<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-2020-0177</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>
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
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 $n \to \infty$. 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
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 implement 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

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 \( n \to \infty \). 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 co-integrating 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 (2002) 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:

\[ \Delta Y_t = \Pi_0 Y_{t-1} + \sum_{j=1}^{p} B_{o,j} \Delta Y_{t-j} + \varepsilon_t, \]  

(2.1)

where \( \Delta Y_t = Y_t - Y_{t-1} \), \( Y_t \) is an \( m \)-dimensional vector-valued time series,
\( \Pi_o = \alpha_o \beta_o' \), with \( \alpha_o \) and \( \beta_o \) being \( m \times r_o \) full-rank matrices, \( B_{o,j} (j = 1, \ldots, p) \) are \( m \times m \) coefficient matrices, where \( p > \text{true order } p_o \) and \( B_{o,j} = 0 \) if \( j > p_o \), and \( \{\varepsilon_t\} \) 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 \( \beta_o' Y_t \) is a stationary time series. Thus, \( \beta_o \) is called the cointegrating vector or long-run cointegrating relations of \( Y_t \).

The rank \( r_o \) of \( \Pi_o \) is called the cointegrating rank of \( Y_t \), and it measures the number of cointegrating relations in the system. The set of nonzero matrices \( B_{o,j} (j = 1, \ldots, p) \) characterizes the transient dynamics in the systems.

We assume that \( \varepsilon_t \) satisfies the following condition:

\[
nP\left( \frac{\varepsilon_1}{a_n} \in \cdot \right) \xrightarrow{v} \mu(\cdot), \quad (2.2)
\]
as \( n \to \infty \), where \( \mu \) is a Radon measure on \((R^m, \mathcal{B}^m)\), \( a_n \) is an increasing sequence diverging to \( \infty \), and \( \xrightarrow{v} \) means the vague convergence in Durrett (2019, pp.121). Here, (2.2) is called the regular variation function, and is equivalent to there existing a probability measure \( \mu^* \) on the unit sphere \( \mathbb{S}^m \).
in $\mathbb{R}^m$, such that, for any $x > 0$,

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

as $t \to \infty$, where $\alpha > 0$ is called the tail index and $\| \cdot \|$ denotes the Euclidean norm; see Resnick (1986). When $\alpha \in (0, 2)$, $\varepsilon_1$ does not have a finite covariance matrix and is called a heavy-tailed random vector. One class of heavy-tailed random vectors is $\varepsilon_1$ with its characteristic function as follows:

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

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

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

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

$$
nP\left(\frac{\|\varepsilon_1\|}{a_n} > y\right) \to c_0 y^{-\alpha},
$$
as \( n \to \infty \), 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 \( \varepsilon_t \) is defined as in (5.1) in Section 5, it has the following density function:

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

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

![Figure 1](image_url)

**Figure 1:** Density \( f(x, y) \)

We simultaneously determine the cointegrating rank \( r_o \) and the lag order \( p_o \), in conjunction with an oracle-like efficient estimation of the coin-
integrating matrix and transient dynamics. When \( r_o = 0 \), we simply take \( \Pi_o = 0 \). 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 \( \Pi_o \). Then, \( \alpha_{o,\perp} \) and \( \beta_{o,\perp} \) are \( m \times (m - r_o) \) full-rank matrices, and are orthogonal complements of \( \alpha_o \) and \( \beta_o \), respectively. Denote \( Q = [\beta_o, \alpha_{o,\perp}]' \).

Following the same arguments as those in Liao and Phillips (2015), we can show that
\[
Q^{-1} = \begin{bmatrix}
\alpha_o(\beta_o\alpha_o)^{-1}, & \beta_{o,\perp}(\alpha_{o,\perp}\beta_{o,\perp})^{-1}
\end{bmatrix},
\]
\[
Q\Pi_o = \begin{bmatrix}
\beta_o'\alpha_o & 0 \\
0 & 0
\end{bmatrix}, \quad \text{and} \quad Q\Pi_o Q^{-1} = \begin{bmatrix}
\beta_o'\alpha_o & 0 \\
0 & 0
\end{bmatrix}.
\]

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

The row vectors of \( Q\Pi_o \) are denoted by \( \Phi'(\Pi_o) = [\Phi_1'(\Pi_o), \Phi_2'(\Pi_o), \ldots, \Phi_m'(\Pi_o)] \). Let \( S_\phi = \{k : \Phi_k(\Pi_o) \neq 0\} \) be the index set of nonzero rows of \( Q\Pi_o \) and, similarly, let \( S_\phi^c = \{k : \Phi_k(\Pi_o) = 0\} \) denote the index set of zero rows of \( Q\Pi_o \). From the definition of \( Q \), we know that \( S_\phi = \{1, \ldots, r_o\} \) and \( S_\phi^c = \{r_o + 1, \ldots, m\} \). Let \( B_o = [B_{o,1}, \ldots, B_{o,p}] \), and denote the index set
of the zero components in $B_o$ as $S_B^c$, such that $\|B_{o,j}\| = 0$, for all $j \in S_B^c$, and $\|B_{o,j}\| \neq 0$ otherwise. The ordinary least squares (OLS) estimate of $(\Pi_o, B_o)$, denoted by $(\hat{\Pi}_{1st}, \hat{B}_{1st})$, is the minimizer of the objective function

$$L_n(\Pi, B) = \sum_{t=1}^{n} \|\Delta Y_t - \Pi Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j}\|^2,$$

where $B = [B_1, \ldots, B_p]$.

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

$$(\hat{\Pi}_n, \hat{B}_n) = \arg\min_{\Pi, B_1, \ldots, B_p \in \mathbb{R}^{m \times m}} \{L_n(\Pi, B)
+ n \sum_{j=1}^{p} \lambda_{b,j,n} \|B_j\| + n \sum_{k=1}^{m} \lambda_{r,k,n} \|\Phi_n k(\Pi)\|\}, \quad (2.3)$$

where $\hat{B}_n = [\hat{B}_{n,1}, \ldots, \hat{B}_{n,p}]$, $\lambda_{b,j,n}$ ($j = 1, \ldots, p$) and $\lambda_{r,k,n}$ ($k = 1, \ldots, m$) are tuning parameters that directly control the penalization, and $\Phi_n k(\Pi)$ is the $k$th row vector of $Q_n \Pi$. The penalty function on the coefficients $B_j$ of the lagged differences is called the group Lasso penalty. The penalty function on $\Pi$ differs from the group Lasso because it works on the rows of the adaptively transformed matrix $Q_n \Pi$, not the rows of $\Pi$ 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 \( Q_n \tilde{\Pi}_n \)) and an implied estimate of the transient dynamic structure (including the order \( p \)) based on the fitted value \( \tilde{\mathbf{B}}_n \).

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

\[
C(z) = \Pi o z + \sum_{j=0}^{p} B_{o,j} (1 - z) z^j \quad \text{with} \quad B_{o,0} = -\mathbf{I}_m;
\]

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

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

\[
\mathbf{Y}_t = \mathbf{C}_B \sum_{s=1}^{t} \mathbf{\varepsilon}_s + \mathbf{\Xi}(L)\mathbf{\varepsilon}_t + \mathbf{C}_B \mathbf{Y}_0, \quad (2.4)
\]

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

$$\Delta Y_t = [\Pi_o B_o] \begin{bmatrix} Y_{t-1} \\ \Delta X_{t-1} \end{bmatrix} + \varepsilon_t.$$ 

Denote a matrix $Q_B$ and its inverse as follows:

$$Q_B = \begin{pmatrix} \beta'_o & 0 \\ 0 & I_{mp} \end{pmatrix} \quad \text{and} \quad Q_B^{-1} = \begin{pmatrix} (\alpha'_o \beta'_o)^{-1} & 0 & \beta_{o,\perp} (\alpha'_o \beta_{o,\perp})^{-1} \\ 0 & I_{mp} & 0 \end{pmatrix}.$$ 

Then,

$$Z_{t-1} \equiv Q_B \begin{bmatrix} Y_{t-1} \\ \Delta X_{t-1} \end{bmatrix} = \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix},$$

where $Z_{1,t-1} = [Y'_{t-1} \beta_o \Delta X'_{t-1}]$ is a stationary process and $Z_{2,t-1} = \alpha'_{o,\perp} Y_{t-1}$ is an $I(1)$ process.

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

**Assumption 2.** $\varepsilon_1$ has a symmetric distribution.

**Theorem 1.** Suppose that (2.2) and Assumptions 1 and 2 are satisfied. If
\[ \delta_{r,n} \equiv \max_{k \in S} \lambda_{r,k,n} = o_p(1) \quad \text{and} \quad \delta_{b,n} \equiv \max_{j \in S} \lambda_{b,j,n} = o_p(1), \]
then the LS shrinkage estimator \((\hat{\Pi}_n, \hat{B}_n)\) is consistent; that is, \((\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o) = o_p(1)\).

Theorem 1 implies that the nonzero eigenvalues of \(\Pi_o\) are estimated as nonzeros asymptotically, which implies that the rank of \(\Pi_o\) 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 \(\tilde{a}_n = \inf\{x : P(\|\varepsilon_1 \varepsilon_2\| > x) < n^{-1}\}\). 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 \((\hat{\Pi}_n, \hat{B}_n)\) 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 
\[ a_n^2/\tilde{a}_n[(\hat{\Pi}_n - \Pi_o), (\hat{B}_n - B_o)] = O_p(1 + n\tilde{a}_n^{-1}\delta_n). \]

(b) if \(0 \leq r_o < m\), \(\alpha \in (1, 2)\) or \(\alpha = 1\) and \(\tilde{L}(n) \to 0\), then 
\[ a_n^2/\tilde{a}_n[(\hat{\Pi}_n - \Pi_o), \hat{\beta}, (\hat{B}_n - B_o)] = O_p(1 + n\tilde{a}_n^{-1}\delta_n), \]
\[ n(\hat{\Pi}_n - \Pi_o)\hat{\beta}_\perp = O_p(1 + n\tilde{a}_n^{-2}\delta_n). \]
(c) if $0 \leq r_o < m$, $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$, then

$$n[(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o)]Q^{-1}_B = O_p(1).$$

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

$$D_{n,B} = \begin{cases} \frac{1}{n}I_{m(1+p)} & \text{if } \alpha \in (0, 1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty, \\ \text{diag}\{ \frac{\tilde{a}_n}{\tilde{a}_n}I_{r_o+mp}, \frac{1}{n}I_d \} & \text{if } \alpha \in (1, 2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0. \end{cases}$$

If the rates of convergence of $\lambda_{r,k,n}(k \in S_p)$ and $\lambda_{b,j,n}(j \in S_B)$ are fast enough such that $n^{1-\frac{1}{\alpha}}(\delta_{r,n} + \delta_{b,n}) = O_p(1)$, then Theorem 2 implies that $(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o) = O_p(n^{-1})$ when $r_o = 0$ and $O_p(n^{-\frac{1}{\alpha}}\tilde{L}(n)^{-1})$ when $r_o = m$, and $[(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o)]Q^{-1}_B D^{-1}_{n,B} = O_p(1)$ otherwise; that is, the LS shrinkage estimators have the same rates of convergence as the OLS estimators $(\Pi_{1st}, B_{1st})$. In the next section, we give a condition on the tuning parameters such that the zero rows of $Q\Pi_o$ and the zero matrices in $B_o$ 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 $Q_n$ as $Q'_n = [Q'_{α,n}, Q'_{α⊥,n}]$, where $Q_{α,n}$ and $Q_{α⊥,n}$ are the first
$\rho_o$ rows and the last $m - \rho_o$ rows of $Q_n$, respectively. Under Lemma 3 and
Theorem 1,

$$Q_{α,n}\hat{\Pi}_n = Q_{α,n}\hat{\Pi}_{1st} + o_p(1) = Λ_{α,n}Q_{α,n} + o_p(1), \quad (3.1)$$

and, similarly,

$$Q_{α⊥,n}\hat{\Pi}_n = Q_{α⊥,n}\hat{\Pi}_{1st} + o_p(1) = Λ_{α⊥,n}Q_{α⊥,n} + o_p(1) = o_p(1), \quad (3.2)$$

where $Λ_{α,n} = \text{diag}[ϕ_1(\hat{\Pi}_{1st}), ..., ϕ_{\rho_o}(\hat{\Pi}_{1st})]$, $Λ_{α⊥,n} = \text{diag}[ϕ_{\rho_o+1}(\hat{\Pi}_{1st}), ..., ϕ_{m}(\hat{\Pi}_{1st})]$, and $ϕ_k(\hat{\Pi}_{1st})$ denotes the kth largest eigenvalues of $\hat{\Pi}_{1st}$, for $k = 1, ..., m$.

Here, (3.1) implies that the first $\rho_o$ rows of $Q_n\hat{\Pi}_n$ are nonzero w.p.a.1, and
(3.2) implies that the last $m - \rho_o$ rows of $Q_n\hat{\Pi}_n$ are arbitrarily close to zero
w.p.a.1. Denote

$$τ = \begin{cases} 2 - \frac{2}{α} & \text{if } α ∈ (0,1) \\ 1 - \frac{1}{α} & \text{if } α ∈ [1,2) \end{cases}.$$ 

We have the following theorem.

**Theorem 3.** If the tuning parameters satisfy $n^{1-\frac{1}{α}}(δ_{b,n} + δ_{r,n}) = O_p(1)$,
\[ n^{1-\frac{2}{\alpha}} \tilde{L}(n)^{-1} \lambda_{r,k,n} \to_p \infty, \text{ for } k \in S^c_o, \text{ and } n^r \tilde{L}(n)^{-1} \lambda_{b,j,n} \to_p \infty, \text{ for } j \in S^c_B, \]

then it follows that for all \( j \in S^c_B, \) as \( n \to \infty, \)

\[ P(Q_{o,1,n,\hat{\Pi}_n} = 0) \to 1 \text{ and } P(\hat{B}_{n,j} = 0_{m \times m}) \to 1. \quad (3.3) \]

Theorem 3 indicates that the zero rows of \( Q \Pi_o \) (and, hence, the zero eigenvalues of \( \Pi_o \)) and the zero matrices in \( B_o \) 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 \( n \to \infty, \)

\[ P(r(\hat{\Pi}_n) = r_o) \to 1 \text{ and } P(\hat{B}_{n,j} = 0) \to 1 \text{ for } j \in S^c_B. \]

We next derive the asymptotic distribution of \((\hat{\Pi}_n, \hat{B}_{SB})\), where \( \hat{B}_{SB} \) denotes the LS shrinkage estimator of the nonzero matrices in \( B_o. \) Let \( I_{SB} = \text{diag}(I_{1,m}, \ldots, I_{d_{SB},m}), \) where \( I_{j,m}(j = 1, \ldots, d_{SB}) \) are \( m \times m \) identity matrices, and \( d_{SB} \) is the dimensionality of the index set \( S_B. \) Define

\[ Q_S = \begin{pmatrix} \beta'_o & 0 \\ 0 & I_{SB} \\ \alpha'_{o,\perp} & 0 \end{pmatrix}. \]
and

\[ D_{n,S} = \begin{cases} \frac{1}{n} I_{m+S_B} & \text{if } \alpha \in (0, 1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty, \\ \text{diag}\{\frac{2}{n} I_{r{o+S_B}}, \frac{1}{n} I_d\} & \text{if } \alpha \in (1, 2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0, \end{cases} \]

where \( I_{S_B} = I_{md_S} \) in \( Q_S \) serves to accommodate the nonzero matrices in \( B_o \). Let \( \Delta X_{S,t} \) denote the nonzero lagged differences in (2.1). The true model can be written as

\[ \Delta Y_t = \Pi_o Y_{t-1} + B_o S_B \Delta X_{S,t-1} + \varepsilon_t, \quad (3.4) \]

where the transformed and reduced regressor variables are

\[ Z_{S,t-1} = Q_S \begin{bmatrix} Y_{t-1} \\ \Delta X_{S,t-1} \end{bmatrix} = \begin{bmatrix} Z_{1S,t-1} \\ Z_{2,t-1} \end{bmatrix}, \quad (3.5) \]

with \( Z_{1S,t-1} = [Y_{t-1}' \ b_o \ \Delta X_{S,t-1}'] \) and \( Z_{2,t-1} = \alpha_{o,1} Y_{t-1} \). By Theorem 4.2 in Johansen (1995) and Assumption 1, we have the following expansions:

\[ Z_{1S,t} = \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} \quad \text{and} \quad Z_{2,t} = [I_d, 0]\sum_{i=1}^{t} \gamma_i, \]

where \( A_i = O(\rho^i) \), \( \gamma_i = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \), \( d = m - r_o \), and \( \psi_i = O(\rho^i) \), with
some $\rho \in (0, 1)$.

To ensure identification, we normalize $\beta_o$ as $\beta_o = [I_{r_o}, O_{r_o}]'$, where $O_{r_o}$ is some $r_o \times (m - r_o)$ matrix such that $\Pi_o = \alpha_o \beta_o' = [\alpha_o, \alpha_o O_{r_o}]$. Let $\beta_\perp = \beta_{o, \perp} (\alpha_{o, \perp} \beta_{o, \perp})^{-1} = [\beta_{\perp, 1}, \beta_{\perp, 2}]'$ and $\tilde{\beta} = \alpha_o (\beta_o' \alpha_o)^{-1} = [\beta_1, \beta_2]'$, where $\beta_{\perp, 2}$ and $\beta_2$ are the last $m - r_o$ rows of $\beta_\perp$ and $\tilde{\beta}$, respectively. Because the cointegrating rank is consistently selected as $r_o$, the LS shrinkage estimator $\hat{\Pi}_n$ can be decomposed as $\hat{\alpha}_n \tilde{\beta}_n'$, where $\hat{\alpha}_n$ is the first $r_o$ columns of $\hat{\Pi}_n$ and $\hat{\beta}_n = [I_{r_o}, \hat{O}_n]$. 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 $= o_p(\sqrt{n})^{-1}$ when $\alpha \in [1, 2]$, then it follows that

(a) $n(\hat{\Pi}_n - O_{r_o}) \rightarrow_d (\alpha'_o \alpha_o)^{-1} \alpha'_o R_2 \Gamma_2^{-1} \beta_{\perp, 2}^{-1}$,

(b) $n^{1/\alpha} \tilde{L}(n)(\hat{\alpha}_n - \alpha_o, \hat{B}_{SB} - B_{o,SB}) \rightarrow_d R_1 \Gamma_1^{-1}$,

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

(c) $n(\hat{\alpha}_n - \alpha_o, \hat{B}_{SB} - B_{o,SB}) \rightarrow_d \alpha_o (\alpha'_o \alpha_o)^{-1} \alpha'_o F_1 + [I_m - \alpha_o (\alpha'_o \alpha_o)^{-1} \alpha'_o] F_2$,

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

where $R_1 = \sum_{i=0}^{\infty} S_{i+2} A'_i$, $\Gamma_{21} = R_2 \sum_{i=0}^{\infty} A'_i$, $\Gamma_{11} = \sum_{i=0}^{\infty} \sum_{j=0}^{i} \psi_j S_i A'_i$, with $\psi = \sum_{i=0}^{\infty} \psi_i$, $\Gamma_{12} = \sum_{i=0}^{\infty} A_i S_i A'_i$, $F_1 = -R_2 \Gamma_2^{-1} \Gamma_{21} \Gamma_1^{-1} - R_2 \Gamma_2^{-1} \beta_{\perp, 2}^{-1} \beta_2 [I_{r_o}, 0]$. 
\[ F_2 = R_2 \Gamma_{22}^{-1} R_2' (\alpha'_o \alpha_o)^{-1} \alpha'_o [I_{r_o}, 0] \Gamma_{11}^{-1}, \] and \( \{ S_i \}, R_2, \) and \( \Gamma_{22} \) are defined in Lemma 2 in the appendix.

**Remark 2.1.** If we replace \( A, B, \tilde{B}_\perp, \) and \( \tilde{B} \) in She and Ling (2020) with \( \alpha_o, \beta'_o, \alpha'_{o, \perp}, \beta_{\perp} \) and \( \beta \), 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 \( \hat{O}_n, \hat{\alpha}_n, \) and \( \hat{B}_{S\beta} \) achieve their oracle property.

Recently, the oracle properties of the Lasso were studied by Kock and Callot (2015) and Basu and Michailidis (2015) for the vector AR model, and by Liang and Schienle (2019) for the VEC model when \( m \to \infty \). 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 \( m \to \infty \).

### 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),

$$
\lambda_{r,k,n} = \frac{\lambda^*_r(\hat{\phi}_k(\hat{\Pi}_{1st}))}{\|\hat{\phi}_k(\hat{\Pi}_{1st})\|_\omega} \quad \text{and} \quad \lambda_{b,j,n} = \frac{m^\omega \lambda^*_b(B_{1st,j})}{\|B_{1st,j}\|_\omega},
$$

(4.1)

where $\lambda^*_r(\cdot)$ and $\lambda^*_b(\cdot)$ are nonincreasing positive sequences, and $\omega$ is some positive finite constant. The extra term $m^\omega$ is used to adjust the effect of the dimensionality of $B_j$ on the adaptive penalty. We introduce the notation $\tau_1$ and $\tau_2$ 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 $\lambda^*_r(\cdot) = o_p(n^{-\tau_1})$, $\lambda^*_b(\cdot) = o_p(n^{-\tau_1})$, and $n^{1-\frac{2}{\alpha}} n^\omega \tilde{L}(n)^{-1} \lambda^*_r, n \to \infty$ and $n^{r_2} \tilde{L}(n)^{-1} \lambda^*_b \to \infty$, for any $j = 1, \ldots, p$ and $k = 1, \ldots, m$, then under Assumption 1 and (2.2), for any $k \in S^c_\phi$ and $j \in S^c_B$, it follows that

$$
\delta_{r,n} + \delta_{b,n} = o_p(n^{-\tau_1}) , \quad n^{1-\frac{2}{\alpha}} \tilde{L}(n)^{-1} \lambda_{r,k,n} \to_p \infty \quad \text{and} \quad n^{r_2} \tilde{L}(n)^{-1} \lambda_{b,j,n} \to_p \infty.
$$
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 $T(k)$ and nonzero $B_{o,j}$ 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:

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

(4.2)

where $\omega \geq \frac{2}{\alpha}$, 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 (7.6) and (7.9) in the appendix. Let $P_n = Q_n^{-1}$, and let $P_n(k)$
be the $k$th column of $P_n$. Denote

$$F_{π,n}(k) = \sum_{t=1}^{n} (\Delta Y_t - \hat{\Pi}_n Y_{t-1} - \sum_{j=1}^{p} \hat{B}_{n,j} \Delta Y_{t-j})' P_n(k) Y'_{t-1}, \quad (4.3)$$

$$F_{b,n}(j) = \sum_{t=1}^{n} (\Delta Y_t - \hat{\Pi}_n Y_{t-1} - \sum_{j=1}^{p} \hat{B}_{n,j} \Delta Y_{t-j}) \Delta Y'_{t-j}, \quad (4.4)$$

for $k = 1, \cdots, m$ and $j = 1, \cdots, p$. Denote $T = Q_n \Pi_o$, and let $T(k)$ be the $k$th row of the matrix $T$. Note that $T_n \equiv Q_n \hat{\Pi}_n$ is an estimator of $T$.

From KKT conditions (7.6) and (7.9) in the appendix, the $k$th row of $T$ is estimated as zero, and the component $B_{o,j}$ in $B_o$ will be estimated as zero only if the following condition holds:

$$\| \tilde{a}_n F_{π,n}(k) \| < \frac{n^{1-\frac{2}{\alpha}} \tilde{L}(n)^{-1} \lambda^*_{r,k,n}}{2\| \phi_k(\hat{\Pi}_{1st}) \|_{\omega}} \quad \text{and} \quad \| \tilde{a}_n F_{b,n}(j) \| < \frac{n^{1-\frac{2}{\alpha}} \tilde{L}(n)^{-1} \lambda^*_{b,j,n}}{2\| \hat{B}_{1st,j} \|_{\omega}}, \quad (4.5)$$

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

$$n^{1-\frac{2}{\alpha}} n^\omega \tilde{L}(n)^{-1} \lambda^*_{r,k,n} \to \infty \quad \text{and} \quad n^{\tau_2} \tilde{L}(n)^{-1} \lambda^*_{b,j,n} \to \infty, \quad (4.6)$$
then (4.5) holds, which implies that a zero $T(k)$ or $B_{o,j}$ is estimated as zero w.p.a.1. By Lemma 3, we have $\phi_k(\hat{\Pi}_{1st}) \rightarrow_p \phi_k(\Pi_o) \neq 0$ and $\hat{B}_{1st,j} \rightarrow_p B_{o,j} \neq 0$ for $k \in S_\phi$ and $j \in S_B$. By Lemma S.1 in the Supplementary Material, if

$$n^{1-\frac{2}{\alpha}} \tilde{L}(n)^{-1} \lambda^*_{r,k,n} \rightarrow 0 \text{ and } n^{r} \tilde{L}(n)^{-1} \lambda^*_{b,j,n} \rightarrow 0, \quad (4.7)$$

then (4.5) cannot hold, which implies that a nonzero $T(k)$ or $B_{o,j}$ is not estimated as zero w.p.a.1. It is not hard to check that (4.2) guarantees (4.6) and that (4.7) holds.

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

$$\hat{c}_{r,k} = 2\|Q_n(k)T_{1,\pi}\| \times \|\hat{T}_{2,\pi}\| \text{ and } \hat{c}_{b,j} = 2\|n^{-\frac{2}{\alpha}} \sum_{t=1}^{n} \Delta Y_{t-j} \Delta Y'_{t-j}\|.$$  

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

\[
\lambda_{r,k,n} = 2n^{-\frac{1}{2}}\|Q_n(k)\hat{T}_{1,\pi}\| \times \|\hat{T}_{2,\pi}\| \times \|\phi_k(\hat{N}_{1st})\|^{-\frac{2}{\alpha}}, \text{ and}
\]

\[
\lambda_{b,j,n} = \begin{cases} 
2n\left(-\frac{1}{2} + \frac{1}{m} - \frac{1}{\alpha^2}\right)n^{-\frac{2}{\alpha}} \sum_{t=1}^{n} \Delta Y_{t-j}\Delta Y'_{t-j} \| \times (\|\hat{B}_{1st,j}\|/m)^{-\frac{2}{\alpha}}, \\
\text{if } \alpha \in (1, 2) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0,
\end{cases}
\]

\[
2n^{-1}n^{-\frac{2}{\alpha}} \sum_{t=1}^{n} \Delta Y_{t-j}\Delta Y'_{t-j} \| \times (\|\hat{B}_{1st,j}\|/m)^{-\frac{2}{\alpha}},
\]

\[
\text{if } \alpha \in (0, 1) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty.
\]

The tail index \( \alpha \) is unknown, in practice. Based on Theorem A.2 in She and Ling (2020), we can estimate \( \frac{2}{\alpha} \) by \( \log(\text{Tr}(\sum_{t=1}^{n} \Delta Y_t\Delta Y'_t)) / \log n \).

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, \( \{Y_t\}_{t=1}^{n} \) are
generated from

\[
\begin{pmatrix}
\Delta Y_{1,t} \\
\Delta Y_{2,t}
\end{pmatrix} = \Pi_0 \begin{pmatrix}
Y_{1,t-1} \\
Y_{2,t-1}
\end{pmatrix} + \sum_{j=1}^{10} B_{o,j} \begin{pmatrix}
\Delta Y_{1,t-1} \\
\Delta Y_{2,t-1}
\end{pmatrix} + \varepsilon_t, \tag{5.1}
\]

with \( \varepsilon_t = |x_t|^{1/\alpha} (\cos \zeta_t, \sin \zeta_t) \), 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 \( Y_0 \) is set to be zero. Furthermore, \( \Pi_0 \) is specified as follows:

\[
\Pi_0 = \begin{bmatrix}
\Pi_{11} & \Pi_{12} \\
\Pi_{21} & \Pi_{22}
\end{bmatrix} = \begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}, \begin{bmatrix}
-1 & -0.5 \\
1 & 0.5
\end{bmatrix}, \text{and} \begin{bmatrix}
-0.5 & 0.1 \\
0.2 & -0.4
\end{bmatrix},
\]

which corresponds to cointegrating ranks of zero, one, and two, respectively.

In addition, \( B_{o,1} \) and \( B_{o,3} \) are taken to be \( \text{diag}(0.4, 0.4) \), and other \( B_{o,j} = 0 \).

We take sample sizes \( n = 100, 400, 800 \), and use 1000 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 $\Pi_o$ 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 $i.i.d.$, the probabilities of selecting the true rank $r_o = 2$...
Table 2: Finite-sample properties of the shrinkage estimates of model (5.1)

<table>
<thead>
<tr>
<th></th>
<th>( \Pi_{11} )</th>
<th></th>
<th>( \Pi_{12} )</th>
<th></th>
<th>( \Pi_{22} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bais</td>
<td>Std</td>
<td>Bais</td>
<td>Std</td>
<td>Bais</td>
</tr>
<tr>
<td>( \alpha = 0.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.01793</td>
<td>0.07266</td>
<td>0.01756</td>
<td>0.09658</td>
<td>0.00584</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00297</td>
<td>0.06269</td>
<td>0.00125</td>
<td>0.07926</td>
<td>0.00158</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00208</td>
<td>0.05298</td>
<td>0.00926</td>
<td>0.08607</td>
<td>0.00118</td>
</tr>
<tr>
<td>( \alpha = 1.3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.01256</td>
<td>0.06628</td>
<td>0.01378</td>
<td>0.09008</td>
<td>0.00332</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00321</td>
<td>0.06513</td>
<td>0.00639</td>
<td>0.06543</td>
<td>0.00177</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00768</td>
<td>0.05298</td>
<td>0.00317</td>
<td>0.0 7626</td>
<td>0.00387</td>
</tr>
</tbody>
</table>

Note: Oracle estimate is the RLSE with \( r_o = 1 \) and the restriction that \( B_{o,2} = 0 \).
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 $B_{o,2} = 0$) 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 vari-
ance 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:

\[
\begin{pmatrix}
\Delta Y_{1,t} \\
\Delta Y_{2,t} \\
\Delta Y_{3,t}
\end{pmatrix} = \Pi_o \begin{pmatrix}
Y_{1,t-1} \\
Y_{2,t-1} \\
Y_{3,t-1}
\end{pmatrix} + \sum_{j=1}^{10} B_{o,j} \begin{pmatrix}
\Delta Y_{1,t-j} \\
\Delta Y_{2,t-j} \\
\Delta Y_{3,t-j}
\end{pmatrix} + \varepsilon_t, \quad (5.2)
\]

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

\[
\Pi_o = \begin{bmatrix}
-0.5 & -0.25 & 0.5 \\
0.1 & 0.05 & -0.1 \\
0.2 & 0.1 & -0.2
\end{bmatrix}, \begin{bmatrix}
-0.5 & -0.2 & 0.7 \\
0.1 & -0.3 & 0.2 \\
0.2 & 0.2 & -0.4
\end{bmatrix}, \text{and} \begin{bmatrix}
-0.4 & 0.0 & 0.0 \\
0.0 & -0.6 & 0.0 \\
0.0 & 0.0 & -0.8
\end{bmatrix},
\]

which corresponds to cointegrating ranks \(r_o = 0, 1, 2, 3\), respectively, and \(B_{o,1}\) and \(B_{o,3}\) are taken to be \(diag(0.4, 0.4, 0.4)\) and other \(B_{o,j} = 0\). The sample size is \(n = 400\) and \(800\), and the number of replications is \(1000\). 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
Table 3: Rank and lagged order selection with adaptive Lasso penalty for model (5.2)

<table>
<thead>
<tr>
<th></th>
<th>( \hat{r}_n = 0 )</th>
<th>( \hat{r}_n = 1 )</th>
<th>( \hat{r}_n = 2 )</th>
<th>( \hat{r}_n = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.667</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>( n = 800 )</td>
<td>0.909</td>
<td>0.990</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>( \alpha = 1.3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.576</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( n = 800 )</td>
<td>0.898</td>
<td>0.994</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Lagged difference selection

<table>
<thead>
<tr>
<th></th>
<th>( \hat{p}_n \in T )</th>
<th>( \hat{p}_n \in C )</th>
<th>( \hat{p}_n \in I )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 0.2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.854</td>
<td>0.017</td>
<td>0.129</td>
</tr>
<tr>
<td>( n = 800 )</td>
<td>0.992</td>
<td>0.008</td>
<td>0.000</td>
</tr>
<tr>
<td>( \alpha = 1.3 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 400 )</td>
<td>0.695</td>
<td>0.017</td>
<td>0.000</td>
</tr>
<tr>
<td>( n = 800 )</td>
<td>0.983</td>
<td>0.017</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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 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 $X_t = (x_{1t}, x_{2t}, x_{3t})'$ denote the original data, and $Y_t = (y_{1t}, y_{2t}, y_{3t})'$ denote the logarithms of the data, that is, $y_{it} = \log(x_{it})$. The data $\{Y_t\}$ 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

Figure 2: The Logarithms of Monthly Prices of Wheat, Corn, and Wheat Flour.
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 \( \{|r_i(t)| : t = 1, \ldots, n\} \) is the decreasing order statistics of \( \{|r_{it}| : t = 1, \ldots, n\} \); 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.

Figure 3: Hill estimator of tail index.
We apply our shrinkage methods to estimate the following VEC model:

\[ \Delta Y_t = \Pi Y_{t-1} + \sum_{k=1}^{10} B_k \Delta Y_{t-k} + \varepsilon_t. \]

(6.1)

The unrestricted LS estimation of \( \Pi \) produced eigenvalues \(-0.0578, -0.0486, \) and \( 0.00003 \), which indicates that \( \Pi \) 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 \( B_k \) are nonzero for \( k = 1, \ldots, 10 \). We apply the LS shrinkage estimation to model (6.1). The results are as follows:

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

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

and the other \( B_k \) are estimated as zero. The eigenvalues of \( \hat{\Pi} \) 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.

7. 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

\[
\sum_{t=1}^{n} \frac{\delta_{\epsilon_{t-1}}}{\alpha_n} \overset{\mathbb{L}}{\to} \sum_{i=1}^{\infty} \delta_{\mathbf{P}_i} = PRM(\mu),
\]
as \( n \to \infty \), where \( PRM(\mu) \) is a Poisson random process with intensity measure \( \mu \) 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

\[
nP\left( \frac{\epsilon_1 \epsilon_2'}{\bar{a}_n} \in \cdot \right) \overset{\mathbb{L}}{\to} \tilde{\mu}(\cdot) \text{ and } \frac{\bar{a}_n}{a_n} \to \infty,
\]
as \( n \to \infty \) when \( E\|\epsilon_1\|^{\alpha} = \infty \), where \( \tilde{\mu} \) is a Radon measure on \((R^{m^2}, B^{m^2})\).

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 \( j = 2, 3, \ldots \) 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

\[ Z_{1,t} = \sum_{i=0}^{\infty} B_i \varepsilon_{t-i} \quad \text{and} \quad Z_{2,t} = [I_d, 0] \sum_{i=1}^{t} \gamma_i, \]

where \( B_i = O(\rho^i) \) with some \( \rho \in (0, 1) \). Denote

\[ R_{1n} = \sum_{t=1}^{n} \varepsilon_t Z'_{1,t-1}, \quad R_{2n} = \sum_{t=1}^{n} \varepsilon_t Z'_{2,t-1}, \]

\[ S_{11n} = \sum_{t=1}^{n} Z_{1,t-1} Z'_{1,t-1}, \quad S_{21n} = \sum_{t=1}^{n} Z_{2,t-1} Z'_{1,t-1} \quad \text{and} \quad S_{22n} = \sum_{t=1}^{n} Z_{2,t-1} Z'_{2,t-1}. \]

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 \( E\|\varepsilon_1\|^{\alpha} = \)
Then

\[ (a) \quad \frac{1}{a_n} R_{1n} \to_d \sum_{i=0}^{\infty} S_{i+2} B_i', \]

\[ (b) \quad \frac{1}{a_n^2} R_{2n} \to_d R_2, \]

\[ (c) \quad \frac{1}{n a_n^2} S_{22n} \to_d \Gamma_{22}, \]

\[ (d) \quad \frac{1}{a_n^2} S_{11n} \to_d \sum_{i=0}^{\infty} B_i S_1 B_i', \]

\[ (e) \quad \frac{1}{a_n^2} S_{21n} \to_d \{ R_2' \sum_{i=0}^{\infty} B_i' + [I_d, 0] \sum_{i=0}^{\infty} \sum_{j=0}^{i} \psi_j S_1 B_i' \}. \]

where \( R_2 = [\int_0^1 P(r) P'(r)dr]' \psi'[I_d, 0]' \) and \( \Gamma_{22} = [I_d, 0] \psi'[\int_0^1 P(r) P'(r)dr] \psi'[I_d, 0]' \),

\[ S_1 = \sum_{i=1}^{\infty} P_i^{(1)} P_i^{(1)'} \] with \( P_i^{(1)} = P_i \), \( S_j = \sum_{i=1}^{\infty} P_i^{(j)} \) for all \( j > 1 \) and \( P(r) \) is a stable process.

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

**Lemma 3.** Suppose that (2.2) and Assumptions 1-2 hold, and \( E\|\varepsilon_1\|^{\alpha} = \infty \) and \( S_1 \) is positive definite almost surely. Then

\[ (a) \quad [(\widehat{\Pi}_{1st}, \widehat{B}_{1st}) - (\Pi_o, B_o)] Q_B^{-1} D_{n,B}^{-1} \to_d (B_{m,1}, B_{m,2}), \] where \( B_{m,2} = \ldots \]
\[ R_2 \Gamma_{22}^{-1} \]

\[ B_{m,1} = \begin{cases} 
R_1^* \Gamma_{11}^{-1} & \text{if } \alpha \in (1, 2) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0, \\
-R_2 \Gamma_{22}^{-1} \Gamma_{21}^* \Gamma_{11}^{-1} & \text{if } \alpha \in (0, 1) \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty, 
\end{cases} \]

\[ R_1^* = \sum_{i=0}^{\infty} S_{i+2} B_i', \quad \Gamma_{11}^* = \sum_{l=0}^{\infty} B_l S_1 B_l' \quad \text{and} \quad \Gamma_{21}^* = R_2' \sum_{i=0}^{\infty} B_i' + I_d, 0] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_j S_1 B_i', \]

(b). for \( k = 1, \ldots, m \), the eigenvalues of \( \tilde{\Pi}_{1st} \) satisfy \( \phi_k(\tilde{\Pi}_{1st}) \to_p \phi_k(\Pi_o), \)

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

\[ n(\phi_1(\tilde{\Pi}_{1st}), \ldots, \phi_{m-r_o}(\tilde{\Pi}_{1st})) \to_d (\tilde{\phi}_{o,1}, \ldots, \tilde{\phi}_{o,m-r_o}), \]

where the \( \tilde{\phi}_{o,j} (j = 1, \ldots, m - r_o) \) are solutions of the equation \( |\mu I_{m-r_o} - \alpha_{o,\perp}' R_2 \Gamma_{22}^{-1}| = 0. \)

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 \( P_n \) as \( P_n = [P_{\alpha,n}; P_{\alpha,\perp,n}] \), where \( P_{\alpha,n} \) is the first \( r_o \) columns of \( P_n \) (\( P_{\alpha,\perp,n} \) is defined accordingly). Then

\[ Q_{\alpha,\perp,n} P_{\alpha,\perp,n} = I_{m-r_o}; \quad Q_{\alpha,n} P_{\alpha,\perp,n} = 0_{r_o \times (m-r_o)} \quad \text{and} \quad Q_{\alpha,\perp,n} \tilde{\Pi}_{1st} = \Lambda_{\alpha,\perp,n} Q_{\alpha,\perp,n}, \]

where \( \Lambda_{\alpha,\perp,n} \) is a diagonal matrix with the ordered last (smallest) \( m - r_o \)
eigenvalues of $\hat{\Pi}_{1st}$. Define a useful estimator of $\Pi_o$ as

$$\Pi_{n,f} = \hat{\Pi}_{1st} - P_{\alpha \perp,n} \Lambda_{\alpha \perp,n} Q_{\alpha \perp,n}. $$

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

$$Q_{\alpha,n} \Pi_{n,f} = Q_{\alpha,n} \hat{\Pi}_{1st} - Q_{\alpha,n} P_{\alpha \perp,n} \Lambda_{\alpha \perp,n} Q_{\alpha \perp,n} = \Lambda_{\alpha,n} Q_{\alpha,n}, \quad (7.1)$$

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

$$Q_{\alpha \perp,n} \Pi_{n,f} = Q_{\alpha \perp,n} \hat{\Pi}_{1st} - Q_{\alpha \perp,n} P_{\alpha \perp,n} \Lambda_{\alpha \perp,n} Q_{\alpha \perp,n} = 0_{(m-r_o)\times m}. \quad (7.2)$$

From Lemma 2 (b), (7.1) and (7.2), we can deduce that $Q_{\alpha,n} \Pi_{n,f}$ is a $r_o \times m$ matrix which is nonzero w.p.a.1 and $Q_{\alpha \perp,n} \Pi_{n,f}$ is always a $(m - r_o) \times m$ zero matrix for all $n$. By Lemma 3 (a) and (c), it follows that

$$(\Pi_{n,f} - \Pi_o) Q^{-1} D_n^{-1} = O_p(1), \quad (7.3)$$
where $D_n = n^{-1}I_m$ when $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$, and $\text{diag}\{\tilde{a}_n/a_n^2I_{r_o}, n^{-1}I_{m-r_o}\}$ when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$.

**Proof of Theorem 3.** Let $T = Q_n\Pi$. We can rewrite the LS shrinkage estimation problem in (2.3) as

\[
(\hat{T}_n, \hat{B}_n) = \arg \min_{T, B_1, \ldots, B_p \in R^{m \times m}} \sum_{t=1}^n \|\Delta Y_t - P_nTY_{t-1} - \sum_{j=1}^p B_j \Delta Y_{t-j}\|^2 + n \sum_{j=1}^p \lambda_{b,j,n} \|B_j\| + n \sum_{k=1}^m \lambda_{r,k,n} \|\Phi_{n,k}(P_nT)\|. \tag{7.4}
\]

By definition of (7.4), $\hat{\Pi}_n = P_n\hat{T}_n$ and $\hat{T}_n = Q_n\hat{\Pi}_n$ for all $n$, and

\[
\hat{T}_n = \begin{pmatrix} Q_{\alpha,n}\hat{\Pi}_n \\ Q_{\alpha,n}\hat{\Pi}_{1st} \end{pmatrix} = \begin{pmatrix} Q_{\alpha,n}\hat{\Pi}_{1st} \\ Q_{\alpha,n}\hat{\Pi}_{1st} \end{pmatrix} + o_p(1),
\]

and thus the first result follows if we can show that the last $m - r_o$ rows of $\hat{T}_n$ are estimated as zeros w.p.a.1. Note that $\Phi_{n,k}(P_nT)$ is the $k$-th row of $Q_n(P_nT)$, i.e., $\Phi_{n,k}(P_nT) = T(k)$. The problem in (7.4) can be rewritten
ESTIMATION FOR HEAVY-TAILED VEC MODELS

\begin{equation}
(\hat{T}_n, \hat{B}_n) = \arg\min_{T, B_1, \ldots, B_p \in \mathbb{R}^{m \times m}} \sum_{t=1}^{n} \| \Delta Y_t - P_n T Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j} \|^2
+n \sum_{j=1}^{p} \lambda_{b,j,n} \| B_j \| + n \sum_{k=1}^{m} \lambda_{r,k,n} \| T(k) \|.
\end{equation}

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

\begin{equation}
\begin{cases}
F_{\pi,n}(k) = \frac{n \lambda_{r,k,n}}{2} \frac{\hat{T}_n(k)}{\| T_n(k) \|} \quad \text{if} \quad \hat{T}_n(k) \neq 0, \\
\| n^{-1} F_{\pi,n}(k) \| < \frac{\lambda_{r,k,n}}{2} \quad \text{if} \quad \hat{T}_n(k) = 0,
\end{cases}
\end{equation}

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

\begin{equation}
\| a_n^{-2} F_{\pi,n}(k) \| = \frac{n a_n^{-2} \lambda_{r,k,n}}{2}.
\end{equation}
By Lemma 2 (b), (c) and (e), and Theorem 1, it follows that

\[ \frac{1}{a_n^2} F_{x,n}(k_0) = \frac{1}{a_n^2} \sum_{t=1}^{n} [\varepsilon_t - (\hat{\theta}_n - \theta_o)Q_B^{-1}Z_{t-1}]'P_n(k_0)Y'_{t-1} \]

\[ = \frac{1}{a_n^2} \sum_{t=1}^{n} [\varepsilon_t - (\hat{\theta}_n - \theta_o)Q_B^{-1}Z_{t-1}]'P_n(k_0)Y'_{t-1} \]

\[ = \frac{P_n(k_0)' \sum_{t=1}^{n} \varepsilon_t Y'_{t-1}}{a_n^2} - \frac{P_n(k_0)'(\hat{\theta}_n - \theta_o)Q_B^{-1} \sum_{t=1}^{n} Z_{t-1}Y'_{t-1}}{a_n^2} = O_p(1). \] (7.8)

By assumptions of tuning parameters, \( na_n^{-2} \lambda_{r,k,n} \rightarrow p \infty \). Furthermore, by (7.7) and (7.8), we must have

\[ P(\mathbf{Q}_n(k_0) \hat{\mathbf{P}}_n = 0) = P(\mathbf{\hat{T}}_n(k_0) = 0) \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty, \]

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

\[ P(\mathbf{Q}_{\alpha,k,n} \hat{\mathbf{P}}_n = 0) \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty. \]

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

\[ \begin{cases} F_{b,n}(j) = \frac{n \lambda_{b,n} \hat{\mathbf{B}}_{n,j}}{2\| \mathbf{B}_{n,j} \|} & \text{if} \ \hat{\mathbf{B}}_{n,j} \neq 0, \\ \|n^{-1}F_{b,n}(j)\| < \frac{\lambda_{b,n}}{2} & \text{if} \ \hat{\mathbf{B}}_{n,j} = 0, \end{cases} \] (7.9)

for any \( j = 1, \ldots, p \), where \( F_{b,n}(j) \) is defined in (4.4).
0_{m \times m}$ for some $j \in S_B^c$, we get the following equation from the optimality conditions
\[
\|F_{b,n}(j)\| = \frac{n\lambda_{b,j,n}}{2}.
\] (7.10)
Let $\tilde{\delta}_n = \tilde{a}_n^{-1}$ when $\alpha \in (1, 2)$ or $\alpha = 1$ and $\tilde{L}(n) \to 0$, and $\tilde{\delta}_n = na_n^{-2}$ when $\alpha \in (0, 1)$ or $\alpha = 1$ and $\tilde{L}(n) \to \infty$. Then, (7.10) is equivalent to
\[
\left\| \tilde{\delta}_n F_{b,n}(j) \right\| = \frac{n\tilde{\delta}_n \lambda_{b,j,n}}{2}.
\] (7.11)

By Lemma 2 (a), (d) and (e), and Theorem 1, we have
\[
\tilde{\delta}_n F_{b,n}(j) = \tilde{\delta}_n \sum_{t=1}^{n} [\varepsilon_t - (\hat{\theta}_n - \theta_o) Q_{B}^{-1} Z_{t-1}] \Delta Y'_{t-j}
= \tilde{\delta}_n \sum_{t=1}^{n} \varepsilon_t \Delta Y'_{t-j} - \tilde{\delta}_n (\hat{\theta}_n - \theta_o) Q_{B}^{-1} \sum_{t=1}^{n} Z_{t-1} \Delta Y'_{t-j} = O_p(1). \tag{7.12}
\]
However, under the assumptions of tuning parameters, $n\tilde{\delta}_n \lambda_{b,j,n} \to_p \infty$, which together with results in (7.11) and (7.12) implies that
\[
P(\hat{B}_{n,j} = 0_{m \times m}) \to 1 \text{ as } n \to \infty,
\]
for any $j \in S_B^c$, which finishes the proof. \qed

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

By definition of $\mathbf{\beta}_o$ and $\mathbf{\beta}_{o,\bot}$,

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

which implies that

$$
\mathbf{\beta}'_{1,\bot} = -\mathbf{\beta}'_{2,\bot} \mathbf{O}'_{r_o} \quad \text{and} \quad \mathbf{\beta}_{2,\bot} = (\mathbf{I}_{m-r_o} + \mathbf{O}'_{r_o} \mathbf{O}_{r_o})^{-\frac{1}{2}}. \quad (7.13)
$$

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

$$
\delta^*_n (\mathbf{\hat{B}}_n - \mathbf{B}_o) = O_p(1), \quad (7.14)
$$

$$
n(\mathbf{\hat{\beta}}_n - \mathbf{\beta}_o) = O_p(1), \quad (7.15)
$$

$$
\delta^*_n (\mathbf{\hat{\alpha}}_n - \mathbf{\alpha}_o) = O_p(1). \quad (7.16)
$$

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

$$
[\delta^*_n (\mathbf{\hat{\Pi}}_n - \mathbf{\Pi}_o) \mathbf{\alpha}_o (\mathbf{\beta}'_o \mathbf{\alpha}_o)^{-1}, \delta^*_n (\mathbf{\hat{B}}_n - \mathbf{B}_o), n(\mathbf{\hat{\Pi}}_n - \mathbf{\Pi}_o) \mathbf{\beta}_{o,\bot} (\mathbf{\alpha}'_{0,\bot} \mathbf{\beta}_{0,\bot})^{-1}] = O_p(1),
$$
which implies that (7.14) holds, and

$$\delta_n^*[\nabla_n - \alpha_0)\beta'_n + \alpha_0(\beta_n - \beta_0)'\alpha_0(\beta'_o\alpha_o)^{-1} = \delta_n^*(\Pi_n - \Pi_0)\alpha_o(\beta'_o\alpha_o)^{-1} = O_p(1), \tag{7.17}$$

$$n\alpha_n(\beta'_n - \beta_o)'(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = n(\Pi_n - \Pi_0)\alpha_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = O_p(1). \tag{7.18}$$

By the definitions of $\hat{\beta}_n$ and $\beta_{o,\perp}$ and (7.18), we can deduce that

$$\beta'_o\alpha_n[n(\hat{\Pi}_n - O_{r_o})]'(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = O_p(1). \tag{7.19}$$

By (7.13) and (7.19), we have

$$n(\hat{\Pi}_n - O_{r_o}) = [\beta'_o\alpha_o + o_p(1)]^{-1}O_p(1)(\alpha'_{o,\perp}\beta_{o,\perp})(I_m - r_o + O'_{r_o}r_o)^2 = O_p(1). \tag{7.20}$$

Again, by the definition of $\hat{\beta}_n$, (7.20) means that $n(\hat{\beta}_n - \beta_o) = O_p(1)$, that is, (7.15) holds. By (7.15) and (7.17), we know that (7.16) holds. When $\alpha \in (0, 1)$ or $\alpha = 1$ and $\bar{L}(n) \to \infty$ 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 $\bar{L}(n) \to 0$, we can show
(7.14)-(7.16) hold.

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

$$
V_n(\theta_S) = \sum_{t=1}^{n} \| \Delta Y_t - \alpha \beta' Y_{t-1} + \sum_{j \in S_B} B_j \Delta Y_{t-j} \|^2 
+ n \sum_{k \in S_B} \lambda_{r,k,n} \| \Phi_{n,k}(\alpha \beta') \|
+ n \sum_{j \in S_B} \lambda_{b,j,n} \| B_j \|.
$$

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

Then

$$
\left[ (\hat{\Pi}_n - \Pi_o), (\hat{B}_{S_B} - B_{o,S_B}) \right] Q_{S}^{-1} D_{n,S}^{-1}
= [\delta_n^* (\hat{\beta}_n - \beta_o)' \alpha_o (\beta_o' \alpha_o)^{-1} + \delta_n^* (\hat{\alpha}_n - \alpha_o), \delta_n^* (\hat{B}_{S_B} - B_{o,S_B}),
\delta_n^* (\hat{\beta}_n - \beta_o)' \beta_{o,\perp} (\alpha_{o,\perp} \beta_{o,\perp})^{-1}]
= \left[ n^{-1} \delta_n^* [0_{r_o}, U_{2,n}^*] \alpha_o (\beta_o' \alpha_o)^{-1} + U_{1,n}^* [0_{r_o}, U_{2,n}^*] \beta_{o,\perp} (\alpha_{o,\perp} \beta_{o,\perp})^{-1} \right].
$$

Denote $U = (U_1, U_2, U_3) \in R^{m \times r_o} \times R^{r_o \times (m-r_o)} \times R^{m \times md_S}$ and

$$
\Pi_n(U) = [n^{-1} \delta_n^* [0_{r_o}, U_2] \alpha_o (\beta_o' \alpha_o)^{-1} + U_1 \alpha_o [0_{r_o}, U_2] \beta_{o,\perp} (\alpha_{o,\perp} \beta_{o,\perp})^{-1}].
$$
Then, \( U_n^* = (U_{1,n}^*, U_{2,n}^*, U_{3,n}^*) \) minimizes the following criterion function

\[
V_n(U) = \sum_{t=1}^{n} (\| \varepsilon_t - \Pi_n(U)D_{n,S}Z_{S,t-1} \|^2 - \| \varepsilon_t \|^2)
\]

\[
+ n \sum_{k \in S_o} \lambda_{r,k,n} [\| \Phi_n,k[\Pi_n(U)D_{n,S}Q_SL_1 + \Pi_o]\| - \| \Phi_n,k(\Pi_o)\|]
\]

\[
+ n \sum_{j \in S_B} \lambda_{b,j,n} [\| \Pi_n(U)D_{n,S}Q_SL_{j+1} + B_{o,j}\| - \| B_{o,j}\|],
\]

where \( L_j = \text{diag}(A_{j,1}, \ldots, A_{j,d_{SB}+1}) \) with \( A_{j,j} = I_m \) and \( A_{i,j} = 0 \) for \( i \neq j \) and \( j = 1, \ldots, d_{SB}+1 \).

For any compact set \( K \in R^{m \times r_o} \times R^{r_o \times (m-r_o)} \times R^{m \times md_{SB}} \) and any \( U \in K \), there is \( \Pi_n(U)D_{n,S}Q_S = O_p(\delta_n^{-1}) \). Then we can deduce that

\[
n \sum_{k \in S_o} \lambda_{r,k,n} [\| \Phi_n,k[\Pi_n(U)D_{n,S}Q_SL_1 + \Pi_o]\| - \| \Phi_n,k(\Pi_o)\|]
\]

\[
\leq n \lambda_{r,k,n} [\| \Phi_n,k(\Pi_n(U)D_{n,S}Q_S)\| = O_p(n \delta_n^{-1} \lambda_{r,k,n}) = o_p(1), \quad (7.21)
\]

and \( n \sum_{j \in S_B} \lambda_{b,j,n} [\| \Pi_n(U)D_{n,S}Q_SL_{j+1} + B_{o,j}\| - \| B_{o,j}\|] = o_p(1), \quad (7.22)
\]

uniformly over \( U \in K \).

Denote \( \vartheta = \{[\text{vec}O_{r_o}'], [\text{vec}(\alpha_o, B_{o,S_B})']\}' \). From (7.21) and (7.22), we deduce that \( \hat{\alpha}_n, \hat{O}_n \) and \( \hat{B}_{SB} \) minimize the following criterion function.
ESTIMATION FOR HEAVY-TAILED VEC MODELS

w.p.a.1

\[ L(\theta) = \sum_{t=1}^{n} \| \varepsilon_t(\theta) \|^2 \quad \text{and} \quad \varepsilon_t(\theta) = \Delta Y_t - \alpha_\theta [I_{r_o}, O_{r_o}] Y_{t-1} - B_{s_o} S_t \Delta X_{S,t-1}. \]

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

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)
REFERENCES

References


REFERENCES


*J. Econometrics*, 186, 325-344.

Kock, A.B. and Callot, L. (2015), Oracle inequalities for high dimensional vector autoregres-


REFERENCES


Zhang, R., and Ling, S. (2015), Asymptotic Inference for AR Models With Heavy-Tailed G-
REFERENCES


School of Mathematics and Statistics
Beijing Institute of Technology, Beijing, China
E-mail: 7520200236@bit.edu.cn

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

Department of Statistics
Shanxi University of Finance and Economics
E-mail: mizc@sxufe.edu.cn