nonzero rows of $\mathbf{Q}\mathbf{\Pi}_o$ and, similarly, let $\mathcal{S}_\phi^c = \{k : \Phi_k(\mathbf{\Pi}_o) = 0\}$ denote the index set of zero rows of $\mathbf{Q}\mathbf{\Pi}_o$. From the definition of \mathbf{Q} , we know that $\mathcal{S}_\phi = \{1, \dots, r_o\}$ and $\mathcal{S}_\phi^c = \{r_o + 1, \dots, m\}$. Let $\mathbf{B}_o = [\mathbf{B}_{o,1}, \dots, \mathbf{B}_{o,p}]$, and denote the index set of the zero components in \mathbf{B}_o as \mathcal{S}_B^c , such that $\|\mathbf{B}_{o,j}\| = 0$, for all $j \in \mathcal{S}_B^c$, and $\|\mathbf{B}_{o,j}\| \neq 0$ otherwise. The ordinary least squares (OLS) estimate of $(\mathbf{\Pi}_o, \mathbf{B}_o)$, denoted by $(\hat{\mathbf{\Pi}}_{1st}, \hat{\mathbf{B}}_{1st})$, is the minimizer of the objective function

$$L_n(\mathbf{\Pi}, \mathbf{B}) = \sum_{t=1}^n \left\| \Delta \mathbf{Y}_t - \mathbf{\Pi} \mathbf{Y}_{t-1} - \sum_{j=1}^p \mathbf{B}_j \Delta \mathbf{Y}_{t-j} \right\|^2,$$

where $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_p]$.

Denote \mathbf{Q}_n as the normalized left eigenvector matrix of eigenvalues of $\widehat{\Pi}_{1st}$, and the last $m - r_o$ row of \mathbf{Q}_n is an estimator of $\alpha'_{o,\perp}$. The true parameters are estimated by the penalized LS estimation

$$(\widehat{\Pi}_n, \widehat{\mathbf{B}}_n) = \underset{\Pi, \mathbf{B}_1, \dots, \mathbf{B}_p \in R^{m \times m}}{\operatorname{argmin}} \left\{ L_n(\Pi, \mathbf{B}) + n \sum_{j=1}^p \lambda_{b,j,n} \|\mathbf{B}_j\| + n \sum_{k=1}^m \lambda_{r,k,n} \|\Phi_{n,k}(\Pi)\| \right\}, \quad (2.3)$$

where $\widehat{\mathbf{B}}_n = [\widehat{\mathbf{B}}_{n,1}, \dots, \widehat{\mathbf{B}}_{n,p}]$, $\lambda_{b,j,n}$ ($j = 1, \dots, p$) and $\lambda_{r,k,n}$ ($k = 1, \dots, m$) are tuning parameters that directly control the penalization, and $\Phi_{n,k}(\Pi)$ is the k th row vector of $\mathbf{Q}_n \Pi$. The penalty function on the coefficients \mathbf{B}_j of the lagged differences is called the group Lasso penalty. The penalty function on Π differs from the group Lasso because it works on the rows of the adaptively transformed matrix $\mathbf{Q}_n \Pi$, not the rows of Π directly. Given the tuning parameters, this procedure delivers an estimator of model (2.1) with an implied estimate of r_o (based on the number of nonzero rows of $\mathbf{Q}_n \widehat{\Pi}_n$) and an implied estimate of the transient dynamic structure (including the order p) based on the fitted value $\widehat{\mathbf{B}}_n$.

The determinant of a square matrix \mathbf{A} is denoted by $|A|$, and the $M \times M$ identity matrix is denoted by \mathbf{I}_M . We first state the following assumption.

Assumption 1. (i) *The determinantal equation $|C(z)| = 0$ has roots on or outside the unit circle, where*

$$C(z) = \Pi_o z + \sum_{j=0}^p \mathbf{B}_{o,j} (1-z) z^j \text{ with } \mathbf{B}_{o,0} = -\mathbf{I}_m;$$

(ii) *the matrix $\alpha'_{o,\perp} (\mathbf{I}_m - \sum_{j=1}^p \mathbf{B}_{o,j}) \beta_{o,\perp}$ is nonsingular.*

From Ahn and Reinsel (1990) and Johansen (1995), Assumption 1 leads to the following partial sum Granger representation:

$$\mathbf{Y}_t = \mathbf{C}_B \sum_{s=1}^t \varepsilon_s + \Xi(L) \varepsilon_t + \mathbf{C}_B \mathbf{Y}_0, \quad (2.4)$$

where $\mathbf{C}_B = \beta_{o,\perp} (\alpha'_{o,\perp} (\mathbf{I}_m - \sum_{j=1}^p \mathbf{B}_{o,j}) \beta_{o,\perp})^{-1} \alpha'_{o,\perp}$, and $\Xi(L) \varepsilon_t$ is a stationary process. From the partial sum in (2.4), one can deduce that $\beta'_o \mathbf{Y}_t$ and $\Delta \mathbf{Y}_{t-j}$ are stationary. Denote $\Delta \mathbf{X}_{t-1} = [\Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p}]'$. Model (2.1) can be written

as

$$\Delta \mathbf{Y}_t = [\mathbf{\Pi}_o \ \mathbf{B}_o] \begin{bmatrix} \mathbf{Y}_{t-1} \\ \Delta \mathbf{X}_{t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t.$$

Denote a matrix \mathbf{Q}_B and its inverse as follows:

$$\mathbf{Q}_B = \begin{pmatrix} \boldsymbol{\beta}'_o & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mp} \\ \boldsymbol{\alpha}'_{o,\perp} & \mathbf{0} \end{pmatrix} \text{ and } \mathbf{Q}_B^{-1} = \begin{pmatrix} \boldsymbol{\alpha}_o(\boldsymbol{\beta}'_o\boldsymbol{\alpha}_o)^{-1} & \mathbf{0} & \boldsymbol{\beta}_{o,\perp}(\boldsymbol{\alpha}'_{o,\perp}\boldsymbol{\beta}_{o,\perp})^{-1} \\ \mathbf{0} & \mathbf{I}_{mp} & \mathbf{0} \end{pmatrix}.$$

Then,

$$\mathbf{Z}_{t-1} \equiv \mathbf{Q}_B \begin{bmatrix} \mathbf{Y}_{t-1} \\ \Delta \mathbf{X}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{1,t-1} \\ \mathbf{Z}_{2,t-1} \end{bmatrix}, \quad (2.5)$$

where $\mathbf{Z}'_{1,t-1} = [\mathbf{Y}'_{t-1}\boldsymbol{\beta}_o \ \Delta \mathbf{X}'_{t-1}]$ is a stationary process and $\mathbf{Z}_{2,t-1} = \boldsymbol{\alpha}'_{o,\perp}\mathbf{Y}_{t-1}$ is an I(1) process.

To study the asymptotic properties of $(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n)$, we need one additional assumption, where we use a stochastic integral result in Kurtz and Protter (1991) for limiting properties.

Assumption 2. $\boldsymbol{\varepsilon}_1$ has a symmetric distribution.

Theorem 1. Suppose that (2.2) and Assumptions 1 and 2 are satisfied. If $\delta_{r,n} \equiv \max_{k \in \mathcal{S}_\phi} \lambda_{r,k,n} = o_p(1)$ and $\delta_{b,n} \equiv \max_{j \in \mathcal{S}_B} \lambda_{b,j,n} = o_p(1)$, then the LS shrinkage estimator $(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n)$ is consistent; that is, $(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n) - (\boldsymbol{\Pi}_o, \mathbf{B}_o) = o_p(1)$.

Theorem 1 implies that the nonzero eigenvalues of $\boldsymbol{\Pi}_o$ are estimated as nonzeros asymptotically, which implies that the rank of $\boldsymbol{\Pi}_o$ will not be under-selected. However, consistency of the estimates of the nonzero eigenvalues is not necessary for a consistent cointegration rank selection. As mentioned by Liao and Phillips (2015), what is essential is that the probability limits of these estimates are not zero, or at least that their rates of convergence are slower than those of the estimates of the zero eigenvalues.

Define $\tilde{a}_n = \inf\{x : P(\|\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}'_2\| > x) < n^{-1}\}$. Note that $a_n^2/\tilde{a}_n = n^{1/\alpha}\tilde{L}(n)$, where $\tilde{L}(n)$ is a slowly varying function. The rate of convergence of $(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n)$ is given in the following theorem.

Theorem 2. Let $\delta_n = \delta_{r,n} + \delta_{b,n}$. Under the conditions of Theorem 1,

(a) if $r_o = m$, then $a_n^2/\tilde{a}_n[(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n) - (\boldsymbol{\Pi}_o, \mathbf{B}_o)] = O_p(1 + n\tilde{a}_n^{-1}\delta_n)$.

(b) if $0 \leq r_o < m$, $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$, then

$$\frac{a_n^2}{\tilde{a}_n}[(\widehat{\boldsymbol{\Pi}}_n - \boldsymbol{\Pi}_o)\bar{\boldsymbol{\beta}}, \widehat{\mathbf{B}}_n - \mathbf{B}_o] = O_p(1 + n\tilde{a}_n^{-1}\delta_n),$$

$$n(\widehat{\boldsymbol{\Pi}}_n - \boldsymbol{\Pi}_o)\bar{\boldsymbol{\beta}}_{\perp} = O_p(1 + na_n^{-2}\delta_n).$$

(c) if $0 \leq r_o < m$, $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$, then

$$n[(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n) - (\boldsymbol{\Pi}_o, \mathbf{B}_o)]\mathbf{Q}_B^{-1} = O_p(1).$$

The terms $\delta_{r,n}$ and $\delta_{b,n}$ represent the shrinkage bias that the penalty function introduces to the LS shrinkage estimator. Denote

$$\mathbf{D}_{n,B} = \begin{cases} \frac{1}{n}\mathbf{I}_{m(1+p)} & \text{if } \alpha \in (0, 1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow \infty, \\ \text{diag}\left\{\frac{\tilde{a}_n}{a_n^2}\mathbf{I}_{r_o+mp}, \frac{1}{n}\mathbf{I}_d\right\} & \text{if } \alpha \in (1, 2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow 0. \end{cases}$$

If the rates of convergence of $\lambda_{r,k,n}(k \in \mathcal{S}_{\phi})$ and $\lambda_{b,j,n}(j \in \mathcal{S}_B)$ are fast enough such that $n^{1-1/\alpha}(\delta_{r,n} + \delta_{b,n}) = O_p(1)$, then Theorem 2 implies that $(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n) - (\boldsymbol{\Pi}_o, \mathbf{B}_o) = O_p(n^{-1})$ when $r_o = 0$ and $O_p(n^{-\frac{1}{\alpha}}\tilde{L}(n)^{-1})$ when $r_o = m$, and $[(\widehat{\boldsymbol{\Pi}}_n, \widehat{\mathbf{B}}_n) - (\boldsymbol{\Pi}_o, \mathbf{B}_o)]\mathbf{Q}_B^{-1}\mathbf{D}_{n,B}^{-1} = O_p(1)$ otherwise; that is, the LS shrinkage estimators have the same rates of convergence as the OLS estimators $(\widehat{\boldsymbol{\Pi}}_{1st}, \widehat{\mathbf{B}}_{1st})$. In the next section, we give a condition on the tuning parameters such that the zero rows of $\mathbf{Q}\boldsymbol{\Pi}_o$ and the zero matrices in \mathbf{B}_o are estimated as zero with probability approaching one (w.p.a.1).

3. Oracle Properties

This section shows that the LS shrinkage estimator is oracle efficient in the sense that it has the same asymptotic distribution as the RLSE when the true cointegration rank and lagged differences are known. We subdivide the matrix \mathbf{Q}_n as $\mathbf{Q}'_n = [\mathbf{Q}'_{\alpha,n}, \mathbf{Q}'_{\alpha_{\perp},n}]$, where $\mathbf{Q}_{\alpha,n}$ and $\mathbf{Q}_{\alpha_{\perp},n}$ are the first r_o rows and the last $m - r_o$ rows of \mathbf{Q}_n , respectively. Under Lemma 3 and Theorem 1,

$$\mathbf{Q}_{\alpha,n}\widehat{\boldsymbol{\Pi}}_n = \mathbf{Q}_{\alpha,n}\widehat{\boldsymbol{\Pi}}_{1st} + o_p(1) = \boldsymbol{\Lambda}_{\alpha,n}\mathbf{Q}_{\alpha,n} + o_p(1), \tag{3.1}$$

and, similarly,

$$\mathbf{Q}_{\alpha_{\perp},n}\widehat{\boldsymbol{\Pi}}_n = \mathbf{Q}_{\alpha_{\perp},n}\widehat{\boldsymbol{\Pi}}_{1st} + o_p(1) = \boldsymbol{\Lambda}_{\alpha_{\perp},n}\mathbf{Q}_{\alpha_{\perp},n} + o_p(1) = o_p(1), \tag{3.2}$$

where $\boldsymbol{\Lambda}_{\alpha,n} = \text{diag}[\phi_1(\widehat{\boldsymbol{\Pi}}_{1st}), \dots, \phi_{r_o}(\widehat{\boldsymbol{\Pi}}_{1st})]$, $\boldsymbol{\Lambda}_{\alpha_{\perp},n} = \text{diag}[\phi_{r_o+1}(\widehat{\boldsymbol{\Pi}}_{1st}), \dots, \phi_m(\widehat{\boldsymbol{\Pi}}_{1st})]$, and $\phi_k(\widehat{\boldsymbol{\Pi}}_{1st})$ denotes the k th largest eigenvalues of $\widehat{\boldsymbol{\Pi}}_{1st}$, for $k = 1, \dots, m$. Here, (3.1) implies that the first r_o rows of $\mathbf{Q}_n\widehat{\boldsymbol{\Pi}}_n$ are nonzero w.p.a.1, and (3.2) implies that the last $m - r_o$ rows of $\mathbf{Q}_n\widehat{\boldsymbol{\Pi}}_n$ are arbitrarily close to zero

w.p.a.1. Denote

$$\tau = \begin{cases} 2 - \frac{2}{\alpha} & \text{if } \alpha \in (0, 1) \\ 1 - \frac{1}{\alpha} & \text{if } \alpha \in [1, 2) \end{cases}.$$

We have the following theorem.

Theorem 3. *If the tuning parameters satisfy $n^{1-1/\alpha}(\delta_{b,n} + \delta_{r,n}) = O_p(1)$, $n^{1-2/\alpha} \tilde{L}(n)^{-1} \lambda_{r,k,n} \rightarrow_p \infty$, for $k \in \mathcal{S}_\phi^c$, and $n^\tau \tilde{L}(n)^{-1} \lambda_{b,j,n} \rightarrow_p \infty$, for $j \in \mathcal{S}_B^c$, then it follows that for all $j \in \mathcal{S}_B^c$, as $n \rightarrow \infty$,*

$$P(\mathbf{Q}_{\alpha,\perp,n} \hat{\boldsymbol{\Pi}}_n = \mathbf{0}) \rightarrow 1 \quad \text{and} \quad P(\hat{\mathbf{B}}_{n,j} = \mathbf{0}_{m \times m}) \rightarrow 1. \quad (3.3)$$

Theorem 3 indicates that the zero rows of $\mathbf{Q}\boldsymbol{\Pi}_o$ (and, hence, the zero eigenvalues of $\boldsymbol{\Pi}_o$) and the zero matrices in \mathbf{B}_o are estimated as zeros w.p.a.1. This implies a consistent selection of the cointegration rank r_o and the lag order p_o .

Corollary 1. *Under the conditions of Theorem 3, it follows that as $n \rightarrow \infty$,*

$$P(r(\hat{\boldsymbol{\Pi}}_n) = r_o) \rightarrow 1 \quad \text{and} \quad P(\hat{\mathbf{B}}_{n,j} = \mathbf{0}) \rightarrow 1 \quad \text{for } j \in \mathcal{S}_B^c.$$

We next derive the asymptotic distribution of $(\hat{\boldsymbol{\Pi}}_n, \hat{\mathbf{B}}_{\mathcal{S}_B})$, where $\hat{\mathbf{B}}_{\mathcal{S}_B}$ denotes the LS shrinkage estimator of the nonzero matrices in \mathbf{B}_o . Let $\mathbf{I}_{\mathcal{S}_B} = \text{diag}(\mathbf{I}_{1,m}, \dots, \mathbf{I}_{d_{\mathcal{S}_B},m})$, where $\mathbf{I}_{j,m}$ ($j = 1, \dots, d_{\mathcal{S}_B}$) are $m \times m$ identity matrices, and $d_{\mathcal{S}_B}$ is the dimensionality of the index set \mathcal{S}_B . Define

$$\mathbf{Q}_{\mathcal{S}} = \begin{pmatrix} \boldsymbol{\beta}'_o & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathcal{S}_B} \\ \boldsymbol{\alpha}'_{o,\perp} & \mathbf{0} \end{pmatrix},$$

and

$$\mathbf{D}_{n,\mathcal{S}} = \begin{cases} \frac{1}{n} \mathbf{I}_{m+\mathcal{S}_B} & \text{if } \alpha \in (0, 1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow \infty, \\ \text{diag} \left\{ \frac{\tilde{a}_n}{\tilde{a}_n^2} \mathbf{I}_{r_o+\mathcal{S}_B}, \frac{1}{n} \mathbf{I}_d \right\} & \text{if } \alpha \in (1, 2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow 0, \end{cases}$$

where $\mathbf{I}_{\mathcal{S}_B} = \mathbf{I}_{md_{\mathcal{S}_B}}$ in $\mathbf{Q}_{\mathcal{S}}$ serves to accommodate the nonzero matrices in \mathbf{B}_o . Let $\Delta \mathbf{X}_{\mathcal{S},t}$ denote the nonzero lagged differences in (2.1). The true model can be written as

$$\Delta \mathbf{Y}_t = \boldsymbol{\Pi}_o \mathbf{Y}_{t-1} + \mathbf{B}_{o,\mathcal{S}_B} \Delta \mathbf{X}_{\mathcal{S},t-1} + \boldsymbol{\varepsilon}_t, \quad (3.4)$$

where the transformed and reduced regressor variables are

$$\mathbf{Z}_{S,t-1} = \mathbf{Q}_S \begin{bmatrix} \mathbf{Y}_{t-1} \\ \Delta \mathbf{X}_{S,t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{1S,t-1} \\ \mathbf{Z}_{2,t-1} \end{bmatrix}, \tag{3.5}$$

with $\mathbf{Z}'_{1S,t-1} = [\mathbf{Y}'_{t-1}\boldsymbol{\beta}_o \ \Delta \mathbf{X}'_{S,t-1}]$ and $\mathbf{Z}_{2,t-1} = \boldsymbol{\alpha}'_{o,\perp} \mathbf{Y}_{t-1}$. By Theorem 4.2 in Johansen (1995) and Assumption 1, we have the following expansions:

$$\mathbf{Z}_{1S,t} = \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-i} \quad \text{and} \quad \mathbf{Z}_{2,t} = [\mathbf{I}_d, \mathbf{0}] \sum_{i=1}^t \boldsymbol{\gamma}_i,$$

where $\mathbf{A}_i = O(\rho^i)$, $\boldsymbol{\gamma}_t = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t-i}$, $d = m - r_o$, and $\boldsymbol{\psi}_i = O(\rho^i)$, with some $\rho \in (0, 1)$.

To ensure identification, we normalize $\boldsymbol{\beta}_o$ as $\boldsymbol{\beta}_o = [\mathbf{I}_{r_o}, \mathbf{O}_{r_o}]'$, where \mathbf{O}_{r_o} is some $r_o \times (m - r_o)$ matrix such that $\boldsymbol{\Pi}_o = \boldsymbol{\alpha}_o \boldsymbol{\beta}'_o = [\boldsymbol{\alpha}_o, \boldsymbol{\alpha}_o \mathbf{O}_{r_o}]$. Let $\bar{\boldsymbol{\beta}}_{\perp} = \boldsymbol{\beta}_{o,\perp} (\boldsymbol{\alpha}'_{o,\perp} \boldsymbol{\beta}_{o,\perp})^{-1} = [\bar{\boldsymbol{\beta}}'_{\perp,1}, \bar{\boldsymbol{\beta}}'_{\perp,2}]'$ and $\bar{\boldsymbol{\beta}} = \boldsymbol{\alpha}_o (\boldsymbol{\beta}'_o \boldsymbol{\alpha}_o)^{-1} = [\bar{\boldsymbol{\beta}}'_1, \bar{\boldsymbol{\beta}}'_2]'$, where $\bar{\boldsymbol{\beta}}_{\perp,2}$ and $\bar{\boldsymbol{\beta}}_2$ are the last $m - r_o$ rows of $\bar{\boldsymbol{\beta}}_{\perp}$ and $\bar{\boldsymbol{\beta}}$, respectively. Because the cointegrating rank is consistently selected as r_o , the LS shrinkage estimator $\hat{\boldsymbol{\Pi}}_n$ can be decomposed as $\hat{\boldsymbol{\alpha}}_n \hat{\boldsymbol{\beta}}'_n$, where $\hat{\boldsymbol{\alpha}}_n$ is the first r_o columns of $\hat{\boldsymbol{\Pi}}_n$ and $\hat{\boldsymbol{\beta}}_n = [\mathbf{I}_{r_o}, \hat{\mathbf{O}}_n]$. Here, \rightarrow_d denotes convergence in distribution. We have the following result.

Theorem 4. *Under the conditions of Theorem 3, if $\delta_{r,n} + \delta_{b,n} = o_p(1)$ when $\alpha \in (0, 1)$ and $= o_p(n^{1/\alpha-1})$ when $\alpha \in [1, 2)$, then it follows that*

- (a) $n(\hat{\mathbf{O}}_n - \mathbf{O}_{r_o}) \rightarrow_d (\boldsymbol{\alpha}'_o \boldsymbol{\alpha}_o)^{-1} \boldsymbol{\alpha}'_o \mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1} \bar{\boldsymbol{\beta}}_{\perp,2}^{-1}$,
- (b) $n^{1/\alpha} \tilde{L}(n)(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_o, \hat{\mathbf{B}}_{S_B} - \mathbf{B}_{o,S_B}) \rightarrow_d \mathbf{R}_1 \boldsymbol{\Gamma}_{11}^{-1}$, when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$,
- (c) $n(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_o, \hat{\mathbf{B}}_{S_B} - \mathbf{B}_{o,S_B}) \rightarrow_d \boldsymbol{\alpha}_o (\boldsymbol{\alpha}'_o \boldsymbol{\alpha}_o)^{-1} \boldsymbol{\alpha}'_o \mathbf{F}_1 + [\mathbf{I}_m - \boldsymbol{\alpha}_o (\boldsymbol{\alpha}'_o \boldsymbol{\alpha}_o)^{-1} \boldsymbol{\alpha}'_o] \mathbf{F}_2$, when $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$,

where $\mathbf{R}_1 = \sum_{i=0}^{\infty} \mathbf{S}_{i+2} \mathbf{A}'_i$, $\boldsymbol{\Gamma}_{21} = \mathbf{R}'_2 \sum_{i=0}^{\infty} \mathbf{A}'_i + [\mathbf{I}_d, \mathbf{0}] \sum_{i=0}^{\infty} \sum_{j=0}^i \boldsymbol{\psi}_j \mathbf{S}_1 \mathbf{A}'_i$, with $\boldsymbol{\psi} = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i$, $\boldsymbol{\Gamma}_{11} = \sum_{l=0}^{\infty} \mathbf{A}_l \mathbf{S}_1 \mathbf{A}'_l$, $\mathbf{F}_1 = -\mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1} \boldsymbol{\Gamma}_{21} \boldsymbol{\Gamma}_{11}^{-1} - \mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1} \bar{\boldsymbol{\beta}}_{\perp,2}^{-1} \bar{\boldsymbol{\beta}}_2 [\mathbf{I}_{r_o}, \mathbf{0}]$, $\mathbf{F}_2 = \mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1} \mathbf{R}'_2 (\boldsymbol{\alpha}'_o \boldsymbol{\alpha}_o)^{-1} \boldsymbol{\alpha}'_o [\mathbf{I}_{r_o}, \mathbf{0}] \boldsymbol{\Gamma}_{11}^{-1}$, and $\{\mathbf{S}_i\}$, \mathbf{R}_2 , and $\boldsymbol{\Gamma}_{22}$ are defined in Lemma 2 in the appendix.

Remark 1. If we replace \mathbf{A} , \mathbf{B} , \mathbf{B}_{\perp} , $\bar{\mathbf{B}}_{\perp}$, and $\bar{\mathbf{B}}$ in She and Ling (2020) with $\boldsymbol{\alpha}_o$, $\boldsymbol{\beta}'_o$, $\boldsymbol{\alpha}'_{o,\perp}$, $\bar{\boldsymbol{\beta}}_{\perp}$ and $\bar{\boldsymbol{\beta}}$, respectively, then the limiting distributions in Theorem 4 are the same as those in She and Ling (2020) when r_o and p_o are known; that is, the estimates $\hat{\mathbf{O}}_n$, $\hat{\boldsymbol{\alpha}}_n$, and $\hat{\mathbf{B}}_{S_B}$ achieve their oracle property.

Recently, the oracle properties of the Lasso were studied by Kock and Calot (2015) and Basu and Michailidis (2015) for the vector AR model, and by Liang and Schienle (2019) for the VEC model when $m \rightarrow \infty$. However, they need to assume that the vector noise $\{\varepsilon_t\}$ is i.i.d. normal or $E\|\varepsilon_t\|^{4+\delta} < \infty$, with $\delta > 0$. These assumptions do not reflect heavy-tailed time series, such as that in model (2.1). This remains a challenging problem for the Lasso procedure of the heavy-tailed VEC model when the dimension $m \rightarrow \infty$.

4. Adaptive Selection of Tuning Parameters

This section develops a data-driven procedure for selecting the tuning parameters $\{\lambda_{r,k,n}\}_{k=1}^m$ and $\{\lambda_{b,j,n}\}_{j=1}^p$. As presented in Theorem 3, the conditions require that the tuning parameters related to the zero and nonzero components have different asymptotic behaviors. It is clear that some sort of adaptive penalization should appear in $\lambda_{b,j,n}$ and $\lambda_{r,k,n}$. One popular choice of a penalty is the adaptive Lasso penalty in Zou (2006),

$$\lambda_{r,k,n} = \frac{\lambda_{r,k,n}^*}{\|\phi_k(\widehat{\boldsymbol{\Pi}}_{1st})\|^\omega} \quad \text{and} \quad \lambda_{b,j,n} = \frac{m^\omega \lambda_{b,j,n}^*}{\|\widehat{\mathbf{B}}_{1st,j}\|^\omega}, \tag{4.1}$$

where $\lambda_{r,k,n}^*$ and $\lambda_{b,j,n}^*$ are nonincreasing positive sequences, and ω is some positive finite constant. The extra term m^ω is used to adjust the effect of the dimensionality of \mathbf{B}_j on the adaptive penalty. We introduce the notation τ_1 and τ_2 as

$$\tau_1 = \begin{cases} 0 & \text{if } \alpha \in (0, 1), \\ 1 - \frac{1}{\alpha} & \text{if } \alpha \in [1, 2), \end{cases} \quad \text{and} \quad \tau_2 = \begin{cases} 2 - \frac{2}{\alpha} + \omega & \text{if } \alpha \in (0, 1), \\ 1 - \frac{1}{\alpha} + \frac{\omega}{\alpha} & \text{if } \alpha \in [1, 2). \end{cases}$$

The following lemma gives the conditions under which the tuning parameters $\lambda_{r,k,n}$ and $\lambda_{b,j,n}$ satisfy the assumptions of Theorem 3 and Theorem 4.

Lemma 1. *If $\lambda_{r,k,n}^* = o_p(n^{-\tau_1})$, $\lambda_{b,j,n}^* = o_p(n^{-\tau_1})$, and $n^{1-2/\alpha} n^\omega \tilde{L}(n)^{-1} \lambda_{r,k,n}^* \rightarrow \infty$ and $n^{\tau_2} \tilde{L}(n)^{-1} \lambda_{b,j,n}^* \rightarrow \infty$, for any $j = 1, \dots, p$ and $k = 1, \dots, m$, then under Assumption 1 and (2.2), for any $k \in \mathcal{S}_\phi^c$ and $j \in \mathcal{S}_B^c$, it follows that*

$$\delta_{r,n} + \delta_{b,n} = o_p(n^{-\tau_1}), \quad n^{1-2/\alpha} \tilde{L}(n)^{-1} \lambda_{r,k,n} \rightarrow_p \infty \quad \text{and} \quad n^\tau \tilde{L}(n)^{-1} \lambda_{b,j,n} \rightarrow_p \infty.$$

We now discuss the choice of $\lambda_{r,k,n}^*$ and $\lambda_{b,j,n}^*$. From Lemma 1, we see that the conditions on the tuning parameters that ensure the oracle properties in the LS shrinkage estimation only restrict the rates at which the sequences $\lambda_{r,k,n}^*$ and $\lambda_{b,j,n}^*$ go to zero. However, these conditions are not precise enough to provide

a clear choice of tuning parameters in finite samples. On the one hand, $\lambda_{r,k,n}^*$ and $\lambda_{b,j,n}^*$ should converge to zero as fast as possible so that nonzero $\mathbf{T}(k)$ and nonzero $\mathbf{B}_{o,j}$ are not estimated as zero, with a high probability. This reduces the shrinkage bias in the estimation of the nonzero components of the model. On the other hand, these sequences should converge to zero as slowly as possible so that the zero components in the model are estimated as zeros with a high probability in finite samples. To achieve a balance, we recommend choosing $\lambda_{r,k,n}^*$ and $\lambda_{b,j,n}^*$ as follows:

$$\lambda_{r,k,n}^* = c_{r,k}n^{-1/2(\omega+1-2/\alpha)} \text{ and } \lambda_{b,j,n}^* = c_{b,j}n^{-\tau_2/2}, \tag{4.2}$$

where $\omega \geq 2/\alpha$, and $c_{r,k}$ and $c_{b,j}$ are some constants. It is not hard to see that (4.2) satisfies the condition of Lemma 1.

To understand (4.2), we further discuss the Karuch–Kuhn–Tucker (KKT) conditions (A.6) and (A.9) in the appendix. Let $\mathbf{P}_n = \mathbf{Q}_n^{-1}$, and let $\mathbf{P}_n(k)$ be the k th column of \mathbf{P}_n . Denote

$$F_{\pi,n}(k) = \sum_{t=1}^n \left(\Delta \mathbf{Y}_t - \widehat{\boldsymbol{\Pi}}_n \mathbf{Y}_{t-1} - \sum_{j=1}^p \widehat{\mathbf{B}}_{n,j} \Delta \mathbf{Y}_{t-j} \right)' \mathbf{P}_n(k) \mathbf{Y}'_{t-1}, \tag{4.3}$$

$$F_{b,n}(j) = \sum_{t=1}^n \left(\Delta \mathbf{Y}_t - \widehat{\boldsymbol{\Pi}}_n \mathbf{Y}_{t-1} - \sum_{j=1}^p \widehat{\mathbf{B}}_{n,j} \Delta \mathbf{Y}_{t-j} \right) \Delta \mathbf{Y}'_{t-j}, \tag{4.4}$$

for $k = 1, \dots, m$ and $j = 1, \dots, p$. Denote $\mathbf{T} = \mathbf{Q}_n \boldsymbol{\Pi}_o$, and let $\mathbf{T}(k)$ be the k th row of the matrix \mathbf{T} . Note that $\mathbf{T}_n \equiv \mathbf{Q}_n \widehat{\boldsymbol{\Pi}}_n$ is an estimator of \mathbf{T} . From KKT conditions (A.6) and (A.9) in the appendix, the k th row of \mathbf{T} is estimated as zero, and the component $\mathbf{B}_{o,j}$ in \mathbf{B}_o will be estimated as zero only if the following condition holds:

$$\|a_n^{-2} F_{\pi,n}(k)\| < \frac{n^{1-2/\alpha} \tilde{L}(n)^{-1} \lambda_{r,k,n}^*}{2 \|\phi_k(\widehat{\boldsymbol{\Pi}}_{1st})\|^\omega} \text{ and} \tag{4.5}$$

$$\|\tilde{a}_n^{-1} F_{b,n}(j)\| < \frac{n^{1-1/\alpha} \tilde{L}(n)^{-1} \lambda_{b,j,n}^*}{2 \|\widehat{\mathbf{B}}_{1st,j}\|^\omega}. \tag{4.6}$$

First, by Lemma 3, we have $n\phi_k(\widehat{\boldsymbol{\Pi}}_{1st}) = O_p(1)$, for $k \in \mathcal{S}_\phi^c$, and $\widehat{\mathbf{B}}_{1st,j} = O_p(n^{-1/\alpha} \tilde{L}(n)^{-1})$ if $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$, and is equal to $O_p(n^{-1})$ if $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$, for $j \in \mathcal{S}_B^c$. By Lemma S.1 in the Supplementary Material, if

$$n^{1-2/\alpha} n^\omega \tilde{L}(n)^{-1} \lambda_{r,k,n}^* \rightarrow \infty \text{ and } n^{\tau_2} \tilde{L}(n)^{-1} \lambda_{b,j,n}^* \rightarrow \infty, \tag{4.7}$$

then (4.5) holds, which implies that a zero $\mathbf{T}(k)$ or $\mathbf{B}_{o,j}$ is estimated as zero w.p.a.1. By Lemma 3, we have $\phi_k(\widehat{\boldsymbol{\Pi}}_{1st}) \rightarrow_p \phi_k(\boldsymbol{\Pi}_o) \neq 0$ and $\widehat{\mathbf{B}}_{1st,j} \rightarrow_p \mathbf{B}_{o,j} \neq 0$ for $k \in \mathcal{S}_\phi$ and $j \in \mathcal{S}_B$. By Lemma S.1 in the Supplementary Material, if

$$n^{1-2/\alpha} \tilde{L}(n)^{-1} \lambda_{r,k,n}^* \rightarrow 0 \text{ and } n^\tau \tilde{L}(n)^{-1} \lambda_{b,j,n}^* \rightarrow 0, \quad (4.8)$$

then (4.5) cannot hold, which implies that a nonzero $\mathbf{T}(k)$ or $\mathbf{B}_{o,j}$ is not estimated as zero w.p.a.1. It is not hard to check that (4.2) guarantees (4.7) and that (4.8) holds.

We next discuss how to choose $c_{r,k}$ and $c_{b,j}$. First, take $\lambda_{b,j,n}^* = 2 \log(n) n^{-\tau_2/2}$ and $\lambda_{r,k,n}^* = 2 \log(n) n^{-(1/2)(\omega+1-2/\alpha)}$, which satisfy the conditions of Lemma 1. We perform a first-step LS shrinkage estimation to obtain estimates $(\hat{\mathbf{T}}_{1,\pi}, \hat{\mathbf{T}}_{2,\pi})$ of $(\mathbf{T}_{1,\pi_0}, \mathbf{T}_{2,\pi_0})$. Following similar arguments to those in Liao and Phillips (2015), we select $c_{r,k}$ and c_b as

$$\hat{c}_{r,k} = 2 \|\mathbf{Q}_n(k) \hat{\mathbf{T}}_{1,\pi}\| \times \|\hat{\mathbf{T}}_{2,\pi}\| \text{ and } \hat{c}_{b,j} = 2 \left\| n^{-2/\alpha} \sum_{t=1}^n \Delta \mathbf{Y}_{t-j} \Delta \mathbf{Y}'_{t-j} \right\|.$$

We further choose $\omega = 2/\alpha$. The data-dependent tuning parameters for the LS shrinkage estimation are given as follows:

$$\lambda_{r,k,n} = 2n^{-1/2} \|\mathbf{Q}_n(k) \hat{\mathbf{T}}_{1,\pi}\| \times \|\hat{\mathbf{T}}_{2,\pi}\| \times \|\phi_k(\widehat{\boldsymbol{\Pi}}_{1st})\|^{-2/\alpha}, \text{ and}$$

$$\lambda_{b,j,n} = \begin{cases} 2n^{(-1/2+1/2\alpha-1/\alpha^2)} \left\| n^{-2/\alpha} \sum_{t=1}^n \Delta \mathbf{Y}_{t-j} \Delta \mathbf{Y}'_{t-j} \right\| \times \left(\frac{\|\widehat{\mathbf{B}}_{1st,j}\|}{m} \right)^{-2/\alpha}, & \text{if } \alpha \in (1, 2) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow 0, \\ 2n^{-1} \left\| n^{-2/\alpha} \sum_{t=1}^n \Delta \mathbf{Y}_{t-j} \Delta \mathbf{Y}'_{t-j} \right\| \times \left(\frac{\|\widehat{\mathbf{B}}_{1st,j}\|}{m} \right)^{-2/\alpha}, & \text{if } \alpha \in (0, 1) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow \infty. \end{cases}$$

The tail index α is unknown, in practice. Based on Theorem A.2 in She and Ling (2020), we can estimate $2/\alpha$ by $\log(\text{Tr}[\sum_{t=1}^n \Delta \mathbf{Y}_t \Delta \mathbf{Y}'_t]) / \log n$. The simulation results in the next section show that these data-dependent tuning parameters work well.

5. Simulation Study

This section examines the performance of the shrinkage estimates in terms of their cointegrating rank selection and efficient estimation in finite samples. First, we investigate the model when $m = 2$. Here, $\{\mathbf{Y}_t\}_{t=1}^n$ are generated from

Table 1. Rank and lagged order selection with adaptive Lasso penalty for model (5.1).

		Cointegration rank selection								
		$r_o = 0$			$r_o = 1$			$r_o = 2$		
		n=100	n=400	n=800	n=100	n=400	n=800	n=100	n=400	n=800
$\alpha = 0.2$	$\hat{r}_n = 0$	0.731	0.991	0.997	0.014	0.003	0.000	0.007	0.001	0.000
	$\hat{r}_n = 1$	0.262	0.009	0.003	0.956	0.997	1.000	0.013	0.002	0.000
	$\hat{r}_n = 2$	0.007	0.000	0.000	0.030	0.000	0.000	0.980	0.997	1.000
$\alpha = 1.3$	$\hat{r}_n = 0$	0.801	1.000	1.000	0.001	0.000	0.000	0.000	0.000	0.000
	$\hat{r}_n = 1$	0.198	0.000	0.000	0.833	0.979	0.999	0.002	0.000	0.000
	$\hat{r}_n = 2$	0.001	0.000	0.000	0.166	0.021	0.001	0.998	1.000	1.000
		Lagged difference selection								
		$r_o = 0$			$r_o = 1$			$r_o = 2$		
		n=100	n=400	n=800	n=100	n=400	n=800	n=100	n=400	n=800
$\alpha = 0.2$	$\hat{p}_n \in T$	0.800	0.971	1.000	0.916	0.996	0.995	0.946	0.978	0.991
	$\hat{p}_n \in C$	0.185	0.029	0.000	0.070	0.000	0.000	0.038	0.000	0.000
	$\hat{p}_n \in I$	0.015	0.000	0.000	0.014	0.004	0.005	0.016	0.022	0.009
$\alpha = 1.3$	$\hat{p}_n \in T$	0.802	0.922	0.996	0.734	0.998	1.000	0.645	0.827	0.995
	$\hat{p}_n \in C$	0.198	0.078	0.004	0.266	0.002	0.000	0.355	0.144	0.005
	$\hat{p}_n \in I$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.000

Note: "T" denotes selection of the true lags model (i.e., a model with $p = 3$ and $\hat{\mathbf{B}}_{n,2} = 0$), "C" denotes the selection of a consistent lags model (i.e., a model with $p = 3$ and $\hat{\mathbf{B}}_{n,2} \neq 0$), and "I" denotes the selection of an inconsistent lags model (i.e., a model with p not selected as 3).

$$\begin{pmatrix} \Delta \mathbf{Y}_{1,t} \\ \Delta \mathbf{Y}_{2,t} \end{pmatrix} = \mathbf{\Pi}_o \begin{pmatrix} \mathbf{Y}_{1,t-1} \\ \mathbf{Y}_{2,t-1} \end{pmatrix} + \sum_{j=1}^{10} \mathbf{B}_{o,j} \begin{pmatrix} \Delta \mathbf{Y}_{1,t-1} \\ \Delta \mathbf{Y}_{2,t-1} \end{pmatrix} + \boldsymbol{\varepsilon}_t, \quad (5.1)$$

with $\boldsymbol{\varepsilon}_t = |x_t|^{1/\alpha}(\cos \zeta_t, \sin \zeta_t)$, where $x_t \sim i.i.d.$ Cauchy distribution and $\zeta_t \sim i.i.d. U[0, 2\pi]$, and they are independent. The initial observation \mathbf{Y}_0 is set to be zero. Furthermore, $\mathbf{\Pi}_o$ is specified as follows:

$$\mathbf{\Pi}_o = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -0.5 \\ 1 & 0.5 \end{bmatrix}, \text{ and } \begin{bmatrix} -0.5 & 0.1 \\ 0.2 & -0.4 \end{bmatrix},$$

which corresponds to cointegrating ranks of zero, one, and two, respectively. In addition, $\mathbf{B}_{o,1}$ and $\mathbf{B}_{o,3}$ are taken to be $diag(0.4, 0.4)$, and other $\mathbf{B}_{o,j} = \mathbf{0}$. We take sample sizes $n = 100, 400, 800$, and use 1,000 replications. The tail index $\alpha = 0.2$ and $\alpha = 1.3$. Model (5.1) is over-parameterized according to the true model that generates the data set.

First, we are interested in how well the shrinkage method selects the correct rank of $\mathbf{\Pi}_o$ and the order of the lagged differences. Table 1 shows the finite-sample probabilities of the shrinkage method for a joint rank and lagged order selection. When the sample size is small (i.e., $n = 100$) and the data are *i.i.d.*,

Table 2. Finite-sample properties of the shrinkage estimates of model (5.1).

		$r_o = 1$ and $n=400$							
		Π_{11}		Π_{21}		Π_{12}		Π_{22}	
		Bais	Std	Bais	Std	Bais	Std	Bais	Std
$\alpha = 0.2$	Lasso	0.01793	0.07266	0.01756	0.09658	0.00584	0.05347	0.00734	0.03170
	OLS	0.00297	0.06269	0.00125	0.07926	0.00158	0.03152	0.00018	0.03996
	Oracle	0.00208	0.05298	0.00926	0.08607	0.00118	0.02684	0.00478	0.04335
$\alpha = 1.3$	Lasso	0.01256	0.06628	0.01378	0.09008	0.00332	0.03327	0.00531	0.04602
	OLS	0.00321	0.06513	0.00639	0.06543	0.00177	0.03268	0.00301	0.03280
	Oracle	0.00768	0.05298	0.00317	0.07626	0.00387	0.02718	0.00162	0.03815
		$B_{1,11}$		$B_{1,21}$		$B_{1,12}$		$B_{1,22}$	
		Bais	Std	Bais	Std	Bais	Std	Bais	Std
$\alpha = 0.2$	Lasso	0.10046	0.09332	0.00005	0.12230	0.00117	0.07091	0.09441	0.10929
	OLS	0.00293	0.09335	0.00079	0.10421	0.00161	0.07151	0.00105	0.07690
	Oracle	0.00119	0.09599	0.00032	0.10431	0.00005	0.08780	0.00500	0.08054
$\alpha = 1.3$	Lasso	0.08917	0.06717	0.00204	0.06268	0.01042	0.06485	0.08347	0.08091
	OLS	0.00143	0.07762	0.00784	0.07191	0.00246	0.09634	0.00809	0.06506
	Oracle	0.00449	0.06652	0.00530	0.08124	0.01033	0.07279	0.00195	0.07513
		$B_{2,11}$		$B_{2,21}$		$B_{2,12}$		$B_{2,22}$	
		Bais	Std	Bais	Std	Bais	Std	Bais	Std
$\alpha = 0.2$	Lasso	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	OLS	0.00599	0.08441	0.00241	0.10716	0.00846	0.09003	0.00545	0.08565
	Oracle	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\alpha = 1.3$	Lasso	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	OLS	0.00229	0.07093	0.00576	0.07302	0.00598	0.08524	0.00285	0.07477
	Oracle	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
		$B_{3,11}$		$B_{3,21}$		$B_{3,12}$		$B_{3,22}$	
		Bais	Std	Bais	Std	Bais	Std	Bais	Std
$\alpha = 0.2$	Lasso	0.12461	0.07388	0.02298	0.07245	0.02132	0.04331	0.09549	0.08743
	OLS	0.00187	0.10214	0.00207	0.11513	0.00198	0.10941	0.00376	0.09365
	Oracle	0.00120	0.08595	0.00322	0.08927	0.00023	0.06855	0.00418	0.09054
$\alpha = 1.3$	Lasso	0.10135	0.06506	0.02273	0.06214	0.02229	0.04619	0.04344	0.07441
	OLS	0.00339	0.06372	0.00365	0.07404	0.00223	0.07452	0.00082	0.06865
	Oracle	0.00449	0.06652	0.00530	0.085124	0.01033	0.07279	0.00185	0.07513

Note: Oracle estimate is the RLSE with $r_o = 1$ and the restriction that $\mathbf{B}_{o,2} = 0$.

the probabilities of selecting the true rank $r_o = 2$ when $\alpha \in (1, 2)$ and $\alpha \in (0, 1)$ are almost equal to one. The probability of selecting the true rank $r_o = 1$ when $\alpha \in (0, 1)$ is close to one. The probability of falsely selecting the true rank $r_o = 0$ and $r_o = 1$ when $\alpha \in (0, 1)$ is increased. However, as the sample size grows, the probability of selecting the true rank moves closer to one. When the sample size is increased to 800, the probabilities of selecting the true rank $r_o = 1$ when

$\alpha \in (1, 2)$ and $r_o = 0$ when $\alpha \in (0, 1)$ are almost equal to one. The probabilities of selecting the true rank of the other cases are equal to one. Evidently, the method performs well in selecting the true rank and true lagged differences in all scenarios. This result shows that selecting the true rank also performs well when adding lags to the model.

Table 2 presents the finite-sample properties of the LS shrinkage estimation when $r_o = 1$ and $n = 400$. The corresponding OLS estimates and oracle estimates (i.e., RLSE with $r_o = 1$ and the restriction that $\mathbf{B}_{o,2} = \mathbf{0}$) are also reported in this table. Additional simulation results when $r_o = 0$ and $r_o = 2$ are presented in the Supplementary Material. Table 2 shows that our method performs well in estimating the parameters, overall. When compared with the oracle estimates, some components in the LS shrinkage even have smaller variances, though their finite-sample biases are slightly larger. Moreover, in general, the LS shrinkage estimate has a smaller variance than that of the OLS estimate, though the finite-sample bias of the shrinkage estimate of the nonzero component is slightly larger.

We now carry out a simulation study for the following VEC model:

$$\begin{pmatrix} \Delta \mathbf{Y}_{1,t} \\ \Delta \mathbf{Y}_{2,t} \\ \Delta \mathbf{Y}_{3,t} \end{pmatrix} = \mathbf{\Pi}_o \begin{pmatrix} \mathbf{Y}_{1,t-1} \\ \mathbf{Y}_{2,t-1} \\ \mathbf{Y}_{3,t-1} \end{pmatrix} + \sum_{j=1}^{10} \mathbf{B}_{o,j} \begin{pmatrix} \Delta \mathbf{Y}_{1,t-j} \\ \Delta \mathbf{Y}_{2,t-j} \\ \Delta \mathbf{Y}_{3,t-j} \end{pmatrix} + \boldsymbol{\varepsilon}_t, \quad (5.2)$$

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, $\mathbf{\Pi}_o$ is specified as follows:

$$\mathbf{\Pi}_o = \mathbf{0}_{3 \times 3}, \begin{bmatrix} -0.5 & -0.25 & 0.5 \\ 0.1 & 0.05 & -0.1 \\ 0.2 & 0.1 & -0.2 \end{bmatrix}, \begin{bmatrix} -0.5 & -0.2 & 0.7 \\ 0.1 & -0.3 & 0.2 \\ 0.2 & 0.2 & -0.4 \end{bmatrix}, \text{ and } \begin{bmatrix} -0.4 & 0.0 & 0.0 \\ 0.0 & -0.6 & 0.0 \\ 0.0 & 0.0 & -0.8 \end{bmatrix},$$

which corresponds to cointegrating ranks $r_o = 0, 1, 2, 3$, respectively, and $\mathbf{B}_{o,1}$ and $\mathbf{B}_{o,3}$ are taken to be $diag(0.4, 0.4, 0.4)$ and other $\mathbf{B}_{o,j} = \mathbf{0}$. The sample size is $n = 400$ and 800 , and the number of replications is $1,000$. Table 3 reports the finite-sample probabilities of the rank and lagged order selection for model (5.2). The results show that our method performs well in selecting the true rank and true lagged differences in all scenarios. The estimates of the other parameters when $r_o = 1$ and 2 are reported in the Supplementary Material.

6. An Empirical Example

This section uses the technique in Section 2 for time series modeling of the long-run and short-run behavior of the prices of wheat, corn, and wheat flour

Table 3. Rank and lagged order selection with adaptive Lasso penalty for model (5.2).

		Cointegration rank selection							
		$r_o = 0$		$r_o = 1$		$r_o = 2$		$r_o = 3$	
		n=400	n=800	n=400	n=800	n=400	n=800	n=400	n=800
$\alpha = 0.2$	$\hat{r}_n = 0$	0.667	0.909	0.001	0.000	0.000	0.000	0.001	0.000
	$\hat{r}_n = 1$	0.305	0.087	0.972	0.990	0.003	0.001	0.003	0.002
	$\hat{r}_n = 2$	0.028	0.004	0.027	0.010	0.995	0.999	0.010	0.000
	$\hat{r}_n = 3$	0.000	0.000	0.000	0.000	0.002	0.000	0.986	0.998
$\alpha = 1.3$	$\hat{r}_n = 0$	0.576	0.898	0.000	0.000	0.000	0.000	0.000	0.000
	$\hat{r}_n = 1$	0.379	0.094	0.942	0.994	0.001	0.000	0.001	0.000
	$\hat{r}_n = 2$	0.045	0.008	0.058	0.006	0.997	1.000	0.001	0.001
	$\hat{r}_n = 3$	0.000	0.000	0.000	0.000	0.002	0.000	0.998	0.999
		Lagged difference selection							
		$r_o = 0$		$r_o = 1$		$r_o = 2$		$r_o = 3$	
		n=400	n=800	n=400	n=800	n=400	n=800	n=400	n=800
$\alpha = 0.2$	$\hat{p}_n \in T$	0.854	0.992	0.963	0.989	0.958	0.992	0.830	0.987
	$\hat{p}_n \in C$	0.017	0.008	0.026	0.002	0.002	0.002	0.153	0.000
	$\hat{p}_n \in I$	0.129	0.000	0.011	0.009	0.040	0.006	0.017	0.013
$\alpha = 1.3$	$\hat{p}_n \in T$	0.695	0.983	0.903	0.960	0.697	0.967	0.775	0.987
	$\hat{p}_n \in C$	0.305	0.017	0.097	0.039	0.177	0.004	0.245	0.007
	$\hat{p}_n \in I$	0.000	0.000	0.000	0.001	0.126	0.029	0.000	0.006

in the United States. The sample used in the empirical study includes monthly data over the period June 1987 to May 2017, with 360 observations. These series for the period January 1961 through October 1972 have been considered by Ahn and Reinsel (1988) in investigating the reduced-rank AR model. 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 logarithms of the data, that is, $y_{it} = \log(x_{it})$. The data $\{\mathbf{Y}_t\}$ are shown in Figure 2. Evidently, the time series display the co-movement over the entire period. Therefore, we can use the VEC model to analyze the data and try to reveal some cointegrating relations. That is, we expect the cointegration rank r_o to satisfy $0 < r_o < 3$.

We first use the first k -largest data and the Hill's estimator to estimate the tail index of the log-returns (i.e., $r_{it} = y_{it} - y_{it-1}$) of each price, that is,

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

where $\{|r_i|_{(t)} : t = 1, \dots, n\}$ is the decreasing order statistics of $\{|r_{it}| : t =$

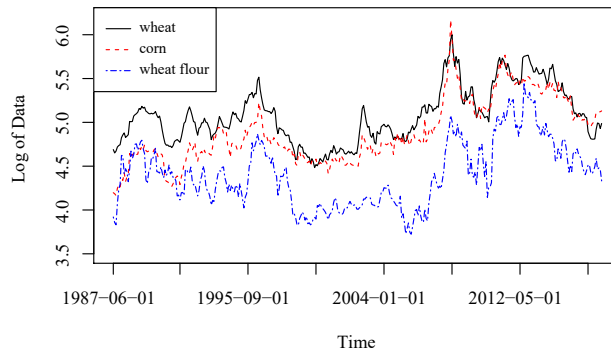


Figure 2. The Logarithms of Monthly Prices of Wheat, Corn, and Wheat Flour.

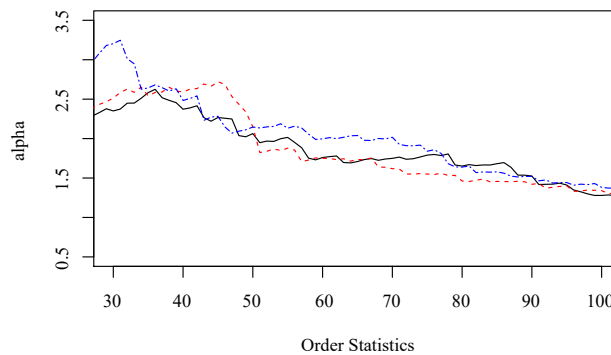


Figure 3. Hill estimator of tail index.

$1, \dots, n$ }; see Resnick (1997). Because $\hat{\alpha}_i(k)$ relies on the choice of k , Figure 3 shows the plots of these estimated tail indices in term of k . It shows that the tail index of each log-return is most likely less than two but larger than one. It seems reasonable to assume these data are heavy-tailed time series.

We apply our shrinkage methods to estimate the following VEC model:

$$\Delta \mathbf{Y}_t = \mathbf{\Pi} \mathbf{Y}_{t-1} + \sum_{k=1}^{10} \mathbf{B}_k \Delta \mathbf{Y}_{t-k} + \varepsilon_t. \quad (6.1)$$

The unrestricted LS estimation of $\mathbf{\Pi}$ produced eigenvalues -0.0578 , -0.0486 , and 0.00003 , which indicates that $\mathbf{\Pi}$ might contain at least one zero eigenvalue, because the last one, 0.00003 , is very close to zero. The LS estimates of the lag coefficients \mathbf{B}_k are nonzero for $k = 1, \dots, 10$. We apply the LS shrinkage estimation to model (6.1). The results are as follows:

$$\hat{\boldsymbol{\Pi}} = \begin{bmatrix} 0.0295 & -0.0069 & -0.0262 \\ 0.1056 & -0.0797 & -0.0327 \\ 0.1485 & -0.0303 & -0.1369 \end{bmatrix},$$

$$\hat{\mathbf{B}}_1 = \begin{bmatrix} 0.1041 & 0.0013 & 0.0054 \\ 0.1046 & -0.0757 & -0.0004 \\ 0.1429 & -0.0308 & -0.0286 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{B}}_5 = \begin{bmatrix} 0.0376 & 0.0013 & -0.0191 \\ 0.0964 & -0.0211 & -0.0071 \\ 0.0534 & 0.0452 & -0.0609 \end{bmatrix},$$

and the other \mathbf{B}_k are estimated as zero. The eigenvalues of $\hat{\boldsymbol{\Pi}}$ are -0.1171 , 0.0684 , and 0 , which implies that the cointegrating rank r_o is two. These results corroborate the manifestation of the co-movement in the three time series through the presence of two cointegrating vectors in the fitted model. In model (6.1), we set $p = 10$. However, the results are the same when we set p from 10 to 15. It seems that our approach is quite stable.

Supplementary Material

In the online Supplementary Material, we provide one preliminary result, additional simulation results, and proofs for Theorem 1, Theorem 2, Lemma 1, and Lemma 3.

Acknowledgments

The work was partially supported by Hong Kong Research Grants Commission Grants (16500117, 16303118, 16301620, 16300621, Theme-based Research Scheme T31-604/18-N), the Australian Research Council, and the National Natural Science Foundation of China (11731015, 11571148) (Shiqing Ling).

A. Appendix

We first state some preliminary results to prove the main theorems. From Proposition 3.1 in Resnick (1986), we can see that the condition (2.2) is equivalent to the following convergence

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

as $n \rightarrow \infty$, where $PRM(\mu)$ is a Poisson random process with intensity measure μ and $\{\mathbf{P}_i\}$ is a sequence of 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

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

as $n \rightarrow \infty$ 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 of random vectors such that $\sum_{i=1}^\infty \delta_{\mathbf{P}_i^{(j)}}$ is the point representation of $PRM(\tilde{\mu})$ for $j = 2, 3, \dots$ and they are independent each other for different j .

Since $\mathbf{Z}_{1,t-1}$ is a stationary process and $\mathbf{Z}_{2,t-1}$ comprises the $I(1)$ components under Assumption 1 and Theorem 4.2 in Johansen (1995). Then we have the following expansions

$$\mathbf{Z}_{1,t} = \sum_{i=0}^\infty \mathbf{B}_i \boldsymbol{\varepsilon}_{t-i} \text{ and } \mathbf{Z}_{2,t} = [\mathbf{I}_d, \mathbf{0}] \sum_{i=1}^t \boldsymbol{\gamma}_i,$$

where $\mathbf{B}_i = O(\rho^i)$ with some $\rho \in (0, 1)$. Denote

$$\begin{aligned} \mathbf{R}_{1n} &= \sum_{t=1}^n \boldsymbol{\varepsilon}_t \mathbf{Z}'_{1,t-1}, & \mathbf{R}_{2n} &= \sum_{t=1}^n \boldsymbol{\varepsilon}_t \mathbf{Z}'_{2,t-1}, \\ \mathbf{S}_{11n} &= \sum_{t=1}^n \mathbf{Z}_{1,t-1} \mathbf{Z}'_{1,t-1}, & \mathbf{S}_{21n} &= \sum_{t=1}^n \mathbf{Z}_{2,t-1} \mathbf{Z}'_{1,t-1} \text{ and } \mathbf{S}_{22n} = \sum_{t=1}^n \mathbf{Z}_{2,t-1} \mathbf{Z}'_{2,t-1}. \end{aligned}$$

By Theorem A.1 and Lemma B.1 in She and Ling (2020), it is straightforward to show the following lemma.

Lemma 2. *Suppose that (2.2) and Assumptions 1–2 hold, and $E\|\boldsymbol{\varepsilon}_1\|^\alpha = \infty$. Then*

- (a). $\frac{1}{\tilde{a}_n} \mathbf{R}_{1n} \rightarrow_d \sum_{i=0}^\infty \mathbf{S}_{i+2} \mathbf{B}'_i,$
- (b). $\frac{1}{a_n^2} \mathbf{R}_{2n} \rightarrow_d \mathbf{R}_2,$
- (c). $\frac{1}{na_n^2} \mathbf{S}_{22n} \rightarrow_d \boldsymbol{\Gamma}_{22},$
- (d). $\frac{1}{a_n^2} \mathbf{S}_{11n} \rightarrow_d \sum_{l=0}^\infty \mathbf{B}_l \mathbf{S}_1 \mathbf{B}'_l,$
- (e). $\frac{1}{a_n^2} \mathbf{S}_{21n} \rightarrow_d \left\{ \mathbf{R}'_2 \sum_{i=0}^\infty \mathbf{B}'_i + [\mathbf{I}_d, \mathbf{0}] \sum_{i=0}^\infty \sum_{j=0}^i \boldsymbol{\psi}_j \mathbf{S}_1 \mathbf{B}'_i \right\}.$

where $\mathbf{R}_2 = [\int_0^1 \mathbf{P}(r) d\mathbf{P}'(r)]' \boldsymbol{\psi}' [\mathbf{I}_d, \mathbf{0}]'$ and $\boldsymbol{\Gamma}_{22} = [\mathbf{I}_d, \mathbf{0}] \boldsymbol{\psi} [\int_0^1 \mathbf{P}(r) \mathbf{P}'(r) dr] \boldsymbol{\psi}' [\mathbf{I}_d, \mathbf{0}]',$

$\mathbf{S}_1 = \sum_{i=1}^{\infty} \mathbf{P}_i^{(1)} \mathbf{P}_i^{(1)'}$ with $\mathbf{P}_i^{(1)} = \mathbf{P}_i$, $\mathbf{S}_j = \sum_{i=1}^{\infty} \mathbf{P}_i^{(j)}$ for all $j > 1$ and $\mathbf{P}(r)$ is a stable process.

We next give the limiting distribution of the OLS estimator $(\widehat{\boldsymbol{\Pi}}_{1st}, \widehat{\mathbf{B}}_{1st})$ and the asymptotic properties of the eigenvalues of $\widehat{\boldsymbol{\Pi}}_{1st}$.

Lemma 3. *Suppose that (2.2) and Assumptions 1–2 hold, and $E\|\boldsymbol{\varepsilon}_1\|^\alpha = \infty$ and \mathbf{S}_1 is positive definite almost surely. Then*

(a). $[(\widehat{\boldsymbol{\Pi}}_{1st}, \widehat{\mathbf{B}}_{1st}) - (\boldsymbol{\Pi}_o, \mathbf{B}_o)] \mathbf{Q}_B^{-1} \mathbf{D}_{n,B}^{-1} \rightarrow_d (\mathbf{B}_{m,1}, \mathbf{B}_{m,2})$, where $\mathbf{B}_{m,2} = \mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1}$,

$$\mathbf{B}_{m,1} = \begin{cases} \mathbf{R}_1^* \boldsymbol{\Gamma}_{11}^{*-1} & \text{if } \alpha \in (1, 2) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow 0, \\ -\mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1} \boldsymbol{\Gamma}_{21}^* \boldsymbol{\Gamma}_{11}^{*-1} & \text{if } \alpha \in (0, 1) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \rightarrow \infty, \end{cases}$$

$$\mathbf{R}_1^* = \sum_{i=0}^{\infty} \mathbf{S}_{i+2} \mathbf{B}'_i, \quad \boldsymbol{\Gamma}_{11}^* = \sum_{l=0}^{\infty} \mathbf{B}_l \mathbf{S}_1 \mathbf{B}'_l \text{ and } \boldsymbol{\Gamma}_{21}^* = \mathbf{R}'_2 \sum_{i=0}^{\infty} \mathbf{B}'_i + [\mathbf{I}_d, \mathbf{0}] \sum_{i=0}^{\infty} \sum_{j=0}^i \boldsymbol{\psi}_j \mathbf{S}_1 \mathbf{B}'_i,$$

(b). for $k = 1, \dots, m$, the eigenvalues of $\widehat{\boldsymbol{\Pi}}_{1st}$ satisfy $\phi_k(\widehat{\boldsymbol{\Pi}}_{1st}) \rightarrow_p \phi_k(\boldsymbol{\Pi}_o)$,

(c). the last $m - r_o$ eigenvalues of $\widehat{\boldsymbol{\Pi}}_{1st}$ satisfy

$$n(\phi_1(\widehat{\boldsymbol{\Pi}}_{1st}), \dots, \phi_{m-r_o}(\widehat{\boldsymbol{\Pi}}_{1st})) \rightarrow_d (\tilde{\phi}_{o,1}, \dots, \tilde{\phi}_{o,m-r_o}),$$

where the $\tilde{\phi}_{o,j} (j = 1, \dots, m - r_o)$ are solutions of the equation $|\mu \mathbf{I}_{m-r_o} - \boldsymbol{\alpha}'_{o,\perp} \mathbf{R}_2 \boldsymbol{\Gamma}_{22}^{-1}| = 0$.

Lemma 3 is used to prove Theorem 1, 3 and Lemma 1. Its proof is given in Supplemental material of this paper.

We subdivide the matrix \mathbf{P}_n as $\mathbf{P}_n = [\mathbf{P}_{\alpha,n}, \mathbf{P}_{\alpha_\perp,n}]$, where $\mathbf{P}_{\alpha,n}$ is the first r_o columns of \mathbf{P}_n ($\mathbf{P}_{\alpha_\perp,n}$ is defined accordingly). Then

$$\mathbf{Q}_{\alpha_\perp,n} \mathbf{P}_{\alpha_\perp,n} = \mathbf{I}_{m-r_o}, \quad \mathbf{Q}_{\alpha,n} \mathbf{P}_{\alpha_\perp,n} = \mathbf{0}_{r_o \times (m-r_o)} \text{ and } \mathbf{Q}_{\alpha_\perp,n} \widehat{\boldsymbol{\Pi}}_{1st} = \boldsymbol{\Lambda}_{\alpha_\perp,n} \mathbf{Q}_{\alpha_\perp,n},$$

where $\boldsymbol{\Lambda}_{\alpha_\perp,n}$ is a diagonal matrix with the ordered last (smallest) $m - r_o$ eigenvalues of $\widehat{\boldsymbol{\Pi}}_{1st}$. Define a useful estimator of $\boldsymbol{\Pi}_o$ as

$$\boldsymbol{\Pi}_{n,f} = \widehat{\boldsymbol{\Pi}}_{1st} - \mathbf{P}_{\alpha_\perp,n} \boldsymbol{\Lambda}_{\alpha_\perp,n} \mathbf{Q}_{\alpha_\perp,n}.$$

$\boldsymbol{\Pi}_{n,f}$ may be interpreted as a modification to the unrestricted estimate $\widehat{\boldsymbol{\Pi}}_{1st}$ which removes components in the eigen-representation of the unrestricted estimate that correspond to the smallest $m - r_o$ eigenvalues. Then

$$\mathbf{Q}_{\alpha,n} \boldsymbol{\Pi}_{n,f} = \mathbf{Q}_{\alpha,n} \widehat{\boldsymbol{\Pi}}_{1st} - \mathbf{Q}_{\alpha,n} \mathbf{P}_{\alpha_\perp,n} \boldsymbol{\Lambda}_{\alpha_\perp,n} \mathbf{Q}_{\alpha_\perp,n} = \boldsymbol{\Lambda}_{\alpha,n} \mathbf{Q}_{\alpha,n}, \tag{A.1}$$

where $\mathbf{\Lambda}_{\alpha,n}$ is a diagonal matrix with the ordered first (largest) r_o eigenvalues of $\widehat{\mathbf{\Pi}}_{1st}$, and more importantly

$$\mathbf{Q}_{\alpha_{\perp},n}\mathbf{\Pi}_{n,f} = \mathbf{Q}_{\alpha_{\perp},n}\widehat{\mathbf{\Pi}}_{1st} - \mathbf{Q}_{\alpha_{\perp},n}\mathbf{P}_{\alpha_{\perp},n}\mathbf{\Lambda}_{\alpha_{\perp},n}\mathbf{Q}_{\alpha_{\perp},n} = \mathbf{0}_{(m-r_o)\times m}. \quad (\text{A.2})$$

From Lemma 2 (b), (A.1) and (A.2), we can deduce that $\mathbf{Q}_{\alpha,n}\mathbf{\Pi}_{n,f}$ is a $r_o \times m$ matrix which is nonzero w.p.a.1 and $\mathbf{Q}_{\alpha_{\perp},n}\mathbf{\Pi}_{n,f}$ is always a $(m - r_o) \times m$ zero matrix for all n . By Lemma 3 (a) and (c), it follows that

$$(\mathbf{\Pi}_{n,f} - \mathbf{\Pi}_o)\mathbf{Q}^{-1}\mathbf{D}_n^{-1} = O_p(1), \quad (\text{A.3})$$

where $\mathbf{D}_n = n^{-1}\mathbf{I}_m$ when $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$, and $= \text{diag}\{\tilde{a}_n/a_n^2 \mathbf{I}_{r_o}, n^{-1}\mathbf{I}_{m-r_o}\}$ when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$.

Proof of Theorem 3. Let $\mathbf{T} = \mathbf{Q}_n\mathbf{\Pi}$. We can rewrite the LS shrinkage estimation problem in (2.3) as

$$\begin{aligned} (\widehat{\mathbf{T}}_n, \widehat{\mathbf{B}}_n) = & \underset{\mathbf{T}, \mathbf{B}_1, \dots, \mathbf{B}_p \in R^{m \times m}}{\text{argmin}} \sum_{t=1}^n \left\| \Delta \mathbf{Y}_t - \mathbf{P}_n \mathbf{T} \mathbf{Y}_{t-1} - \sum_{j=1}^p \mathbf{B}_j \Delta \mathbf{Y}_{t-j} \right\|^2 \\ & + n \sum_{j=1}^p \lambda_{b,j,n} \|\mathbf{B}_j\| + n \sum_{k=1}^m \lambda_{r,k,n} \|\Phi_{n,k}(\mathbf{P}_n \mathbf{T})\|. \end{aligned} \quad (\text{A.4})$$

By definition of (A.4), $\widehat{\mathbf{\Pi}}_n = \mathbf{P}_n \widehat{\mathbf{T}}_n$ and $\widehat{\mathbf{T}}_n = \mathbf{Q}_n \widehat{\mathbf{\Pi}}_n$ for all n , and

$$\widehat{\mathbf{T}}_n = \begin{pmatrix} \mathbf{Q}_{\alpha,n} \widehat{\mathbf{\Pi}}_n \\ \mathbf{Q}_{\alpha_{\perp},n} \widehat{\mathbf{\Pi}}_n \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_{\alpha,n} \widehat{\mathbf{\Pi}}_{1st} \\ \mathbf{Q}_{\alpha_{\perp},n} \widehat{\mathbf{\Pi}}_{1st} \end{pmatrix} + o_p(1),$$

and thus the first result follows if we can show that the last $m - r_o$ rows of $\widehat{\mathbf{T}}_n$ are estimated as zeros w.p.a.1. Note that $\Phi_{n,k}(\mathbf{P}_n \mathbf{T})$ is the k -th row of $\mathbf{Q}_n(\mathbf{P}_n \mathbf{T})$, i.e., $\Phi_{n,k}(\mathbf{P}_n \mathbf{T}) = \mathbf{T}(k)$. The problem in (A.4) can be rewritten as

$$\begin{aligned} (\widehat{\mathbf{T}}_n, \widehat{\mathbf{B}}_n) = & \underset{\mathbf{T}, \mathbf{B}_1, \dots, \mathbf{B}_p \in R^{m \times m}}{\text{argmin}} \sum_{t=1}^n \left\| \Delta \mathbf{Y}_t - \mathbf{P}_n \mathbf{T} \mathbf{Y}_{t-1} - \sum_{j=1}^p \mathbf{B}_j \Delta \mathbf{Y}_{t-j} \right\|^2 \\ & + n \sum_{j=1}^p \lambda_{b,j,n} \|\mathbf{B}_j\| + n \sum_{k=1}^m \lambda_{r,k,n} \|\mathbf{T}(k)\|. \end{aligned} \quad (\text{A.5})$$

Let $\widehat{\mathbf{T}}_n(k)$ be the k -th row of $\widehat{\mathbf{T}}_n$. The Karuch-Kuhn-Tucker (KKT) optimality conditions for $\widehat{\mathbf{T}}_n$ are

$$\begin{cases} F_{\pi,n}(k) = \frac{n\lambda_{r,k,n}}{2} \frac{\widehat{\mathbf{T}}_n(k)}{\|\widehat{\mathbf{T}}_n(k)\|} \text{ if } \widehat{\mathbf{T}}_n(k) \neq \mathbf{0}, \\ \|n^{-1}F_{\pi,n}(k)\| < \frac{\lambda_{r,k,n}}{2} \text{ if } \widehat{\mathbf{T}}_n(k) = \mathbf{0}, \end{cases} \tag{A.6}$$

for $k = 1, \dots, m$, where $F_{\pi,n}(k)$ is defined in (4.3). Conditioned on the event $\{\widehat{\mathbf{T}}_n(k_0) \neq \mathbf{0}\}$ for some k_0 satisfying $r_o < k_0 \leq m$, we obtain the following equation

$$\|a_n^{-2}F_{\pi,n}(k)\| = \frac{na_n^{-2}\lambda_{r,k,n}}{2}. \tag{A.7}$$

By Lemma 2 (b), (c) and (e), and Theorem 1, it follows that

$$\begin{aligned} \frac{1}{a_n^2}F_{\pi,n}(k_0) &= \frac{1}{a_n^2} \sum_{t=1}^n [\varepsilon_t - (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_o)\mathbf{Q}_B^{-1}\mathbf{Z}_{t-1}]'\mathbf{P}_n(k_0)\mathbf{Y}'_{t-1} \\ &= \frac{1}{a_n^2} \sum_{t=1}^n [\varepsilon_t - (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_o)\mathbf{Q}_B^{-1}\mathbf{Z}_{t-1}]'\mathbf{P}_n(k_0)\mathbf{Y}'_{t-1} \\ &= \frac{\mathbf{P}_n(k_0)' \sum_{t=1}^n \varepsilon_t \mathbf{Y}'_{t-1}}{a_n^2} - \frac{\mathbf{P}_n(k_0)'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_o)\mathbf{Q}_B^{-1} \sum_{t=1}^n \mathbf{Z}_{t-1} \mathbf{Y}'_{t-1}}{a_n^2} = O_p(1). \end{aligned} \tag{A.8}$$

By assumptions of tuning parameters, $na_n^{-2}\lambda_{r,k,n} \rightarrow_p \infty$. Furthermore, by (A.7) and (A.8), we must have

$$P(\mathbf{Q}_n(k_0)\widehat{\boldsymbol{\Pi}}_n = \mathbf{0}) = P(\widehat{\mathbf{T}}_n(k_0) = \mathbf{0}) \rightarrow 1 \text{ as } n \rightarrow \infty,$$

for any k_0 such that $r_o < k_0 \leq m$. Thus, we obtain

$$P(\mathbf{Q}_{\alpha_{\perp},n}\widehat{\boldsymbol{\Pi}}_n = \mathbf{0}) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

We next show the second part in (3.3). The KKT optimality conditions for $\widehat{\mathbf{B}}_{n,j}$ are

$$\begin{cases} F_{b,n}(j) = \frac{n\lambda_{b,j,n}\widehat{\mathbf{B}}_{n,j}}{2\|\widehat{\mathbf{B}}_{n,j}\|} \text{ if } \widehat{\mathbf{B}}_{n,j} \neq \mathbf{0}, \\ \|n^{-1}F_{b,n}(j)\| < \frac{\lambda_{b,j,n}}{2} \text{ if } \widehat{\mathbf{B}}_{n,j} = \mathbf{0}, \end{cases} \tag{A.9}$$

for any $j = 1, \dots, p$, where $F_{b,n}(j)$ is defined in (4.4). On the event $\{\widehat{\mathbf{B}}_{n,j} \neq \mathbf{0}_{m \times m}\}$ for some $j \in \mathcal{S}_B^c$, we get the following equation from the optimality conditions

$$\|F_{b,n}(j)\| = \frac{n\lambda_{b,j,n}}{2}. \tag{A.10}$$

Let $\tilde{\delta}_n = \tilde{a}_n^{-1}$ when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$, and $\tilde{\delta}_n = na_n^{-2}$ when

$\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$. Then, (A.10) is equivalent to

$$\left\| \tilde{\delta}_n F_{b,n}(j) \right\| = \frac{n \tilde{\delta}_n \lambda_{b,j,n}}{2}. \tag{A.11}$$

By Lemma 2 (a), (d) and (e), and Theorem 1, we have

$$\begin{aligned} \tilde{\delta}_n F_{b,n}(j) &= \tilde{\delta}_n \sum_{t=1}^n [\varepsilon_t - (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_o) \mathbf{Q}_B^{-1} \mathbf{z}_{t-1}] \Delta \mathbf{Y}'_{t-j} \\ &= \tilde{\delta}_n \sum_{t=1}^n \varepsilon_t \Delta \mathbf{Y}'_{t-j} - \tilde{\delta}_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_o) \mathbf{Q}_B^{-1} \sum_{t=1}^n \mathbf{z}_{t-1} \Delta \mathbf{Y}'_{t-j} = O_p(1). \end{aligned} \tag{A.12}$$

However, under the assumptions of tuning parameters, $n \tilde{\delta}_n \lambda_{b,j,n} \rightarrow_p \infty$, which together with results in (A.11) and (A.12) implies that

$$P(\hat{\mathbf{B}}_{n,j} = \mathbf{0}_{m \times m}) \rightarrow 1 \text{ as } n \rightarrow \infty,$$

for any $j \in \mathcal{S}_B^c$, which finishes the proof.

Proof of Theorem 4. Without loss of generality, we assume the first r_o columns of $\boldsymbol{\Pi}_o$ are linearly independent. Let $\boldsymbol{\beta}_{o,\perp} = (\boldsymbol{\beta}'_{1,o\perp}, \boldsymbol{\beta}'_{2,o\perp})'$, where $\boldsymbol{\beta}_{1,o\perp}$ is a $r_o \times (m - r_o)$ matrix and $\boldsymbol{\beta}_{2,o\perp}$ is a $(m - r_o) \times (m - r_o)$ matrix. By definition of $\boldsymbol{\beta}_o$ and $\boldsymbol{\beta}_{o,\perp}$,

$$\boldsymbol{\beta}'_{1,o\perp} + \boldsymbol{\beta}'_{2,o\perp} \mathbf{O}'_{r_o} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\beta}'_{1,o\perp} \boldsymbol{\beta}_{1,o\perp} + \boldsymbol{\beta}'_{2,o\perp} \boldsymbol{\beta}_{2,o\perp} = \mathbf{I}_{m-r_o},$$

which implies that

$$\boldsymbol{\beta}'_{1,o\perp} = -\boldsymbol{\beta}'_{2,o\perp} \mathbf{O}'_{r_o} \text{ and } \boldsymbol{\beta}_{2,o\perp} = (\mathbf{I}_{m-r_o} + \mathbf{O}'_{r_o} \mathbf{O}_{r_o})^{-1/2}. \tag{A.13}$$

Let $\delta_n^* = a_n^2 / \tilde{a}_n$ when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$, and $= n$ when $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$. We first show

$$\delta_n^* (\hat{\mathbf{B}}_n - \mathbf{B}_o) = O_p(1), \tag{A.14}$$

$$n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_o) = O_p(1), \tag{A.15}$$

$$\delta_n^* (\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_o) = O_p(1). \tag{A.16}$$

When $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$, and $n^{1-1/\alpha}(\delta_{r,n} + \delta_{b,n}) = o_p(1)$, by Theorem 2, we have

$$[\delta_n^* (\hat{\boldsymbol{\Pi}}_n - \boldsymbol{\Pi}_o) \boldsymbol{\alpha}_o (\boldsymbol{\beta}'_o \boldsymbol{\alpha}_o)^{-1}, \delta_n^* (\hat{\mathbf{B}}_n - \mathbf{B}_o), n(\hat{\boldsymbol{\Pi}}_n - \boldsymbol{\Pi}_o) \boldsymbol{\beta}_{o,\perp} (\boldsymbol{\alpha}'_{0,\perp} \boldsymbol{\beta}_{0,\perp})^{-1}] = O_p(1),$$

which implies that (A.14) holds, and

$$\delta_n^*[(\hat{\alpha}_n - \alpha_o)\hat{\beta}'_n + \alpha_o(\hat{\beta}_n - \beta_o)']\alpha_o(\beta'_o\alpha_o)^{-1} \tag{A.17}$$

$$\begin{aligned} &= \delta_n^*(\hat{\Pi}_n - \Pi_0)\alpha_o(\beta'_o\alpha_o)^{-1} = O_p(1), \\ n\hat{\alpha}_n(\hat{\beta}_n - \beta_o)'\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} & \tag{A.18} \\ &= n(\hat{\Pi}_n - \Pi_0)\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = O_p(1). \end{aligned}$$

By the definitions of $\hat{\beta}_n$ and $\beta_{o,\perp}$ and (A.18), we can deduce that

$$\beta'_o\hat{\alpha}_n[n(\hat{\mathbf{O}}_n - \mathbf{O}_{r_o})]'\beta_{2,o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = O_p(1). \tag{A.19}$$

By (A.13) and (A.19), we have

$$n(\hat{\mathbf{O}}_n - \mathbf{O}_{r_o}) = [\beta'_o\alpha_o + o_p(1)]^{-1}O_p(1)(\alpha'_{o,\perp}\beta_{o,\perp})(\mathbf{I}_{m-r_o} + \mathbf{O}'_{r_o}\mathbf{O}_{r_o})^{1/2} = O_p(1). \tag{A.20}$$

Again, by the definition of $\hat{\beta}_n$, (A.20) means that $n(\hat{\beta}_n - \beta_o) = O_p(1)$, that is, (A.15) holds. By (A.15) and (A.17), we know that (A.16) holds. When $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow \infty$ and $\delta_{r,n} + \delta_{b,n} = o_p(1)$, by Theorem 2, similar to the case when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \rightarrow 0$, we can show (A.14)–(A.16) hold.

From Theorem 3, we deduce that $\hat{\alpha}_n$, $\hat{\beta}_n$ and $\hat{\mathbf{B}}_{S_B}$ minimize the following criterion function w.p.a.1,

$$\begin{aligned} V_n(\theta_S) &= \sum_{t=1}^n \left\| \Delta \mathbf{Y}_t - \alpha\beta' \mathbf{Y}_{t-1} - \sum_{j \in S_B} \mathbf{B}_j \Delta \mathbf{Y}_{t-j} \right\|^2 \\ &\quad + n \sum_{k \in S_\phi} \lambda_{r,k,n} \|\Phi_{n,k}(\alpha\beta')\| + n \sum_{j \in S_B} \lambda_{b,j,n} \|\mathbf{B}_j\|. \end{aligned}$$

Define $\mathbf{U}_{1,n}^* = \delta_n^*(\hat{\alpha}_n - \alpha_o)$, $\mathbf{U}_{2,n}^* = n(\hat{\mathbf{O}}_n - \mathbf{O}_{r_o})$ and $\mathbf{U}_{3,n}^* = \delta_n^*(\hat{\mathbf{B}}_{S_B} - \mathbf{B}_{o,S_B})$. Then

$$\begin{aligned} &[(\hat{\Pi}_n - \Pi_0), (\hat{\mathbf{B}}_{S_B} - \mathbf{B}_{o,S_B})]\mathbf{Q}_S^{-1}\mathbf{D}_{n,S}^{-1} \\ &= [\delta_n^*\hat{\alpha}_n(\hat{\beta}_n - \beta_o)'\alpha_o(\beta'_o\alpha_o)^{-1} + \delta_n^*(\hat{\alpha}_n - \alpha_o), \delta_n^*(\hat{\mathbf{B}}_{S_B} - \mathbf{B}_{o,S_B}), \\ n\hat{\alpha}_n(\hat{\beta}_n - \beta_o)'\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}] \\ &= [n^{-1}\delta_n^*\hat{\alpha}_n[\mathbf{0}_{r_o}, \mathbf{U}_{2,n}^*]\alpha_o(\beta'_o\alpha_o)^{-1} + \mathbf{U}_{1,n}^*, \mathbf{U}_{3,n}^*, \hat{\alpha}_n[\mathbf{0}_{r_o}, \mathbf{U}_{2,n}^*]\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}]. \end{aligned}$$

Denote $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) \in R^{m \times r_o} \times R^{r_o \times (m-r_o)} \times R^{m \times m_{S_B}}$ and

$$\mathbf{\Pi}_n(\mathbf{U}) = [n^{-1}\delta_n^*\hat{\alpha}_n[\mathbf{0}_{r_o}, \mathbf{U}_2]\alpha_o(\beta'_o\alpha_o)^{-1} + \mathbf{U}_1, \mathbf{U}_3, \hat{\alpha}_n[\mathbf{0}_{r_o}, \mathbf{U}_2]\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}].$$

Then, $\mathbf{U}_n^* = (\mathbf{U}_{1,n}^*, \mathbf{U}_{2,n}^*, \mathbf{U}_{3,n}^*)$ minimizes the following criterion function

$$\begin{aligned} V_n(\mathbf{U}) &= \sum_{t=1}^n (\|\boldsymbol{\varepsilon}_t - \boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Z}_{S,t-1}\|^2 - \|\boldsymbol{\varepsilon}_t\|^2) \\ &\quad + n \sum_{k \in \mathcal{S}_\phi} \lambda_{r,k,n} [\|\Phi_{n,k}[\boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Q}_S\mathbf{L}_1 + \boldsymbol{\Pi}_o]\| - \|\Phi_{n,k}(\boldsymbol{\Pi}_o)\|] \\ &\quad + n \sum_{j \in \mathcal{S}_B} \lambda_{b,j,n} [\|\boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Q}_S\mathbf{L}_{j+1} + \mathbf{B}_{o,j}\| - \|\mathbf{B}_{o,j}\|], \end{aligned}$$

where $\mathbf{L}_j = \text{diag}(\mathbf{A}_{j,1}, \dots, \mathbf{A}_{j,d_{S_B+1}})$ with $\mathbf{A}_{j,j} = \mathbf{I}_m$ and $\mathbf{A}_{i,j} = \mathbf{0}$ for $i \neq j$ and $j = 1, \dots, d_{S_B+1}$.

For any compact set $K \in R^{m \times r_o} \times R^{r_o \times (m-r_o)} \times R^{m \times m d_{S_B}}$ and any $\mathbf{U} \in K$, there is $\boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Q}_S = O_p(\delta_n^{*-1})$. Then we can deduce that

$$\begin{aligned} &n \sum_{k \in \mathcal{S}_\phi} \lambda_{r,k,n} [\|\Phi_{n,k}[\boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Q}_S\mathbf{L}_1 + \boldsymbol{\Pi}_o]\| - \|\Phi_{n,k}(\boldsymbol{\Pi}_o)\|] \\ &\leq n \lambda_{r,k,n} \|\Phi_{n,k}(\boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Q}_S)\| = O_p(n \delta_n^{*-1} \lambda_{r,k,n}) = o_p(1), \end{aligned} \quad (\text{A.21})$$

$$\text{and } n \sum_{j \in \mathcal{S}_B} \lambda_{b,j,n} [\|\boldsymbol{\Pi}_n(\mathbf{U})\mathbf{D}_{n,S}\mathbf{Q}_S\mathbf{L}_{j+1} + \mathbf{B}_{o,j}\| - \|\mathbf{B}_{o,j}\|] = o_p(1), \quad (\text{A.22})$$

uniformly over $\mathbf{U} \in K$.

Denote $\boldsymbol{\vartheta} = \{[\text{vec}(\mathbf{O}'_{r_o})]', [\text{vec}(\boldsymbol{\alpha}_o, \mathbf{B}_{o,S_B})]'\}'$. From (A.21) and (A.22), we deduce that $\hat{\boldsymbol{\alpha}}_n$, $\hat{\mathbf{O}}_n$ and $\hat{\mathbf{B}}_{S_B}$ minimize the following criterion function w.p.a.1

$$L(\boldsymbol{\vartheta}) = \sum_{t=1}^n \|\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\|^2 \text{ and } \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \Delta \mathbf{Y}_t - \boldsymbol{\alpha}_o[\mathbf{I}_{r_o}, \mathbf{O}_{r_o}] \mathbf{Y}_{t-1} - \mathbf{B}_{o,S_B} \Delta \mathbf{X}_{S,t-1}.$$

Then, using the similar argument as for Theorem 3.1 in She and Ling (2020), we can obtain the limiting distribution.

References

- Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive 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.
- Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* **43**, 1535–1567.
- Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press, Cambridge.

- Caner, M. (1998). Tests for cointegration with infinite variance errors. *Journal of Econometrics* **86**, 155–175.
- Cavaliere, G., Nielsen, H. B. and Rahbek, A. (2015). Bootstrap testing of hypotheses on cointegration relations in vector autoregressive models. *Econometrica* **83**, 813–831.
- Chan, N. H., Ling, S. and Yau, C. Y. (2020). Lasso-based variable selection of ARMA models. *Statistica Sinica* **30**, 1925–1948.
- Chen, K. and Chan K. S. (2011). Subset ARMA selection via the adaptive Lasso. *Statistics and its Interface* **4**, 197–205.
- 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.
- Durrett, R. (2019). *Probability: Theory and Examples*. 5th Edition. Cambridge University Press, Cambridge.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55**, 251–276.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.
- Fan, J. and Li, R. (2002). Variable selection for Cox’s proportional hazards model and frailty model. *The Annals of Statistics* **30**, 74–99.
- Granger, C. W. J. (1983). Co-integrated variables and error-correcting models. USC Discussion Paper. University of California, San Diego.
- Huang, J., Ma, S. and Zhang, C. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica* **18**, 1603–1618.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- Knight, K. and Fu, W. (2000). Asymptotics for Lasso-type estimators. *The Annals of Statistics* **28**, 1356–1378.
- Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* **186**, 325–344.
- Kock, A. B. (2016). Consistent and conservative model selection with the adaptive Lasso in stationary and nonstationary autoregressions. *Econometric Theory* **32**, 243–259.
- Kristensen, D. and Rahbek, A. (2013). Testing and inference in nonlinear cointegrating vector error correction models. *Econometric Theory* **29**, 1238–1288.
- Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability* **19**, 1035–1070.
- Liang, C. and Schienle, M. (2019). Determination of vector error correction models in high dimensions. *Journal of Econometrics* **208**, 418–441.
- Liao, Z. and Phillips, P. C. B. (2015). Automated estimation of vector error correction models. *Econometric Theory* **31**, 581–646.
- 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.

- Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies* **53**, 473–495.
- Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics* **20**, 971–1001.
- Reinsel, G. C. and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, Likelihood ratio test, and forecasting. *Journal of Time Series Analysis* **13**, 353–375.
- 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). Inference in heavy-tailed vector error correction models. *Journal of Econometrics* **214**, 433–450.
- Song, S. and Bickel, P. J. (2011). Large vector autoregressions. *arXiv preprint 1106.3915*.
- Stock, J. H. and Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* **61**, 783–820.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Methodological)* **58**, 267–288.
- Wang, H., Li, G. and Tsai, C. L. (2007). Regression coefficients and autoregressive order shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **69**, 63–78.
- Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics* **40**, 727–758.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **68**, 49–67.
- Zhang, R. and Ling, S. (2015). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* **31**, 880–890.
- Zou, H. (2006). The adaptive Lasso and its oracle properties *Journal of the American Statistical Association* **101**, 1418–1429.

Feifei Guo

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China.

E-mail: 7520200236@bit.edu.cn

Shiqing Ling

Department of Mathematics, Hong Kong University of Science and Technology Kowloon, Hong Kong.

E-mail: maling@ust.hk

Zichuan Mi Department of Statistics, Shanxi University of Finance and Economics, Taiyuan City, Shanxi, China.

E-mail: mizc@sxufe.edu.cn

(Received May 2020; accepted March 2021)