**Statistica Sinica Preprint No: SS-2020-0177**

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Automated Estimation of Heavy-Tailed Vector Error Correction Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>SS-2020-0177</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td>10.5705/ss.202020.0177</td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
<td>Feifei Guo, Shiqing Ling and Zichuan Mi</td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
<td>Shiqing Ling</td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
<td><a href="mailto:maling@ust.hk">maling@ust.hk</a></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
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 via adaptive shrinkage techniques to determine the co-integrating rank and estimate parameters simultaneously in the VEC model with unknown order $p$ when its noises are i.i.d. heavy-tailed random vectors with tail index $\alpha \in (0, 2)$. It is showed that the estimated co-integrating rank and order $p$ equal to the true rank and the true order $p_0$, respectively, with probability trending to 1 as the sample size $n \to \infty$, while other estimated parameters achieve the oracle property, that is, they have the same rate of convergence and the same limiting distribution as those of estimated parameters when the co-integrating rank and the true order $p_0$ are known. This paper also proposes a data-driven procedure of selecting the tuning parameters. Simulation studies are carried to evaluate the performance of this procedure in finite samples. Our techniques are applied to explore the long-run and short-run behavior of prices of wheat, corn and wheat in USA.

Key words: Cointegration, Heavy-tailed, Reduced rank LSE, LASSO.
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. The early research can be found in Phillips and Durlauf (1986), Ahn and Reinsel (1990), Reinsel and Ahn (1992), Stock and Watson (1993), Johansen (1988, 1995), among many others. Recently, Wang and Phillips (2012) proposed a test for nonlinear non-stationary models. Kristensen and Rahbek (2013) showed the testing and inference in nonlinear cointegrating vector error correction models. Cavaliere, Nielsen, and Rahbek (2015) considered a bootstrap test on the cointegration rank relation in vector autoregressive 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 via the shrinkage technique–group lasso approach. This approach does not need to estimate the long-variance of linear processes via a non-parametric approach as in Phillips and Solo (1992) or to pre-determine the order of VEC models as in Johansen (1995).

The research on cointegration systems mainly focuses on the time se-
ries with a finite second or even higher moment. The heavy-tailed time series do not have a finite second moment, and they have been well observed in financial market, engineering, network system and other areas, see Resnick (1997). Davis and Resnick (1985, 1986) showed that the limiting distribution of the least square estimator (LSE) of the parameters in a heavy-tailed autoregressive (AR) process is a function of two stable random variables with the rate of convergence much faster than $\sqrt{n}$. Mikosch et al. (1995) studied the whittle estimators for the heavy-tailed ARMA model and gave its asymptotic properties. Zhang and Ling (2015) established the asymptotic properties of the AR model with heavy-tailed G-GARCH noises. 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, this theory cannot be applied for testing the cointegrating rank of the heavy-tailed VEC model. Thus, except for a very special case in Caner (1998), it still remains an open problem to determine the co-integrating rank in the VEC model when its noise is a heavy-tailed random vector.

This paper is to develop an automated approach via adaptive shrinkage
techniques to determine the co-integrating rank and estimate parameters simultaneously in the VEC model with unknown order $p$ when its noises are i.i.d. heavy-tailed random vectors with tail index $\alpha \in (0, 2)$. It is showed that the estimated co-integrating rank and order $p$ equal to the true rank and the true order $p_o$, respectively, with probability trending to 1 as the sample size $n \to \infty$, while other estimated parameters achieve the oracle property, that is, they have the same rate of convergence and the same limiting distribution as those of estimated parameters when the co-integrating rank and the true order $p_o$ are known. This paper also proposes a data-driven procedure of selecting the tuning parameters.

The Lasso approach was developed by Tibshirani (1996) for selecting variables and estimating parameters. It has been extensively studied and many variants have been proposed, e.g., 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 adaptive Lasso of high-dimensional regression. Chen and Chan (2011) considered adaptive Lasso for ARMA model selection and obtained asymptotic normality for the estimated pa-
rameters. Song and Bickel (2011) studied the Lasso estimator for a large vector AR model. Kock (2016) investigated adaptive Lasso the for AR models. Chan, Ling and Yau (2020) studied Lasso-based variable selection of stationary and unit-root ARMA models. The results in current paper may provide a new insight to the Lasso approach for both stationary and non-stationary heavy-tailed time series.

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

2. Model and LS Shrinkage Estimation

We consider the following VEC representation of a cointegrated system

\[
\Delta Y_t = \Pi_o Y_{t-1} + \sum_{j=1}^{p} B_{o,j} \Delta Y_{t-j} + \varepsilon_t, \tag{2.1}
\]

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, \(p > \text{true order } p_o\) and \(B_{o,j} = 0\) if \(j > p_o\).
estimation for heavy-tailed vec models

\( p_0, \) and \( \{ \varepsilon_t \} \) is a sequence of independent and identically distributed (i.i.d.) \( m \)-dimensional random vector. Model (2.1) is a partially non-stationary 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. \( \beta_o \) 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 is to measure the number of cointegrating relations in the system. The set of nonzero matrices \( B_{o,j} \) \( (j = 1, \ldots, p) \) is characterizing 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 \( (R^m, B^m) \), \( a_n \) is an increasing sequence divergence to \( \infty \) and \( \xrightarrow{v} \) means the vague convergence in Durrett (2019, pp.121). (2.2) is called regular variation function and it is equivalent to that there exists a probability measure \( \mu^* \) on the unit sphere \( S^m \) 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 \to \infty \text{, where } \alpha > 0 \text{ called the tail index and } \| \cdot \| \text{ denotes the Euclidean norm, see Resnick(1986). When } \alpha \in (0, 2), \varepsilon_1 \text{ 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 \text{ 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's}, \forall u \in R^m,

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

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

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

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

\text{as } n \to \infty, \text{ where } c_0 \text{ is some constant, see Resnick(1986). We choose } a_n \text{ as follows}

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

\text{Then } a_n = n^{1/\alpha} L(n), \text{ where } L(n) \text{ is a slowly varying function, see Bingham,
Goldie, and Teugels (1989). When \( \varepsilon_t \) is defined as in (5.1) in Section 5, its density function is

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

see Figure 1 for the plot of \( f(x, y) \) when \( \alpha = 0.8 \) and 1.6.

We will determine the cointegrating rank \( r_o \) and lag order \( p_o \) in conjunction with oracle-like efficient estimation of the cointegrating matrix and transient dynamics simultaneously. 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 the \( m \times (m - r_o) \) full rank matrices and
are orthogonal complements of $\alpha_o$ and $\beta_o$. Denote $Q = [\beta_o, \alpha_{o,\perp}]'$. Following the same arguments in Liao and Phillips (2015), we can show that $Q^{-1} = \left[ \alpha_o (\beta_o' \alpha_o)^{-1}, \beta_{o,\perp} (\alpha_{o,\perp}' \beta_{o,\perp})^{-1} \right]$, 

\[
Q \Pi_o = \begin{bmatrix} \beta_o' \alpha_o' \\ 0 \end{bmatrix} \quad \text{and} \quad Q \Pi_o Q^{-1} = \begin{bmatrix} \beta_o' \alpha_o & 0 \\ 0 & 0 \end{bmatrix}.
\]

Thus, the cointegrating rank $r_o$ is the number of the nonzero row vector count of $Q \Pi_o$. It follows that consistent selection of the cointegration rank $r_o$ is equivalent to determine the number of the 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 $S^c_\phi = \{k : \Phi_k(\Pi_o) = 0\}$ 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 $S^c_\phi = \{r_o + 1, \ldots, m\}$. Let $B_o = [B_{o,1}, \ldots, B_{o,p}]$, and denote the index set of the zero components in $B_o$ as $S^c_B$ such that $\|B_{o,j}\| = 0$ for all $j \in S^c_B$ and $\|B_{o,j}\| \neq 0$ otherwise. The ordinary least squares (OLS) estimate of
(\(\Pi_0, B_0\)), 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 \(\alpha'_o\). The true parameters are estimated by the penalized LS estimation

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

where \(\hat{B}_n = [\hat{B}_{n,1}, \ldots, \hat{B}_{n,p}]\), and \(\lambda_{b,j,n} (j = 1, \ldots, p)\) 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 \(\Pi\) is different 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 \( Q_n \hat{\Pi}_n \) and an implied estimate of the transient dynamic structure (including the order \( p \)) based on the fitted value \( \hat{B}_n \).

The determinant of a square matrix \( A \) is denoted by \( |A| \) and the \( M \times M \) identity matrix is denoted by \( I_M \). We first give the following assumption:

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

\[
C(z) = \Pi_0 z + \sum_{j=0}^{p} B_{o,j} (1 - z) z^j \quad \text{with} \quad B_{o,0} = -I_m,
\]

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

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

\[
Y_t = C_B \sum_{s=1}^{t} \varepsilon_s + \Xi(L) \varepsilon_t + C_B Y_0, \quad \text{(2.4)}
\]

where \( C_B = \beta'_{o,\perp} (\alpha'_{o,\perp} (I_m - \sum_{j=1}^{p} B_{o,j}) \beta'_{o,\perp})^{-1} \alpha'_{o,\perp} \) and \( \Xi(L) \varepsilon_t \) is a stationary process. From the partial sum in (2.4), one can deduce that \( \beta'_{o} Y_t \) and \( \Delta Y_{t-j} \) are stationary. Denote \( \Delta X_{t-1} = [\Delta Y_{t-1}', \ldots, \Delta Y_{t-p}']' \). Model
in (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:

\[
Q_B = \begin{pmatrix} \beta_o & 0 \\ 0 & I_{mp} \\ \alpha_{o,\perp}' & 0 \end{pmatrix} \quad \text{and} \quad Q_B^{-1} = \begin{pmatrix} \alpha_o (\beta_o' \alpha_o)^{-1} & 0 & \beta_{o,\perp} (\alpha_{o,\perp}' \beta_{o,\perp})^{-1} \\ 0 & I_{mp} & 0 \end{pmatrix}.
\]

Then

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

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 \( (\widehat{\Pi}_n, \widehat{B}_n) \), we need one more assumption. It is to 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 1-2 are satisfied. If \( \delta_{r,n} \equiv \)
max_{k \in S_b} \lambda_{r,k,n} = o_p(1) \text{ and } \delta_{b,n} = \max_{j \in S_B} \lambda_{b,j,n} = o_p(1), \text{ then the LS shrinkage estimator } (\hat{\Pi}_n, \hat{B}_n) \text{ is consistent, i.e. } (\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, the consistency of 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 the probability limits of the estimates of those nonzero eigenvalues are not zero or at least that their rates of convergence are slower than those of 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 variation 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 \( a_n^2/\tilde{a}_n[(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o)] = O_p(1 + n \tilde{a}_n^{-1} \delta_n). \)

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

(c). if \(0 \leq r_o < m\), \(\alpha \in (0, 1)\) or \(\alpha = 1\) and \(\tilde{L}(n) \to \infty\), then
\[
n[(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o)]Q_B^{-1} = O_p(1).
\]

The term \(\delta_{r,n}\) and \(\delta_{b,n}\) represent the shrinkage bias that the penalty function introduces to the LS shrinkage estimator. Denote
\[
D_{n,B} = \begin{cases} 
\frac{1}{n}I_{m(1+p)} & \text{if } \alpha \in (0, 1), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to \infty, \\
\text{diag}\{\frac{\tilde{a}_n}{a_n^2}I_{r_o + mp}, \frac{1}{n}I_d\} & \text{if } \alpha \in (1, 2), \text{ or } \alpha = 1 \text{ and } \tilde{L}(n) \to 0.
\end{cases}
\]

If the rate of convergence of \(\lambda_{r,k,n}(k \in S_\phi)\) 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
\[
(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o) = O_p(n^{-1}) \text{ when } r_o = 0, \text{ and } O_p(n^{-\frac{1}{\alpha}}\tilde{L}(n)^{-1}) \text{ when } r_o = m, \text{ and } [(\hat{\Pi}_n, \hat{B}_n) - (\Pi_o, B_o)]Q_B^{-1}D_{n,B}^{-1} = O_p(1) \text{ otherwise, that is,}
\]
the LS shrinkage estimators have the same rates of convergence as the OLS estimators \((\hat{\Pi}_{1st}, \hat{B}_{1st})\). In the next section, we give the condition on 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 1 (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 $Q'_n = [Q'_{\alpha,n}, Q'_{\alpha_{\perp},n}]$, where $Q_{\alpha,n}$ and $Q_{\alpha_{\perp},n}$ are the first $r_o$ rows and the last $m - r_o$ rows of $Q_n$ respectively. Under Lemma 3 and Theorem 1,

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

and similarly

$$Q_{\alpha_{\perp},n} \hat{\Pi}_n = Q_{\alpha_{\perp},n} \hat{\Pi}_{1st} + o_p(1) = \Lambda_{\alpha_{\perp},n} Q_{\alpha_{\perp},n} + o_p(1) = o_p(1), \quad (3.2)$$

where $\Lambda_{\alpha,n} = diag[\phi_1(\hat{\Pi}_{1st}), ..., \phi_{r_o}(\hat{\Pi}_{1st})]$, $\Lambda_{\alpha_{\perp},n} = diag[\phi_{r_o+1}(\hat{\Pi}_{1st}), ..., \phi_m(\hat{\Pi}_{1st})]$ and $\phi_k(\hat{\Pi}_{1st})$ denotes the $k$-th largest eigenvalues of $\hat{\Pi}_{1st}$ for $k = 1, \ldots, m$. (3.1) implies that the first $r_o$ rows of $Q_n \hat{\Pi}_n$ are nonzero w.p.a.1, while (3.2) implies that the last $m - r_o$ rows of $Q_n \hat{\Pi}_n$ 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

\[ n^{1-\frac{2}{\alpha}}(\delta_{b,n} + \delta_{r,n}) = O_p(1), \]

\[ n^{1-\frac{2}{\alpha}}\tilde{L}(n)^{-1}\lambda_{r,k,n} \rightarrow_p \infty \text{ for } k \in S^c_B \text{ and } n^{2-\alpha}\tilde{L}(n)^{-1}\lambda_{b,j,n} \rightarrow_p \infty \text{ for } j \in S^c_B, \]

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

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

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 the consistent selection of the cointegration rank \( r_o \) and the lag order \( p_o \). That is the following result.

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

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

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

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

and

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

where \( I_{S_B} = I_{m d_{S_B}} \) 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

\[
Z_{S,t-1} = Q_S \begin{bmatrix}
Y_{t-1} \\
\Delta X_{S,t-1}
\end{bmatrix} = \begin{bmatrix}
Z_{1,S,t-1} \\
Z_{2,t-1}
\end{bmatrix}, \quad (3.5)
\]
with $Z'_{1S,t-1} = [Y'_{t-1}\beta_o \Delta X'_{S,t-1}]$ and $Z_{2,t-1} = \alpha'_{o,\perp}Y_{t-1}$. 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 $A_i = O(\rho^i)$, $\gamma_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$, $d = m - r_o$ 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 $\tilde{\beta}_1 = \beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = [\beta'_{1,1}, \beta'_{1,2}]$, and $\tilde{\beta} = \alpha_o (\beta'_o \alpha_o)^{-1} = [\beta'_1, \beta'_2]'$, where $\beta_{1,2}$ and $\beta_2$ are the last $m - r_o$ rows of $\tilde{\beta}_1$ and $\tilde{\beta}$, respectively.

Since the cointegrating rank is consistently selected as $r_o$, the LS shrinkage estimator $\hat{\Pi}_n$ can be decomposed as $\hat{\alpha}_n \hat{\beta}'_n$, where $\hat{\alpha}_n$ is the first $r_o$ columns of $\hat{\Pi}_n$ and $\hat{\beta}_n = [I_{r_o}, \hat{O}_n]$. $\rightarrow_d$ denotes convergence in distribution. We have the following result.

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

\( (a) \quad n(\hat{O}_n - O_{r_o}) \to_d (\alpha_o' \alpha_o)^{-1} \alpha_o' R_2 \Gamma_{22}^{-1} \beta_{2, 2}^{-1}, \)

\( (b) \quad n^{1/\alpha} \tilde{L}(n)(\hat{\alpha}_n - \alpha_o, \hat{B}_{S_B} - B_{o,S_B}) \to_d R_1 \Gamma_{11}^{-1}, \)

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

\( (c) \quad n(\hat{\alpha}_n - \alpha_o, \hat{B}_{S_B} - B_{o,S_B}) \to_d \alpha_o(\alpha_o' \alpha_o)^{-1} \alpha_o' F_1 + [I_m - \alpha_o(\alpha_o' \alpha_o)^{-1} \alpha_o'] F_2, \)

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

where \( R_1 = \sum_{i=0}^{\infty} S_{i+2} A_i', \quad \Gamma_{21} = R_2' \sum_{i=0}^{\infty} A_i', \quad [I_d, 0] \sum_{i=0}^{\infty} \psi S_1 A_i', \) with

\( \psi = \sum_{i=0}^{\infty} \psi_i, \quad \Gamma_{11} = \sum_{i=0}^{\infty} A_i S_1 A_i', \quad F_1 = -R_2 \Gamma_{22}^{-1} \Gamma_{21} \Gamma_{11}^{-1} R_2 \Gamma_{22}^{-1} \beta_{1, 2}^{-1} \beta_{2} [I_{r_o}, 0], \)

\( F_2 = R_2 \Gamma_{22}^{-1} R_2' (\alpha_o' \alpha_o)^{-1} \alpha_o' [I_{r_o}, 0] \Gamma_{11}^{-1}, \) and \( \{S_i\}, R_2 \) and \( \Gamma_{22} \) are defined in Lemma 2 in Appendix.

**Remark 2.1.** If replacing \( A, B \) and \( B_\perp, B_\perp, \) and \( \hat{B} \) in She and Ling (2020) by \( \alpha_o, \beta_o', \alpha_o', \beta_\perp, \) and \( \beta \), respectively, then limiting distributions in Theorem 4 are the same as those in She and Ling (2020) when \( r_o \) and \( p_o \) are known, i.e., the estimates \( \hat{O}_n, \hat{\alpha}_n \) and \( \hat{B}_{S_B} \) achieve their oracle property. Recently, the oracle properties of 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 \( \{\varepsilon_t\} \) are i.i.d normal or \( E\|\varepsilon_t\|^{4+\delta} < \infty \).
with $\delta > 0$. These assumptions are far away from the heavy-tailed time series as in model (2.1). It remains a challenging problem about 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 of 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 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),

\[ \lambda_{r,k,n} = \frac{\lambda^*_{r,k,n}}{\|\phi_k(\Pi_{1st})\|_\omega} \quad \text{and} \quad \lambda_{b,j,n} = \frac{m^\omega \lambda^*_{b,j,n}}{\|B_{1st,j}\|_\omega}, \quad (4.1) \]

where $\lambda^*_{r,k,n}$ and $\lambda^*_{b,j,n}$ are nonincreasing positive sequences and $\omega$ is some positive finite constant. The extra term $m^\omega$ is used to adjust the effect of 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 \( \lambda_{r,k,n}^* = o_p(n^{-\gamma_1}) \), \( \lambda_{b,j,n}^* = o_p(n^{-\gamma_2}) \), and \( n^{1-\frac{2}{\alpha}} \omega \tilde{L}(n)^{-1} \lambda_{r,k,n}^* \to \infty \) and \( n^{\gamma_2} \tilde{L}(n)^{-1} \lambda_{b,j,n}^* \to \infty \) for any \( j = 1, \ldots, p \) and \( k = 1, \ldots, m \), then under Assumption 1 and (2.2), for any \( k \in S_{\phi}^c \) and \( j \in S_{B}^c \), it follows that

\[
\delta_{r,n} + \delta_{b,n} = o_p(n^{-\gamma_1}), \quad n^{1-\frac{2}{\alpha}} \omega \tilde{L}(n)^{-1} \lambda_{r,k,n} \to_p \infty \quad \text{and} \quad n^{\gamma_2} \tilde{L}(n)^{-1} \lambda_{b,j,n} \to_p \infty.
\]

We now discuss the choice of \( \lambda_{r,k,n}^* \) and \( \lambda_{b,j,n}^* \). From Lemma 1, we see that the conditions of tuning parameters to ensure oracle properties in LS shrinkage estimation only restrict the rates at which the sequences \( \lambda_{r,k,n}^* \) and \( \lambda_{b,j,n}^* \) go to zero. But these conditions are not precise enough to provide a clear choice of tuning parameters in finite samples. On one hand, \( \lambda_{r,k,n}^* \) and \( \lambda_{b,j,n}^* \) should converge to zero as fast as possible so that a nonzero \( T(k) \) is not estimated as zero and a nonzero \( B_{o,j} \) is not estimated as zero with a high probability, which will reduce the shrinkage bias in the estimation of the nonzero components of the model. On the other hand, these sequences should converge to zero as slow as possible so that zero components in the model are estimated as zeros with a high probability in finite samples. To
take a balance, we recommend the choice of $\lambda_{r,k,n}^*$ and $\lambda_{b,j,n}^*$ as follows:

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

(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 Karuch-Kuhn-Tucker (KKT) conditions (7.6) and (7.9) in Appendix. Let $P_n = Q_n^{-1}$ and $P_n(k)$ be the $k$-th column of $P_n$. Denote

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

(4.3)

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

(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 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:

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

(4.5)
First, by Lemma 3, we have \( n\phi_k(\hat{\Pi}_{1st}) = O_p(1) \) for \( k \in S^c_\phi \) and \( \hat{B}_{1st,j} = O_p(n^{-\frac{1}{2}}\tilde{L}(n)^{-1}) \) if \( \alpha \in (1, 2) \) or \( \alpha = 1 \) and \( \tilde{L}(n) \to 0 \) and \( = O_p(n^{-1}) \) if \( \alpha \in (0, 1) \) or \( \alpha = 1 \) and \( \tilde{L}(n) \to \infty \) for \( j \in S^c_B \). By Lemma S.1 in Supplementary, if

\[
n^{1-\frac{2}{n}}n\omega\tilde{L}(n)^{-1}\lambda_{r,k,n}^* \to \infty \text{ and } n^{\tau_2}\tilde{L}(n)^{-1}\lambda_{b,j,n}^* \to \infty,
\]

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 \( \phi_k(\hat{\Pi}_{1st}) \to_p \phi_k(\Pi_o) \neq 0 \) and \( \hat{B}_{1st,j} \to_p B_{o,j} \neq 0 \) for \( k \in S_\phi \) and \( j \in S_B \). By Lemma S.1 in Supplementary, if

\[
n^{1-\frac{2}{n}}\tilde{L}(n)^{-1}\lambda_{r,k,n}^* \to 0 \text{ and } n^{\tau}\tilde{L}(n)^{-1}\lambda_{b,j,n}^* \to 0,
\]

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 (4.7) holds.

We next discuss how to choose \( c_{r,k} \) and \( c_{b,j} \). First take \( \lambda_{b,j,n}^* = 2\log(n)n^{-\frac{\tau_2}{2}} \), and \( \lambda_{r,k,n}^* = 2\log(n)n^{-\frac{1}{2}(\omega+1-\frac{2}{n})} \), which satisfy the conditions of Lemma 4.1. We do a first-step LS shrinkage estimation to get an estimates \( (\hat{T}_{1,\pi}, \hat{T}_{2,\pi}) \) of \( (T_{1,\pi_0}, T_{2,\pi_0}) \). Following the similar arguments in Liao and Phillips (2015),
we can 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 LS shrinkage estimation are given as follows:

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

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

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 $\log(Tr[\sum_{t=1}^{n} \Delta Y_t \Delta Y'_t]) / \log n$. The simulation results in the next section shows that these data dependent turning parameters work well.

5. Simulation Study

This section examines the performance of shrinkage estimates in terms of cointegrating rank selection and efficient estimation in finite samples.
Firstly we investigated the model when $m = 2$. \( \{Y_t\}_{t=1}^n \) are generated from

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

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. $\Pi_o$ is specified as follows

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

which corresponds to the cointegrating rank to be 0, 1 and 2, respectively, $B_{o,1}$ and $B_{o,3}$ are taken to be $\text{diag}(0.4, 0.4)$ and other $B_{o,j} = 0$. We take sample size $n = 100, 400, 800$ 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.
Table 1: Rank and lagged order selection with adaptive Lasso penalty for model (5.1)

<table>
<thead>
<tr>
<th>Cointegration rank selection</th>
<th>( r_o = 0 )</th>
<th>( r_o = 1 )</th>
<th>( r_o = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n=100 )</td>
<td>( n=400 )</td>
<td>( n=800 )</td>
</tr>
<tr>
<td>( \alpha = 0.2 )</td>
<td>( \hat{r}_n = 0 )</td>
<td>0.731</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>( \hat{r}_n = 1 )</td>
<td>0.262</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>( \hat{r}_n = 2 )</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>( \alpha = 1.3 )</td>
<td>( \hat{r}_n = 0 )</td>
<td>0.801</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>( \hat{r}_n = 1 )</td>
<td>0.198</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>( \hat{r}_n = 2 )</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lagged difference selection</th>
<th>( p_o = 0 )</th>
<th>( p_o = 1 )</th>
<th>( p_o = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n=100 )</td>
<td>( n=400 )</td>
<td>( n=800 )</td>
</tr>
<tr>
<td>( \alpha = 0.2 )</td>
<td>( \hat{p}_n \in T )</td>
<td>0.800</td>
<td>0.971</td>
</tr>
<tr>
<td></td>
<td>( \hat{p}_n \in C )</td>
<td>0.185</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>( \hat{p}_n \in I )</td>
<td>0.015</td>
<td>0.000</td>
</tr>
<tr>
<td>( \alpha = 1.3 )</td>
<td>( \hat{p}_n \in T )</td>
<td>0.802</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>( \hat{p}_n \in C )</td>
<td>0.198</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>( \hat{p}_n \in I )</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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

First, we are interested in the performance of the shrinkage method in selecting the correct rank of \( \Pi_o \) and the order of the lagged differences. Table 1 shows finite sample probabilities of the shrinkage method in joint rank and lag order selection. When the sample size is small (i.e. \( n = 100 \)) and the data are \( i.i.d. \), the probability 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>$\Pi_{21}$</th>
<th>$\Pi_{12}$</th>
<th>$\Pi_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bais</td>
<td>Std</td>
<td>Bais</td>
<td>Std</td>
</tr>
<tr>
<td>$\alpha = 0.2$</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>
</tr>
<tr>
<td>OLS</td>
<td>0.00297</td>
<td>0.06269</td>
<td>0.00125</td>
<td>0.07926</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00208</td>
<td>0.05298</td>
<td>0.00926</td>
<td>0.08607</td>
</tr>
<tr>
<td>$\alpha = 1$</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>
</tr>
<tr>
<td>OLS</td>
<td>0.00321</td>
<td>0.06513</td>
<td>0.00639</td>
<td>0.06543</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00768</td>
<td>0.05298</td>
<td>0.00317</td>
<td>0.07626</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$B_{1,11}$</th>
<th>$B_{1,21}$</th>
<th>$B_{1,12}$</th>
<th>$B_{1,22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bais</td>
<td>Std</td>
<td>Bais</td>
<td>Std</td>
</tr>
<tr>
<td>$\alpha = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.10046</td>
<td>0.09332</td>
<td>0.00005</td>
<td>0.12230</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00293</td>
<td>0.09335</td>
<td>0.00079</td>
<td>0.10421</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00119</td>
<td>0.09599</td>
<td>0.00032</td>
<td>0.10431</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.08917</td>
<td>0.06717</td>
<td>0.00204</td>
<td>0.06268</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00143</td>
<td>0.07762</td>
<td>0.00784</td>
<td>0.07191</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00449</td>
<td>0.06652</td>
<td>0.00530</td>
<td>0.08124</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$B_{2,11}$</th>
<th>$B_{2,21}$</th>
<th>$B_{2,12}$</th>
<th>$B_{2,22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bais</td>
<td>Std</td>
<td>Bais</td>
<td>Std</td>
</tr>
<tr>
<td>$\alpha = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00599</td>
<td>0.08441</td>
<td>0.00241</td>
<td>0.10716</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00229</td>
<td>0.07093</td>
<td>0.00576</td>
<td>0.07302</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$B_{3,11}$</th>
<th>$B_{3,21}$</th>
<th>$B_{3,12}$</th>
<th>$B_{3,22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bais</td>
<td>Std</td>
<td>Bais</td>
<td>Std</td>
</tr>
<tr>
<td>$\alpha = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.12461</td>
<td>0.07388</td>
<td>0.02298</td>
<td>0.07245</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00187</td>
<td>0.10214</td>
<td>0.00207</td>
<td>0.11513</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00120</td>
<td>0.08595</td>
<td>0.00322</td>
<td>0.08927</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.10135</td>
<td>0.06506</td>
<td>0.02273</td>
<td>0.06214</td>
</tr>
<tr>
<td>OLS</td>
<td>0.00339</td>
<td>0.06372</td>
<td>0.00365</td>
<td>0.07404</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.00449</td>
<td>0.06652</td>
<td>0.00530</td>
<td>0.08512</td>
</tr>
</tbody>
</table>

Note: Oracle estimate is the RLSE with $r_\alpha = 1$ and the restriction that $B_{o,2} = 0$. 
when $\alpha \in (1, 2)$ and $\alpha \in (0, 1)$ are almost equal to 1. The probabilities of selecting the true rank $r_o = 1$ when $\alpha \in (0, 1)$ is close to 1. The probabilities 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 1. 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 1. The probabilities of selecting the true rank of other cases equal 1. Evidently, the method performs well in selecting the true rank and true lagged differences in all scenarios. The result shows that the selecting true rank also perform well when adding lags to the model.

Table 2 presents the finite sample properties of 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 Supplementary. Table 2 shows that our method performed well in estimating the parameters overall. When compared with oracle estimates, some components in the LS shrinkage even have smaller variances, though their finite sample biases are slightly larger. Moreover, the LS shrinkage estimate generally has smaller variance when compared
with the OLS estimate, though the finite sample bias of the shrinkage estimate of 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 \( \varepsilon_t = |x_t|^{1/\alpha}(\sin \zeta_t \cos \varphi_t, \sin \zeta_t \sin \varphi_t, \cos \zeta_t) \), where \( \varphi_t, \zeta_t \sim i.i.d. U[0, 2\pi] \) and they are independent each other. \( \Pi_o \) is specified as follows:

\[
\Pi_o = o_{3 \times 3}, \quad \begin{bmatrix}
-0.5 & -0.25 & 0.5 \\
0.1 & 0.05 & -0.1 \\
0.2 & 0.1 & -0.2
\end{bmatrix}, \quad \begin{bmatrix}
-0.5 & -0.2 & 0.7 \\
0.1 & -0.3 & 0.2 \\
0.2 & 0.2 & -0.4
\end{bmatrix}, \quad \text{and} \quad \begin{bmatrix}
-0.4 & 0.0 & 0.0 \\
0.0 & -0.6 & 0.0 \\
0.0 & 0.0 & -0.8
\end{bmatrix},
\]

which corresponds to the cointegrating rank \( r_o = 0, 1, 2, 3 \), respectively, and \( B_{o,1} \) and \( B_{o,3} \) are taken to be \( \text{diag}(0.4, 0.4, 0.4) \) and other \( B_{o,j} = 0 \). The sample size is \( n = 400 \) and 800 and the number of replication is 1000. Table 3 reports the finite sample probabilities of rank and lag order selection for model (5.3). It shows that our method performs well in selecting the true rank and true lagged differences in all scenarios. The estimates of other parameters when \( r_o = 1 \) and 2 are reported in Supplementary.
Table 3: Rank and lagged order selection with adaptive Lasso penalty for model (5.2)

<table>
<thead>
<tr>
<th></th>
<th>$\alpha = 0.2$</th>
<th></th>
<th></th>
<th>$\alpha = 1.3$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_o = 0$</td>
<td>$r_o = 1$</td>
<td>$r_o = 2$</td>
<td>$r_o = 3$</td>
<td>$r_o = 0$</td>
<td>$r_o = 1$</td>
</tr>
<tr>
<td></td>
<td>n=400</td>
<td>n=800</td>
<td>n=400</td>
<td>n=800</td>
<td>n=400</td>
<td>n=800</td>
</tr>
<tr>
<td></td>
<td>$\hat{r}_n = 0$</td>
<td>0.667</td>
<td>0.909</td>
<td>0.991</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\hat{r}_n = 1$</td>
<td>0.305</td>
<td>0.087</td>
<td>0.972</td>
<td>0.990</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>$\hat{r}_n = 2$</td>
<td>0.028</td>
<td>0.044</td>
<td>0.027</td>
<td>0.010</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>$\hat{r}_n = 3$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\alpha = 0.2$</th>
<th></th>
<th></th>
<th>$\alpha = 1.3$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{p}_n \in T$</td>
<td>0.854</td>
<td>0.992</td>
<td>0.963</td>
<td>0.989</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>$\hat{p}_n \in C$</td>
<td>0.017</td>
<td>0.008</td>
<td>0.026</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>$\hat{p}_n \in I$</td>
<td>0.129</td>
<td>0.000</td>
<td>0.011</td>
<td>0.009</td>
<td>0.040</td>
</tr>
</tbody>
</table>

6. An Empirical Example

This section uses the technique in Section 2 to time series modeling of long-run and short-run behavior of prices of wheat, corn and wheat flour in the United States. The sample used in the empirical study is monthly data over the period from 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, i.e. $y_{it} = \log(x_{it})$. The log of data \{Y_t\} is shown in Figure 2. Evidently, the time series display the co-movement over the entire period. Therefore, we can use VEC model to analyze the data and try to reveal some cointegrating relations. That is, we would 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 log-returns (i.e. $r_{it} = y_{it} - y_{i,t-1}$) of each prices, that is,

$$\hat{\alpha}_i(k) = \left\{ \frac{1}{k} \sum_{t=1}^{k} \log \left( \frac{|r_{i,t}|}{|r_{i,t+1}|} \right) \right\}^{-1},$$
where \( \{|r_t|: t = 1, \cdots, n\} \) is the decreasing order statistics of \( \{|r_{it}|: t = 1, \cdots, n\} \), see Resnick (1997). Since \( \hat{\alpha}_i(k) \) relies on the choice of \( k \), Figure 3 is the plots of these estimated tail indices in term of \( k \). It shows that the tail index of each log-return most likely is less than 2 but larger than 1. It seems to be reasonable to assume these data are heavy-tailed time series.

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

\[
\Delta Y_t = \Pi Y_{t-1} + \sum_{k=1}^{10} B_k \Delta Y_{t-k} + \epsilon_t \tag{6.1}
\]

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 as the last one 0.00003 is very close to zero. The LS estimates of the lag coefficients \( B_k \) are nonzero for \( k = 1, \cdots, 10 \). We applied LS shrinkage
estimation to model (6.1). The results are as follows:

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

\[
\hat{B}_1 = \begin{bmatrix}
0.1041 & 0.0013 & 0.0054 \\
0.1046 & -0.0757 & -0.0004 \\
0.1429 & -0.0308 & -0.0286
\end{bmatrix}
\]

and \( \hat{B}_5 \) 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 2. These results corroborate the manifestation of 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

\[
\sum_{t=1}^{n} \frac{\delta_{\epsilon_{t-1}}}{\alpha_n} \overset{v}{\to} \sum_{i=1}^{\infty} \delta_{P_i} = PRM(\mu),
\]
as \( n \to \infty \), where \( PRM(\mu) \) is a Poisson random process with intensity measure \( \mu \) and \( \{P_i\} \) is a sequence of random vectors such that \( \sum_{i=1}^{\infty} \delta_{P_i} \) is the point representation of \( PRM(\mu) \). From Davis and Resnick (1985) we can see that
\[
nP\left( \frac{\varepsilon_1^t \varepsilon_2^t}{a_n} \in \cdot \right) \xrightarrow{w} \tilde{\mu}(\cdot) \text{ and } a_n/a_n \to \infty,
\]
as \( n \to \infty \) when \( E\|\varepsilon_1\|^2 = \infty \), where \( \tilde{\mu} \) is a Radon measure on \((\mathbb{R}^{m^2}, \mathcal{B}^{m^2})\).

Let \( \{P_i^{(j)}\} \) be a sequence of random vectors such that \( \sum_{i=1}^{\infty} \delta_{P_i^{(j)}} \) is the point representation of \( PRM(\tilde{\mu}) \) for \( j = 2, 3, \ldots \) and they are independent each other for different \( j \).

Since \( Z_{1,t-1} \) is a stationary process and \( Z_{2,t-1} \) comprises the I(1) components under Assumption 1 and Theorem 4.2 in Johansen(1995). Then we have the following expansions
\[
Z_{1,t} = \sum_{i=0}^{\infty} B_i \varepsilon_{t-i} \text{ and } Z_{2,t} = [I_d, 0] \sum_{i=1}^{t} \gamma_i,
\]
where \( B_i = O(\rho^i) \) with some \( \rho \in (0, 1) \). Denote

\[
R_{1n} = \sum_{t=1}^{n} \varepsilon_t Z_{1,t-1}' \quad \text{and} \quad R_{2n} = \sum_{t=1}^{n} \varepsilon_t Z_{2,t-1}',
\]
\[
S_{11n} = \sum_{t=1}^{n} Z_{1,t-1} Z_{1,t-1}' \quad \text{and} \quad S_{21n} = \sum_{t=1}^{n} Z_{2,t-1} Z_{1,t-1}' \quad \text{and} \quad S_{22n} = \sum_{t=1}^{n} Z_{2,t-1} Z_{2,t-1}'.
\]
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 \( E\|e_1\|^\alpha = \infty \). Then

\[
\begin{align*}
(a) & \quad \frac{1}{a_n} R_{1n} \xrightarrow{d} \sum_{i=0}^\infty S_{i+2} B_i', \\
(b) & \quad \frac{1}{a_n} R_{2n} \xrightarrow{d} R_2, \\
(c) & \quad \frac{1}{b_n^2} S_{22n} \xrightarrow{d} \Gamma_{22}, \\
(d) & \quad \frac{1}{a_n^2} S_{11n} \xrightarrow{d} \sum_{i=0}^\infty B_i S_1 B_i', \\
(e) & \quad \frac{1}{b_n^2} S_{21n} \xrightarrow{d} \{ R_2 \sum_{i=0}^\infty B_i' + \left[ I_d, 0 \right] \sum_{i=0}^\infty \sum_{j=0}^\infty \psi_j S_1 B_i' \}.
\end{align*}
\]

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

\( S_1 = \sum_{i=1}^\infty P_i^{(1)} P_i^{(1)'}, \) 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 (\( \hat{\Pi}_{1st}, \hat{B}_{1st} \)) and the asymptotic properties of the eigenvalues of \( \hat{\Pi}_{1st} \).

**Lemma 3.** Suppose that (2.2) and Assumptions 1-2 hold, and \( E\|e_1\|^\alpha = \infty \) and \( S_1 \) is positive definite almost surely. Then
(a). \([\hat{\Pi}_{1st}, \hat{B}_{1st}] - (\Pi_0, B_0)\] \(Q_B^{-1}D_{n,B}^{-1} \to_d (B_{m,1}, B_{m,2})\), where \(B_{m,2} = R_2 \Gamma_{22}^{-1}\),

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

\(R_1^* = \sum_{i=0}^{\infty} S_i + 2B_i^t\), \(\Gamma_{11}^* = \sum_{i=0}^{\infty} B_i S_i B_i^t\) and \(\Gamma_{21}^* = R_2^* \sum_{i=0}^{\infty} B_i^t[I_{d^*}, 0] \sum_{i=0}^{\infty} \sum_{j=0}^{i} \psi_j S_i B_i^t\),

(b). for \(k = 1, \ldots, m\), the eigenvalues of \(\hat{\Pi}_{1st}\) satisfy \(\phi_k(\hat{\Pi}_{1st}) \to_d \phi_k(\Pi_0)\),

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

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

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

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

\[
Q_{\alpha,n} \Pi_{n,f} = Q_{\alpha,n} \hat{\Pi}_{1st} - Q_{\alpha,n} P_{\alpha \perp n} \Lambda_{\alpha \perp n} Q_{\alpha \perp n} = \Lambda_{\alpha,n} Q_{\alpha,n}, \tag{7.1}
\]

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

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

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

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

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

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...
as

\[(\hat{T}_n, \hat{B}_n) = \arg \min_{T, B_1, ..., B_p \in \mathbb{R}^{m \times m}} \sum_{t=1}^{n} \| \Delta Y_t - P_n T Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j} \|^2 + n \sum_{j=1}^{p} \lambda_{b,j,n} \| B_j \| + n \sum_{k=1}^{m} \lambda_{r,k,n} \| T(k) \|. \tag{7.5} \]

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

\[
\begin{cases} 
F_{\pi,n}(k) = \frac{n \lambda_{r,k,n}}{2} \frac{\hat{T}_n(k)}{\| \hat{T}_n(k) \|} & \text{if } \hat{T}_n(k) \neq 0, \\
\| n^{-1} F_{\pi,n}(k) \| < \frac{\lambda_{r,k,n}}{2} & \text{if } \hat{T}_n(k) = 0, \end{cases} \tag{7.6} \]

for \( k = 1, \ldots, m \), where \( F_{\pi,n}(k) \) is defined in (4.2). 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

\[
\| a_n^{-2} F_{\pi,n}(k) \| = \frac{n a_n^{-2} \lambda_{r,k,n}}{2}. \tag{7.7} \]
By Lemma 2 (b), (c) and (e), and Theorem 1, it follows that

\[
\frac{1}{a_n^2} F_{\pi,n}(k_0) = \frac{1}{a_n^2} \sum_{t=1}^{n} [\varepsilon_t - (\hat{\theta}_n - \theta_o)Q_B^{-1}Z_{t-1}]'P_n(k_0)Y_{t-1}' \tag{7.8}
\]

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

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

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

\[
P(Q_n(k_0)\hat{\Pi}_n = 0) = P(\hat{T}_n(k_0) = 0) \to 1 \text{ as } n \to \infty,
\]

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

\[
P(Q_{a_k,n}\hat{\Pi}_n = 0) \to 1 \text{ as } n \to \infty.
\]

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

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

for any \(j = 1, \ldots, p\), where \(F_{b,n}(j)\) is defined in (4.3). On the event \(\hat{B}_{n,j} \neq 0\),
For some $j \in S^c_B$, we get the following equation from the optimality conditions

$$\|F_{b,n}(j)\| = \frac{n\lambda_{b,j,n}}{2}. \quad (7.10)$$

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

$$\|\tilde{\delta}_n F_{b,n}(j)\| = \frac{n\tilde{\delta}_n \lambda_{b,j,n}}{2}. \quad (7.11)$$

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

$$\tilde{\delta}_n F_{b,n}(j) = \tilde{\delta}_n \sum_{t=1}^{n} [\varepsilon_t - (\tilde{\theta}_n - \theta_o)Q_B^{-1}Z_{t-1}]\Delta Y'_{t-j}$$

$$= \tilde{\delta}_n \sum_{t=1}^{n} \varepsilon_t \Delta Y'_{t-j} - \tilde{\delta}_n (\tilde{\theta}_n - \theta_o)Q_B^{-1} \sum_{t=1}^{n} Z_{t-1} \Delta Y'_{t-j} = O_p(1). \quad (7.12)$$

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

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

for any $j \in S^c_B$, which finishes the proof.

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

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

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

which implies that

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

(7.13)

Let $\delta_n^* = a_n^2 / \tilde{a}_n$ 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

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

(7.14)

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

(7.15)

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

(7.16)

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

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

(7.17)
which implies that (7.14) holds, and

\[ \delta_n^*[(\hat{\alpha}_n - \alpha_o)\hat{\beta}'_n + \alpha_o(\hat{\beta}_n - \beta_o)]\alpha_o(\beta'_o\alpha_o)^{-1} = \delta_n^*(\hat{\Pi}_n - \Pi_0)\alpha_o(\beta'_o\alpha_o)^{-1} = O_p(1), \]

(7.17)

\[ n\hat{\alpha}_n(\hat{\beta}_n - \beta_o)'\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = n(\hat{\Pi}_n - \Pi_0)\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = O_p(1). \]

(7.18)

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

\[ \beta'_o\hat{\alpha}_n[n(\hat{O}_n - O_{ro})]'\beta_{2,o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1} = O_p(1). \]

(7.19)

By (7.13) and (7.19), we have

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

(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 \( \bar{L}(n) \rightarrow \infty \) 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 \( \bar{L}(n) \rightarrow 0 \), we can show
(7.14)-(7.16) hold.

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

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

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

Then

\[
[(\hat{\Pi}_n - \Pi_o), (\hat{B}_{SB} - B_{o,SB})] Q_S^{-1} D_{n,S}^{-1} \\
= [\delta_n^*(\hat{\alpha}_n - \alpha_o)' \alpha_o(\beta_o' \alpha_o)^{-1} + \delta_n^*(\hat{\alpha}_n - \alpha_o), \delta_n^*(\hat{B}_{SB} - B_{o,SB}), n\hat{\alpha}_n(\hat{\beta}_n - \beta_o)' \beta_{o,\perp}(\alpha_o' \beta_{o,\perp})^{-1}] \\
= [n^{-1}\delta_n^*\hat{\alpha}_n[0_{ro}, U_{2,n}^*] \alpha_o(\beta_o' \alpha_o)^{-1} + U_{1,n}^*, U_{2,n}^*, \hat{\alpha}_n[0_{ro}, U_{2,n}^*] \beta_{o,\perp}(\alpha_o' \beta_{o,\perp})^{-1}].
\]

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

\[
\Pi_n(U) = [n^{-1}\delta_n^*\hat{\alpha}_n[0_{ro}, U_{2}] \alpha_o(\beta_o' \alpha_o)^{-1} + U_1, U_3, \hat{\alpha}_n[0_{ro}, U_{2}] \beta_{o,\perp}(\alpha_o' \beta_{o,\perp})^{-1}].
\]
Then, \( U_n^* = (U_{1,n}^*, U_{2,n}^*, U_{3,n}^*) \) minimizes the following criterion function

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

\[
+ n \sum_{k \in S_o} \lambda_{r,k,n} [\| \Phi_{n,k}(\Pi_n(U)D_{n,S}Q_SL_1 + \Pi_0) \| - \| \Phi_{n,k}(\Pi_o) \|]
\]

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

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

For any compact set \( K \subset \mathbb{R}^{m \times r_o} \times \mathbb{R}^{r_o \times (m-r_o)} \times \mathbb{R}^{m \times m d_{S_B}} \) and any \( U \in K \), there is \( \Pi_n(U)D_{n,S}Q_SL = O_p(\delta_{n}^{-1}) \). Then we can deduce that

\[
\sum_{k \in S_o} \lambda_{r,k,n} [\| \Phi_{n,k}(\Pi_n(U)D_{n,S}Q_SL_1 + \Pi_0) \| - \| \Phi_{n,k}(\Pi_o) \|] 
\]

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

and

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

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 \( \hat{B}_{S_B} \) minimize the following criterion function
w.p.a. 1

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

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

Acknowledgments

The authors greatly appreciate the very helpful comments of two anonymous referees, the Associate Editor and the Editor Professor Xiaotong Shen.

The work was partially supported by by Hong Kong Research Grants Commission Grants (16500117, 16303118, 16301620 ) and Australian Research Council and by National Natural Science Foundation of China (11731015, 11571148) (Shiqing Ling)

References


REFERENCES


REFERENCES


Department of Mathematics

Beijing Institute of Technology, Beijing, China

E-mail: 7520200236@bit.edu.cn

Department of Mathematics
REFERENCES

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: mize@sxufe.edu.cn