LSE Research Online

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Qiwei Yao, Rongmao Zhang and Peter Robinson

Identifying cointegration by eigenanalysis

Article (Accepted version)
(Refereed)

Original citation:
Yao, Qiwei and Zhang, Rongmao and Robinson, Peter (2018) Identifying cointegration by eigenanalysis. Journal of the American Statistical Association. ISSN 0162-1459 (In Press)
DOI:

© 2018 Informa UK Limited

This version available at: http://eprints.lse.ac.uk/87431/
Available in LSE Research Online: April 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.lse.ac.uk

Identifying Cointegration by Eigenanalysis*

Rongmao Zhang Peter Robinson Qiwei Yao

Abstract

We propose a new and easy-to-use method for identifying cointegrated components of nonstationary time series, consisting of an eigenanalysis for a certain non-negative definite matrix. Our setting is model-free, and we allow the integer-valued integration orders of the observable series to be unknown, and to possibly differ. Consistency of estimates of the cointegration space and cointegration rank is established both when the dimension of the observable time series is fixed as sample size increases, and when it diverges slowly. The proposed methodology is also extended and justified in a fractional setting. A Monte Carlo study of finite-sample performance, and a small empirical illustration, are reported.

Keywords: Cointegration, Eigenanalysis, $I(d)$, Nonstationary processes, Vector time series.

*Rongmao Zhang is Professor, School of Mathematics, Zhejiang University, Hangzhou, China (E-mail: rmzhang@zju.edu.cn). Peter Robinson is Tooke Professor of Economic Science and Statistics, Department of Economics, London School of Economics, UK (E-mail: p.m.robinson@lse.ac.uk). Qiwei Yao is Professor, Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, U.K.; Guanghua School of Management, Peking University, China (E-mail: q.yao@lse.ac.uk). Partially supported by the NSFC research grants 11371318/11771390, the ZPNSFC research grant LR16A010001, the ESRC research grant ES/J007242/1, and the EPSRC research grant EP/L01226X/1.

1 Introduction

Cointegration entails a dimensionality reduction of certain observable multiple time series that are dominated by common components. In particular a multiple time series can be said to be (linearly) cointegrated if there exists an instantaneous linear combination, or cointegrating error, with lower integration order. Much of the vast literature, following Box and Tiao (1977), Granger (1981), Engle and Granger (1987), has focused on unit root series which have one or more short memory cointegrating errors, but there have been extensions to nonstationary series with other integer orders of integration, allowing also for the possibility of some nonstationary cointegrating errors, as well as to fractional nonstationary, and even stationary, observable series and cointegrating errors, with unknown integration orders. Much of the early literature, in particular, assumed a complete parameterization of second order properties, where in particular the observable series are generated from short memory inputs that have finite autoregressive moving average (ARMA) structure, but it has also been common to study semiparametric settings, with underlying short memory inputs having nonparametric autocorrelation, see e.g. Stock (1987), Phillips (1991), in some cases without sacrificing precision relative to a correctly specified parametric structure.

Given knowledge of the cointegration rank, r, of a p-dimensional observable series, that is the number of cointegrating relations, various methods are available for estimating the unknown parameters of the model, such as the coefficients of the cointegrating errors, and even of unknown integration orders, and for carrying out asymptotically valid, and sometimes even efficient, statistical inference. However, r might not be known to the practitioner, and various approaches for estimating r from the data have been developed, starting from Engle and Granger (1987), Johansen (1991), in their parametric, unit root vector autoregressive (VAR) setting, and continuing with, for example, Aznar and Salvador (2002) and Saikkonen and Lütkepohl (2000). If, however, the order of the VAR is underspecified, or all observable series do not have a single unit root, then typically the resulting specification error will invalidate such approaches, not to mention rules of statistical inference on unknown coefficients in the model. It is possible that one or more of the nonstationary observable processes could have two or more unit roots, or indeed could have fractional orders of integration, as supported by some empirical investigations. References that allow for nonparametric autocorrelation and/or unknown integration orders include Phillips and Ouliaris (1988, 1990), Bierens (1997), Stock (1999), Shintani (2001), Harris and Poskitt (2004), Li, Pan and Yao (2009) in the case of integer integration orders, and Robinson and Yajima (2002), Chen and Hurvich (2006), Robinson (2008) in case of fractional integration orders, including in the latter setting cases where observables are stationary and the cointegrating errors are stationary with less memory.

Like Phillips and Ouliaris (1988), Robinson and Yajima (2002), Harris and Poskitt (2004), Li, Pan and Yao (2009), we employ methods based on eigenanalysis. In our case, in the setting of nonparametric autocorrelation and unknown (and possibly different) integration orders, we employ eigenvalues of a certain non-negative definite matrix function of sample autocovariance matrices of the observable series, for estimating cointegration rank, with the cointegration space then estimated by selection of eigenvectors, and cointegrating errors thereby proxied. Though the initial development assumes that observable series have integer orders and cointegrating errors have short memory, we extend these results to allow for observables to be fractionally nonstationary, and cointegrating errors to be fractionally stationary. In both circumstances we establish consistency of our estimates of cointegration rank and space with p fixed as the length n of our time series diverges. In case of integer integration orders, we also establish consistency allowing p to diverge slowly with n.

The rest of the paper is organized as follows. The proposed methodology is presented in Section 2. Asymptotic theory with integer orders of integration is developed in Section 3. In Section 4, both the proposed method and part of the asymptotic theory are extended to the fractional case. Simulations and a small real data example are reported in Section 5. All statements and proofs are relegated to an Appendix, which also contains a number of technical lemmas.

2 Methods

2.1 Setting

We call a vector process $\mathbf{u}_t$ weakly stationary if (i) $E\mathbf{u}_t$ is a constant vector independent of t, and (ii) $E\|\mathbf{u}_t\|^2 < \infty$, and $\mathrm{Cov}(\mathbf{u}_t, \mathbf{u}_{t+s})$ depends on s only for any integers t, s, where $\|\cdot\|$ denotes the Euclidean norm. Denote by ∇ the difference operator, i.e. $\nabla\mathbf{u}_t = \mathbf{u}_t - \mathbf{u}_{t-1}$, and $\nabla^d\mathbf{u}_t = \nabla(\nabla^{d-1}\mathbf{u}_t)$ for any integer $d \geq 1$. We use the convention $\nabla^0\mathbf{u}_t = \mathbf{u}_t$. Further, if $\mathbf{u}_t$ has spectral density matrix that is finite and positive definite at zero frequency we say $\mathbf{u}_t$ is an $I(0)$ process. An example of an $I(0)$ process is a stationary an invertible vector ARMA, and many $I(0)$ processes satisfy Condition 1 of Section 3.1 below, imposed for our asymptotic theory, including the examples described immediately after Condition 1. Now denote by u_{it} the ith element of $\mathbf{u}_t$ and define $u_{it}^+ = u_{it}1(t \geq 1)$, where $1(\cdot)$ is the indicator function. For an m-dimensional $I(0)$ process $\mathbf{u}_t$ and non-negative integers $d_1, ..., d_m$, we say that $\mathbf{v}_t = \left(\nabla^{-d_1}u_{1t}^+, ..., \nabla^{-d_m}u_{mt}^+\right)'$ is an (m-dimensional) $I(d_1, ..., d_m)$ process, with some abuse of notation when $m = 1$, $d_1 = 0$. Note that for $d_1 = ... = d_m = 0$, $\mathbf{v}_t$ is not $I(0)$ or even weakly stationary or equivalent to $\mathbf{u}_t$ due to the truncation (implying $\mathbf{v}_t = 0$, $t \leq 0$) that is imposed in order to achieve bounded variance in case of positive d_i, but it is 'asymptotically' weakly stationary and $I(0)$. When $d_1 = ... = d_m = 1$, all

elements of $\mathbf{v}_t$ have a single unit root, but we are concerned with processes for which d_i can vary over i.

Now assume a $p \times 1$ observable time series $\mathbf{y}_t$ is $I(d_1, ..., d_p)$ for non-negative integers, and admits the form

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t, \tag{2.1}$$

where $\mathbf{A}$ is an unknown and invertible constant matrix, $\mathbf{x}_t = (\mathbf{x}_{t1}', \mathbf{x}_{t2}')'$ is a latent $p \times 1$ process, $\mathbf{x}_{t2}$ is an $r \times 1$ $I(0)$ process, and $\mathbf{x}_{t1}$ is an $I(c_1, ..., c_{p-r})$ process, where each c_i is an element of the set $\{d_1, ..., d_p\}$. Furthermore no linear combination of $\mathbf{x}_{t1}$ is $I(0)$, as such a stationary variable can be absorbed into $\mathbf{x}_{t2}$. Each component of $\mathbf{x}_{t2}$ is a cointegrating error of $\mathbf{y}_t$ and $r \geq 0$ is the cointegration rank. In the event that there exists no cointegration among the components of $\mathbf{y}_t$, $r = 0$. When $\mathbf{y}_t$ itself is $I(0, \cdots, 0)$, $r = p$. But these are two extreme cases. Note that cointegration requires equality of at least two d_i. For many economic and financial applications, there exist a small number of cointegrated variables, i.e. $r \geq 1$ is a small integer.

The pair $(\mathbf{A}, \mathbf{x}_t)$ in (2.1) is not uniquely defined, as it can be replaced by $(\mathbf{A}\mathbf{H}^{-1}, \mathbf{H}\mathbf{x}_t)$ for any invertible $\mathbf{H}$ of the form

$$\begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{0} & \mathbf{H}_{22} \end{pmatrix}$$

where $\mathbf{H}_{11}, \mathbf{H}_{22}$ are square matrices of size $(p-r)$, r respectively, and $\mathbf{0}$ denotes a matrix with all entries equal to 0. Therefore there is no loss of generality in assuming $\mathbf{A}$ to be orthogonal, because any non-orthogonal $\mathbf{A}$ admits the decomposition $\mathbf{A} = \mathbf{Q}\mathbf{U}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{U}$ is upper-triangular, and we may then replace $(\mathbf{A}, \mathbf{x}_t)$ in (2.1) by $(\mathbf{Q}, \mathbf{U}\mathbf{x}_t)$. In the sequel, we always assume that $\mathbf{A}$ in (2.1) is orthogonal, i.e., $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$, where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. Write

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2),$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are respectively, $p \times (p-r)$ and $p \times r$ matrices. As now $\mathbf{x}_{t2} = \mathbf{A}_2'\mathbf{y}_t$, the linear space spanned by the columns of $\mathbf{A}_2$, denoted by $\mathcal{M}(\mathbf{A}_2)$, is called the *cointegration space*. In fact this cointegration space is uniquely defined by (2.1), though $\mathbf{A}_2$ itself is not.

To highlight the key idea of the new approach, we only consider in this section and also Section 3 below the cointegration with $\mathbf{x}_{t2} \sim I(0)$. The extension of our method to the cases when $\mathbf{x}_{t2} \sim I(d)$ with $0 < d < \min_{1 \leq j \leq p} d_j$ are presented in Section 4 which also allows d_j's and d to be fractional numbers.

2.2 Estimation

The goal is to determine the cointegration rank r in (2.1) and to identify $\mathbf{A}_2$, or more precisely $\mathcal{M}(\mathbf{A}_2)$. Then $\mathcal{M}(\mathbf{A}_1)$ is the orthogonal complement of $\mathcal{M}(\mathbf{A}_2)$, and $\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{y}_t$ for $i = 1, 2$. Our estimation method is motivated by the following observation. For $j \geq 0$, let

$$\widehat{\mathbf{\Sigma}}_j = \frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{y}_{t+j} - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})', \qquad \bar{\mathbf{y}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t.$$

For any $\mathbf{a} \in \mathcal{M}(\mathbf{A}_2)$, $\mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a}$ is the sample autocovariance function at lag j for the weakly stationary univariate time series $\mathbf{a}'\mathbf{y}_t$, and it converges to a finite constant (i.e. the autocovariance function of $\mathbf{a}'\mathbf{y}_t$ at lag j) almost surely under some mild conditions. However for any $\mathbf{a} \notin \mathcal{M}(\mathbf{A}_2)$, $\mathbf{a}'\mathbf{y}_t$ is $I(d)$ for some $d \geq 1$, and

$$\mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a} = O_e(n^{2d-1}) \quad \text{or} \quad O_e(n^{2d}), \tag{2.2}$$

depending on whether $E(\mathbf{a}'\mathbf{y}_t) = 0$ or not, see Theorems 1 & 2 of Peña and Poncela (2006). In the above expression, $U = O_e(V)$ indicates that $P(0 < |U/V| < \infty) \to 1$. Hence intuitively the r directions in the cointegration space $\mathcal{M}(\mathbf{A}_2)$ make $|\mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a}|$ as small as possible for all $j \geq 0$.

To combine information over different lags, define

$$\widehat{\mathbf{W}} = \sum_{j=0}^{j_0}\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{\Sigma}}_j', \tag{2.3}$$

where $j_0 \geq 1$ is a prespecified and fixed integer with respect to n throughout. We use the product $\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{\Sigma}}_j'$ instead of $\widehat{\mathbf{\Sigma}}_j$ to ensure each term in the sum is non-negative definite, and that there is no information cancellation over different lags. Note that $\mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a} = O_e(1)$ if $\mathbf{a} \in \mathcal{M}(\mathbf{A}_2)$, and is at least of the order of n^{2d-1} if $\mathbf{a} \in \mathcal{M}(\mathbf{A}_1)$, where d is the minimum integration order of the components $\mathbf{x}_{t1}$. It can be shown that the $(p-r)$ largest eigenvalues of $\widehat{\mathbf{W}}$ are at least of the order n^{2d-1}, while the other r eigenvalues are $O_e(1)$ (see (7.14), (7.15) below). Hence intuitively $\mathcal{M}(\mathbf{A}_2)$ can be estimated by the linear space spanned by the r eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the r smallest eigenvalues, and $\mathcal{M}(\mathbf{A}_1)$ can be estimated by that spanned by the $(p-r)$ eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the $(p-r)$ largest eigenvalues.

Let $(\widehat{\gamma}_1, \cdots, \widehat{\gamma}_p)$ be the orthonormal eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the eigenvalues arranged in descending order. Define

$$\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2), \quad \widehat{\mathbf{x}}_{t1} = \widehat{\mathbf{A}}_1'\mathbf{y}_t \quad \text{and} \quad \widehat{\mathbf{x}}_{t2} = \widehat{\mathbf{A}}_2'\mathbf{y}_t. \tag{2.4}$$

Then $\mathcal{M}(\widehat{\mathbf{A}}_1)$ and $\mathcal{M}(\widehat{\mathbf{A}}_2)$, the linear spaces spanned by the eigenvectors of $\widehat{\mathbf{W}}$, are consistent estimators for $\mathcal{M}(\mathbf{A}_1)$ and $\mathcal{M}(\mathbf{A}_2)$ respectively; see Theorem 1 below.

The idea of using an eigenanalysis based on a quadratic form of sample autocovariance matrices has been used for factor modelling for dimension reduction (Lam and Yao 2012, and references within), and for segmenting a high-dimensional time series into several both contemporaneously and serially uncorrelated subseries (Chang et al. 2017). One distinctive advantage of using the quadratic form $\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{\Sigma}}_j'$ instead of $\widehat{\mathbf{\Sigma}}_j$ in (2.3) is that there is no information cancellation over different lags. Therefore this approach is insensitive to the choice of j_0 in (2.3). Often small values such as $j_0 = 5$ are sufficient to catch the relevant characteristics, as serial dependence is usually most predominant at small lags. Using different values of j_0 hardly changes the results; see Table 5 in Section 5 below, and also Lam and Yao (2012) and Chang et al. (2017).

2.3 Determining cointegration ranks

The components of $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t \equiv (\widehat{x}_t^1, \cdots, \widehat{x}_t^p)'$, defined in (2.4), are arranged according to descending order of the eigenvalues of $\widehat{\mathbf{W}}$. Therefore, the order of the components reflects inversely the closeness to stationarity of the component series, with $\{\widehat{x}_t^p\}$ most likely being stationary, and $\{\widehat{x}_t^1\}$ most likely being $I(d)$ with largest possible integer $d \geq 1$. Let $S_i(m) = \sum_{k=1}^m \widehat{\rho}_i(k)$, where $\widehat{\rho}_i(\cdot)$ is the sample autocorrelation function (ACF) of $\widehat{x}_t^i$ defined as

$$\widehat{\rho}_i(k) = \Big(\frac{1}{n-k}\sum_{t=1}^{n-k}(\widehat{x}_{t+k}^i - \overline{\widehat{x}}^i)(\widehat{x}_t^i - \overline{\widehat{x}}^i)\Big)\Big/\Big(\frac{1}{n}\sum_{t=1}^{n}(\widehat{x}_t^i - \overline{\widehat{x}}^i)^2\Big), \quad i = 1, 2, \cdots, p,$$

where $\overline{\widehat{x}}^i = \sum_{t=1}^n \widehat{x}_t^i/n$. When $\widehat{x}_t^i$ is stationary and suitable additional conditions hold, $\lim_{m\to\infty} S_i(m) < \infty$ in probility, however, when $\widehat{x}_t^i$ is non-stationary, $\widehat{\rho}_i(k) \to 1$ in probability for any fixed k. Hence $\lim_{m\to\infty} S_i(m) = \infty$. Therefore, we can estimate the cointegration rank r by

$$\widehat{r} = \sum_{i=1}^{p} I\{S_i(m)/m < c_0\} \tag{2.5}$$

for some constant $0 < c_0 < 1$ and large m. For a classical stationary ARMA time series, the autocorrelation $\rho_i(k)$ decays exponentially, i.e., there exists a $\rho \in (0, 1)$ such that $\rho_i(k) = O(\rho^k)$. Hence it is usually sufficient to use a moderate m in (2.5). In our numerical experiments reported in Section 5, we always set $c_0 = 0.3$ and $m = 20$, and the estimator $\widehat{r}$ performs very well and is robust across the different settings.

Remark 1. *For unit-root processes, $\widehat{r}$ defined in (2.5) typically takes the value 0 with probability approaching 1. To appreciate this, let $y_t = y_{t-1} + \varepsilon_t$ be a unit root process and $\widehat{\rho}(k)$ be its sample ACF $\widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0)$, where*

$$\widehat{\gamma}(i) = \frac{1}{n}\sum_{t=1}^{n-i}(Y_t - \overline{Y})(Y_{t+i} - \overline{Y}), \quad \overline{Y} = \sum_{i=1}^{n} Y_i/n.$$

Under some regularity conditions on ε_t, similar to those in Theorem 1 of Bierens (1993), it can be shown that

$$\frac{n}{m+1}\Big(\frac{\sum_{k=1}^{m}\widehat{\rho}(k)}{m}-1\Big)\xrightarrow{d}-\frac{(W(1)-\int_0^1 W(t)\,dt)^2+(\int_0^1 W(t)\,dt)^2+d_m}{4[\int_0^1 W^2(t)\,dt-(\int_0^1 W(t)\,dt)^2]}, \tag{2.6}$$

where

$$d_m=\frac{1}{\sigma^2}\Big(c(0)+2\sum_{i=1}^{m-1}\frac{(m-i)(m-i+1)}{m(m+1)}c(i)\Big),\quad c(i)=\mathrm{cov}(\varepsilon_0,\varepsilon_{\mathrm{i}}),\quad \sigma^2=\lim_{\mathrm{n}\to\infty}\frac{1}{\mathrm{n}}\mathrm{E}\Big(\sum_{\mathrm{s}=1}^{\mathrm{n}}\varepsilon_{\mathrm{s}}\Big)^2.$$

Thus $\sum_{t=1}^{m}\widehat{\rho}(k)/m\xrightarrow{p}1$, provided that n/m is large enough.

We may also estimate r by unit-root tests. For a given integer $r_0\le 1$, testing a hypothesis on cointegration order $H_0: r<r_0$ can be transformed to testing a unit-root hypothesis

$$H_0:\widehat{x}_t^{p-r_0+1}\sim I(d)\text{ for some integer }d\ge 1. \tag{2.7}$$

We can apply the test method of Phillips and Ouliaris (1988) to test (2.7) as d may be greater than 1. When the null hypothesis H_0 is rejected, we conclude r is at least as large as r_0.

2.4 Estimation for high integration orders

Let $r_1,\cdots,r_q$ be q positive integers, and $r_1+\cdots+r_q=p-r$. Let $1\le a_1<\cdots<a_q$ be q integers such that $\mathbf{x}_{t1}=(\mathbf{x}_{t1q},\cdots,\mathbf{x}_{t11})=(\mathbf{A}_{1q}'\mathbf{y}_t,\cdots,\mathbf{A}_{11}'\mathbf{y}_t)$, where $\mathbf{x}_{t1j}$ is an $r_j\times 1$ $I(a_j)$ process. Let

$$\widehat{\mathbf{A}}_1=(\widehat{\mathbf{A}}_{1q},\cdots,\widehat{\mathbf{A}}_{11}), \tag{2.8}$$

where $\widehat{\mathbf{A}}_{1j}$ has r_j columns. Then $\widehat{\mathbf{x}}_{t1j}=\widehat{\mathbf{A}}_{1j}'\mathbf{y}_t$ is the estimated component of $\mathbf{x}_{t1}$ of integration order a_j.

Similar to Section 2.3 above, a unit-root test can be adapted to estimate the sizes $r_1,\cdots,r_q$ and the integration orders $a_1,\cdots,a_q$. We illustrate the idea below by outlining the steps in estimating (a_1,r_1), they can be repeated in order to estimate $(a_2,r_2),(a_3,r_3),\cdots$.

For $\widehat{r}$ defined in (2.5), let $\widehat{a}_1$ be the minimum integer $d\ge 1$ such that a unit-root test rejects $H_0:\nabla^d\widehat{x}_t^{p-\widehat{r}}\sim I(1)$ against $H_1:\nabla^d\widehat{x}_t^{p-\widehat{r}}\sim I(0)$. Then the size r_1 can be estimated by applying estimator (2.5) to the $(p-\widehat{r})\times 1$ series $\{\nabla^{\widehat{a}_1}\widehat{x}_t^j,\ j=1,\cdots,p-\widehat{r}\}$.

3 Asymptotic Properties

In this section, we investigate the asymptotic properties of the proposed statistics. First, we show that with r given, the linear space $\mathcal{M}(\widehat{\mathbf{A}}_2)$ consistently estimate the cointegration space $\mathcal{M}(\mathbf{A}_2)$. We measure the distance between the two spaces by

$$D(\mathcal{M}(\widehat{\mathbf{A}}_2),\mathcal{M}(\mathbf{A}_2))=\sqrt{1-\frac{1}{r}\mathrm{tr}(\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\mathbf{A}_2\mathbf{A}_2')}. \tag{3.1}$$

Then $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) \in [0,1]$, being 0 if and only if $\mathcal{M}(\widehat{\mathbf{A}}_2) = \mathcal{M}(\mathbf{A}_2)$, and 1 if and only if $\mathcal{M}(\widehat{\mathbf{A}}_2)$ and $\mathcal{M}(\mathbf{A}_2)$ are orthogonal. Furthermore, we show that the estimator $\widehat{r}$, defined in (2.5), is consistent. We consider two asymptotic regimes: (i) p is fixed while $n \to \infty$, and (ii) $p \to \infty$ more slowly than n.

Put $\mathbf{x}_{t1} = (x_t^1, \cdots, x_t^{p-r})'$. Under (2.1), x_t^j is $I(d_j)$ for $1 \le j \le p-r$ and $z_t^j \equiv \nabla^{d_j} x_t^j$ is I(0), where $d_j \ge 1$ is an integer. Write $\mathbf{z}_t = (z_t^1, \cdots, z_t^{p-r})'$ and $\boldsymbol{\varepsilon}_t = (\mathbf{z}_t', \mathbf{x}_{t2}')'$. Denote the vector of partial sums of components of $\boldsymbol{\varepsilon}_t$ by

$$\mathbf{S}_n(\mathbf{t}) \equiv (S_n^1(t_1), \cdots, S_n^p(t_p))' = \Big(\frac{1}{\sqrt{n}} \sum_{l=1}^{[nt_1]} (\varepsilon_l^1 - \mathrm{E}\varepsilon_1^1), \cdots, \frac{1}{\sqrt{n}} \sum_{l=1}^{[nt_p]} (\varepsilon_l^p - \mathrm{E}\varepsilon_1^p)\Big)',$$

where $0 < t_1 < \cdots < t_p \le 1$ are constants and $\mathbf{t} = (t_1, \cdots, t_p)'$.

3.1 When $n \to \infty$ and p is fixed

We introduce a regularity condition first.

Condition 1.

(i) There exists a Gaussian process $\mathbf{W}(\mathbf{t}) = (W^1(t_1), \cdots, W^p(t_p))'$ such that as $n \to \infty$,

$$\mathbf{S}_n(\mathbf{t}) \overset{J_1}{\Longrightarrow} \mathbf{W}(\mathbf{t}), \quad \text{on } D^p(0,1),$$

where $\overset{J_1}{\Longrightarrow}$ denotes weak convergence under Skorohod J_1 topology (Chapter 3 in Billingsley 1999), and $\mathbf{W}(\mathbf{1})$ has a positive definite covariance matrix $\boldsymbol{\Omega} = (\sigma_{ij})$.

(ii) The sample autocovariance matrix of $\mathbf{x}_{t2}$ satisfies

$$\max_{0 \le j \le j_0} \Big\| \frac{1}{n} \sum_{t=1}^{n-j} (\mathbf{x}_{t+j,2} - \bar{\mathbf{x}}_2)(\mathbf{x}_{t2} - \bar{\mathbf{x}}_2)' - \mathrm{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}_{1,2}) \Big\|_2 \xrightarrow{p} 0,$$

where $\|\mathbf{H}\|_2 = \max_{\|\mathbf{a}\|=1} \|\mathbf{H}\mathbf{a}\|$ is the L_2-norm of matrix $\mathbf{H}$, $\bar{\mathbf{x}}_2$ is the sample mean of $\mathbf{x}_{t2}$, and $\xrightarrow{p}$ denotes convergence in probability.

Note that our definition of cointegration is formally different from that of Johansen (1995) which is based on ARIMA framework. There are some subtle technical differences between the respective conditions. For example, Condition 1(i) above implies $\det(\mathrm{Var}(\boldsymbol{\varepsilon}_t)) \neq 0$ while Johansen's setting allows the ARIMA process driven by a degenerate innovation process.

In fact, Condition 1 is mild. It is fulfilled when $\{\boldsymbol{\varepsilon}_t\}$ is weakly stationary with $\det(\mathrm{Var}(\boldsymbol{\varepsilon}_t)) \neq 0$, $\mathrm{E}\|\boldsymbol{\varepsilon}_t\|^{2\gamma} < C$ for some constants $\gamma > 1$ and $C < \infty$, and $\{\boldsymbol{\varepsilon}_t\}$ is also α-mixing with mixing coefficients α_m satisfying the condition $\sum_{m=1}^{\infty} \alpha_m^{1-1/\gamma} < \infty$; see Theorem 3.2.3 of Lin and Lu (1997). It is also fulfilled when $\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\eta}_{t-j}$, where $\boldsymbol{\eta}_t$ are i.i.d. with non-singular covariance

matrix and $E\|\boldsymbol{\eta}_t\|^{4\gamma} < \infty$ for some constant $\gamma > 1$, and $\det(\sum_{j=0}^{\infty} \mathbf{C}_j) \neq 0$, $\sum_{j=1}^{\infty} ||\mathbf{C}_j|| < \infty$. See Fakhre-Zakeria and Lee (2000). Note that our setting accommodates the cases when $\mathbf{y}_t$ contains linear deterministic components, as we allow $\mathrm{E}(\boldsymbol{\varepsilon}_t) \neq 0$.

Theorem 1. *Let r be known. Under Condition 1, $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = o_p(1)$. Furthermore,*

(i) $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_e(n^{-2a_1+1})$ provided either (a) $|I_0| \geq 2$ or (b) $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$,

(ii) $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_e(n^{-2a_1})$ provided $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} \neq 0$, and

(iii) $D(\mathcal{M}(\widehat{\mathbf{A}}_{1j}), \mathcal{M}(\mathbf{A}_{1j})) = O_e(n^{-2\alpha_j})$ for $j = 1, \cdots, q$ provided $\mathrm{E}\mathbf{z}_t = 0$,

where $I_0 = \{i : x_t^i \sim I(a_1), 1 \leq i \leq p - r\}$, $|I_0|$ denotes the number of elements in I_0, $\alpha_j = \min\{a_j - a_{j-1}, a_{j+1} - a_j\}$, $a_0 = 1/2$ and a_j, $j = 1, \cdots, q$ are defined in Section 2.4.

Remark 2. *When $\mathrm{E}\mathbf{z}_t \neq 0$, we can express the components x_t^i of $\mathbf{x}_{t1}$ as*

$$(1 - B)^{d_i} x_t^i = (z_t^i - \mathrm{E}z_t^i) + \mathrm{E}z_t^i =: \varepsilon_t^i + \mu_i.$$

Hence

$$x_t^i = (1 - B)^{-d_i} \varepsilon_t^i + \mu_i \prod_{l=0}^{d_i - 1} (t + l)/(d_i!) =: \xi_t^i + \mu_i \prod_{l=0}^{d_i - 1} (t + l)/(d_i!).$$

This entails $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t = \mathbf{A}(\boldsymbol{\xi}_t', \mathbf{x}_{t2}')' + \mathbf{B}(1, t, t^2, \cdots, t^{a_q})'$, where $\mathbf{B}$ is a $p \times a_q$ matrix. We can estimate $\mathbf{B}$ by the least squares method based on $\{\mathbf{y}_t\}$, and identify the cointegration subspaces spanned by $\mathbf{A}_{1j}$ using the detrending series $\widetilde{\mathbf{y}}_t = \mathbf{y}_t - \widehat{\mathbf{B}}(1, t, t^2, \cdots, t^{a_q})'$. It can then be shown that Theorem 1 (iii) still holds.

Theorem 2. *Under Condition 1, $\lim_{m \to \infty} P(\,\widehat{r} = r\,) = 1$.*

3.2 When $n \to \infty$, $p \to \infty$ and $p = O(n^c)$

We extend the asymptotic results in the previous section to the cases when $p \to \infty$ and $p = O(n^c)$ for some $c \in (0, 1/2)$. Technically we employ a normal approximation method to establish the results. See Condition 2(i) below.

Condition 2.

(i) Suppose that there exists an m-dimensional vector $\boldsymbol{e}_t$ with mean zero and independent components such that $\mathbf{z}_t = \mathbf{B}\boldsymbol{e}_t$, where $\mathbf{B}$ is a $(p-r) \times m$ matrix, $m \geq p - r$ and $\|\mathbf{B}\|_2 < \infty$. For each component e_t^i of $\boldsymbol{e}_t$, there exists an independent and standard normal sequence $\{\nu_t^i\}$ for which as $n \to \infty$,

$$\max_{1 \leq i \leq m} \max_{0 \leq t \leq 1} \mathrm{E}\Big[\sum_{s=1}^{[nt]} (e_s^i - \sigma_{ii}\nu_s^i)\Big]^2 = O(n^{2\tau}), \tag{3.2}$$

where $0 < \tau < 1/2$ is a constant, $b_1 \leq \sigma_{ii}^2 \equiv \lim_{n\to\infty} \mathrm{Var}\left(\sum_{s=1}^n e_s^i\right)/n \leq b_2$ for any i, and b_1, b_2 are two positive constants.

(ii) The sample autocovariance matrix of $\mathbf{x}_{t2}$ satisfies

$$\max_{0\leq j\leq j_0} \left\| \frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2}-\bar{\mathbf{x}}_2)(\mathbf{x}_{t2}-\bar{\mathbf{x}}_2)' - \mathrm{Cov}(\mathbf{x}_{1+j,2},\mathbf{x}_{1,2})\right\|_2 \xrightarrow{p} 0.$$

(iii) Suppose $\{\mathbf{z}_t\}$ and $\{\mathbf{x}_{t2}\}$ are independent and for τ given above

$$\max_{p-r<j\leq p} \sum_{s,t=1}^{n} |\mathrm{E}(\varepsilon_s^j \varepsilon_t^j)| = O(n^{1+2\tau}).$$

Remark 3. *The inequalities immediately below (3.2) holds when all components series of $\mathbf{z}_t$ are I(0) with spectral density continuous at zero frequency. This is guaranteed by the fact that their variance is proportional to the Cesaro sum of the Fourier series of the spectral density at zero frequency, and thus converges to the latter (which is positive and finite under I(0)) after normalization.*

Remark 4. *The form $\mathbf{z}_t = \mathbf{B}\boldsymbol{e}_t$ in Condition 2 (i), has been used by Bai and Saranadasa (1996) and Chen and Qin (2010). Many classic vector time series including stationary VAR, VARMA and more generally the linear process*

$$\mathbf{z}_t = \sum_{j=0}^{\infty} \mathbf{B}_j \boldsymbol{e}_{t-j}$$

with $\sum_{j=0}^{\infty} \|\mathbf{B}_j\|_2 < \infty$ follow this from. We require $m \geq p-r$, which ensures that no linear combination of $\mathbf{z}_t$ is $I(0)$. The assumption on the independence between $\{\mathbf{z}_t\}$ and $\{\mathbf{x}_{t2}\}$ in Condition 2(iii) ensures that cross correlation of $\{\mathbf{z}_t\}$ and $\{\mathbf{x}_{t2}\}$ is negligible in deriving the properties of the eigenvalues of $\widehat{\mathbf{W}}$, which can be replaced by the condition that $\mathrm{E}(n^{-(d_i+1/2)}\sum_{t=1}^n x_t^i x_t^h)^2 = o(1/(pr))$.

Remark 5. *Let $p = o(n^{1/2})$. Condition 2 is implied by any of the three assertions below.*

(i) The components of $\boldsymbol{\varepsilon}_t$ are independent of each other, and each component series $\{\varepsilon_t^i\}$ is a martingale difference sequence with $\max_{1\leq i\leq p} \mathrm{E}|\varepsilon_t^i|^q < \infty$ for some $q > 2$. Furthermore, for some $2 < q^* \leq \min\{4, q\}$,

$$\max_{1\leq i\leq p} \mathrm{E}\left|\sum_{t=1}^{n}[(\varepsilon_t^i)^2 - \sigma_{ii}^2]\right| = O(n^{2/q^*}).$$

(ii) The components of $\boldsymbol{\varepsilon}_t$ are independent, $\mathrm{E}\boldsymbol{\varepsilon}_t = 0$, and $\max_{1\leq i\leq p} \mathrm{E}|\varepsilon_t^i|^\kappa < \infty$ for some $\kappa > q \in (2, 4]$. The process $\{\boldsymbol{\varepsilon}_t\}$ is α-mixing with mixing coefficients α_m satisfying

$$\sum_{m=1}^{\infty} \alpha_m^{(\kappa-q)/(\kappa q)} < \infty. \tag{3.3}$$

(iii) The components of $\boldsymbol{\varepsilon}_t$ are independent. Each component ε_t^i satisfies the following conditions.

(a) There exists an i.i.d random sequence $\{\eta_t^i\}$ such that

$$\varepsilon_t^i = \sum_{j=0}^{\infty} c_{ij}\eta_{t-j}^i.$$

(b) $\mathrm{E}\varepsilon_t^i = 0$, $\mathrm{E}|\varepsilon_t^i|^q < \infty$ for some $q > 2$ and $\sum_{j=0}^{\infty} j|c_{ij}| < \infty$.

Theorem 3. *Let r be known and Condition 2 hold. If $p = o(n^{1/2-\tau})$ and τ given in Condition 2,*

$$D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_p(p^{1/2}n^{-2a_1+1}(\lambda^*)^{-1}),$$

where λ^* is the smallest eigenvalue of $\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\,dt$ defined in Lemma 9 in Section 7 below.

Remark 6. *Theorem 3 is derived under the condition $p = o(n^{1/2-\tau})$, while there are no direct constraints on either r or $p-r$. However when $p-r$ is fixed, $\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\,dt$ is a $(p-r)\times(p-r)$ positive definite matrix, and, hence, λ^* is positive and $O_e(1)$. When the integration orders of all the nonstationary components are the same and equal to d_{min}, then $(\lambda^*)^{-1} = O_p((p-r)^{2d_{min}-1})$.*

Theorem 4. *Let Condition 2 hold and $p = o(n^{1/2-\tau})$. Then*

$$\lim_{n\to\infty} P(\,\widehat{r} = r\,) = 1,$$

provided $(\lambda^*)^{-1}p^{1/2}n^{-a_1+1/2} = o(1)$.

4 Fractional cointegration

Fractional cointegration has attracted increasing attention in recent years, see, e.g., Robinson and Hualde (2003), Chen and Hurvich (2006) and Robinson (2008). In this section, we generalize the method presented in Section 2 to cases when the components of $\mathbf{y}_t$ may be fractionally integrated. For simplicity, we now assume p is fixed.

Let $v_t^+ = v_t\mathbf{1}(t > 0)$ and for any $\alpha \in \mathbb{R}$,

$$\Delta^{-\alpha} = \sum_{j=0}^{\infty} a_j(\alpha)B^j, \quad a_j(\alpha) = \frac{\Gamma(j+\alpha)}{\Gamma(\alpha)\Gamma(j+1)}$$

be formally defined as in Hualde and Robinson (2010), where B is the backshift operator. With these definitions we can extend the definition of the $I\,(d_1, ..., d_m)$ process $\mathbf{v}_t$ in Section 2 to non-negative real-valued d_i, such that $d_i \neq k - 1/2$ for any integer k. Note that for $d_i < 1/2$ the ith element of $\mathbf{v}_t$ is 'asymptotically stationary' (due again to the truncation in the definition of $\mathbf{v}_t$), while $d_i > 1/2$ represents the 'nonstationary' region.

With this extended definition to cover fractional time series we again consider a $p\times 1$ observable $I(d_1,\cdots,d_p)$ time series $\mathbf{y}_t$ satisfying (2.1), partitioning $\mathbf{x}_t$ as before. However we also extend the definition of cointegration, saying that $\mathbf{y}_t$ is cointegrated if at least two d_i are equal and exceed 1/2 and there exists a linear combination giving nonzero weight to two or more of these that is $I(c)$ for $0 \le c < d_i$. Thus, let $a_1 > 1/2$ be the smallest integration order of elements of $\mathbf{x}_{t1}$ and let $\delta \in [0, a_1)$ be the integration order of elements of $\mathbf{x}_{t2}$. Thus, each component of $\mathbf{x}_{t2}$ is a cointegrating error of $\mathbf{y}_t$. Let $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$ and $\mathcal{M}(\mathbf{A}_2)$ be defined as in Section 2. Then $\mathcal{M}(\mathbf{A}_2)$ is called the fractional cointegration space and r is called the fractional cointegration rank. We estimate $\mathcal{M}(\mathbf{A}_2)$ and r in the same manner as in Section 2.

Furthermore, let $r_1,\cdots,r_q$ be q positive integers with $r_1+\cdots+r_q = p-r$, and $1/2 < a_1 < \cdots < a_q$. Suppose that $\mathbf{x}_{t1}$ consists of r_j $I(a_j)$ components. Let

$$\widehat{\mathbf{A}}_1 = (\widehat{\mathbf{A}}_{1q},\cdots,\widehat{\mathbf{A}}_{11}), \tag{4.4}$$

where $\widehat{\mathbf{A}}_{1j}$ has r_j columns. Then $\widehat{\mathbf{x}}_{t1j} = \widehat{\mathbf{A}}_{1j}'\mathbf{y}_t$ is the estimated components of $\mathbf{x}_{t1}$ (i.e., $\mathbf{x}_{t1j} = \mathbf{A}_{1j}'\mathbf{y}_t$) of integration order a_j.

Let $\boldsymbol{\varepsilon}_i = (\varepsilon_i^1,\cdots,\varepsilon_i^p)'$ be the p-dimensional $I(0)$ with mean zero such that $\nabla^{d_j}x_i^j = \varepsilon_i^j + \mu_j$. Let $\mathbf{S}_n(t) = \sum_{i=1}^{[nt]}\boldsymbol{\varepsilon}_i$ and $I_1 = \{i : d_i < 1/2,\ 1 \le i \le p\}$.

Condition 3.

(i) $\mathrm{E}||\boldsymbol{\varepsilon}_t||_2^q < \infty$ for some $q > \max(4, 2/(2a_1-1))$ and for any $i,j \in I_1$, as $n\to\infty$,

$$\frac{1}{n}\sum_{t=1}^{n} x_t^i x_t^j \xrightarrow{p} \mathrm{E}[x_1^i x_1^j].$$

(ii) There exists an i.i.d mean zero $p\times 1$ normal vector $\{\mathbf{w}_i\}$ such that as $n\to\infty$,

$$\max_{0\le t\le 1}||S_n(t) - \sum_{i=1}^{[nt]}\mathbf{w}_i||_2 = o_p(n^{1/s}),\ \text{for some}\ \ s>2.$$

Remark 7. *Condition 3 is mild and satisfied by either of the following processes.*

1. *Suppose $\boldsymbol{\varepsilon}_t$ follows a linear process:*

$$\boldsymbol{\varepsilon}_t = \sum_{k=0}^{\infty}\mathbf{C}_k\mathbf{e}_{t-k},\ t=1,2,\cdots$$

and $\{\mathbf{e}_t\}$ are i.i.d vectors with mean zero, $\mathrm{E}\mathbf{e}_t\mathbf{e}_t' = \Sigma_e > 0$, $\mathrm{E}||\mathbf{e}_t||_2^q < \infty$ for some $q>4$, the $p\times p$ coefficient matrices $\mathbf{C}_k$ satisfy $\sum_{k=0}^{\infty}k||\mathbf{C}_k||^2 < \infty$. Then, by Lemma 2 of Marinucci and Robinson (2000), we have (ii) of Condition 3 holds. (i) follows by ergodicity.

2. *Suppose $\boldsymbol{\varepsilon}_t$ follows a generalized random coefficient autoregressive model:*

$$\boldsymbol{\varepsilon}_t = \mathbf{C}_t \boldsymbol{\varepsilon}_{t-1} + \mathbf{e}_t \tag{4.5}$$

and $\{(\mathbf{C}_t, \mathbf{e}_t)\}$ are i.i.d random variables with $\mathrm{E}||\mathbf{C}_1||_2^q < 1$ and $\mathrm{E}||\mathbf{e}||^q < \infty$ for some $q > 2$, then (ii) of Condition 3 holds with $s < \min\{q, 4\}$, see Corollary 3.4 of Liu and Lin (2009). Similarly, (i) follows by ergodicity.

Theorem 5. *Let r be known. Under Condition 3, $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = o_p(1)$. Furthermore,*

(i) when $\delta < 1/2$,

(a) $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_e(n^{-2a_1+1})$ provided either $|I_0| \geq 2$ or $|I_0| = 1$ and $\mu_{I_0} = 0$;

(b) $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_e(n^{-2a_1})$ provided $|I_0| = 1$, $\mu_{I_0} \neq 0$;

(ii) when $\delta > 1/2$ and $\mu_j = 0$ for $j \geq p - r$, $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_e(n^{-2(a_1-\delta)})$;

(iii) when $\mu_j = 0$ for $j = 1, \cdots, p - r$,

$$D(\mathcal{M}(\widehat{\mathbf{A}}_{1j}), \mathcal{M}(\mathbf{A}_{1j})) = O_e(n^{-2\alpha_j}) \quad \textit{for} \quad j = 1, \cdots, q,$$

where I_0 and α_j are defined as in Theorem 1.

Theorem 6. *Let Condition 3 hold. Then $\lim_{n\to\infty} P(\widehat{r} = r) = 1$, provided $1 \leq r < p$.*

5 Numerical properties

We illustrate the proposed method with 4 simulated examples and one real data set. Note that the comparison with Johansen's (1991) likelihood method is carried out for Example 1 and the real data example only, as Examples 2 concerns different integration orders for different components, Example 3 illustrate the method in the presence of an additional deterministic linear trend, and Example 4 is a model of fractional cointegration. Johansen's method is not applicable to Examples 2-4.

Example 1. Let the first three components of $\mathbf{y}_t$ be the same as Exercise 3.1 in Johansen (1995), i.e.

$$\begin{pmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1/2 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t1} \\ x_{t2} \\ x_{t3} \end{pmatrix} =: \mathbf{A}_{11} \begin{pmatrix} x_{t1} \\ x_{t2} \\ x_{t3} \end{pmatrix},$$

where x_{t1} is an $I(1)$ process, x_{t2}, x_{t3} and the innovations in x_{t1} are independent $N(0, 1)$. For $p > 3$, we add to y_{t1}, y_{t2}, y_{t3} above $r - 2$ extra stationary AR(1) components and $p - r - 1$ ARIMA(1,1,1)

components. All the coefficients in AR(1) are 0.5, the coefficients in ARIMA(1,1,1) are $(0.6, 0.8)$, and all the innovations are independent $N(0,1)$. Except for the elements in $\mathbf{A}_{11}$ specified above, all the other elements of $\mathbf{A}$ are generated independently from $U(-3,3)$. For each setting with different combinations of p, r and n (see Table 1), we draw 500 samples. We set $j_0 = 5$ in (2.3), and estimate the cointegration rank r by (2.5) with $c_0 = 0.3$ for each of the 500 samples. Then with $r = \widehat{r}$, we estimate $\widehat{\mathbf{A}}$ by (2.4). Since $\widehat{r}$ is not necessarily equal to r, and $\mathbf{A}$ is not a half orthogonal matrix (as specified above), we extend the definition of discrepancy measure (3.1) as follows:

$$D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2)) = \Big\{1 - \frac{\mathrm{tr}\big(\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\mathbf{B}_2(\mathbf{B}_2'\mathbf{B}_2)^{-1}\mathbf{B}_2'\big)}{\max(r, \widehat{r})}\Big\}^{1/2}, \tag{5.1}$$

where $\mathbf{B}_2$ is the $p \times r$ matrix consisting of the last r columns of $(\mathbf{A}^{-1})'$, as now $\mathbf{x}_{t2} = \mathbf{B}_2\mathbf{y}_t$. Then $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2)) \in [0,1]$, being 1 if and only if $\mathcal{M}(\widehat{\mathbf{A}}_2)$ and $\mathcal{M}(\mathbf{B}_2)$ are mutually orthogonal, and 0 if and only if the two subspaces are the same. When $\widehat{r} = r$ and $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$, $\mathbf{B}_2 = \mathbf{A}_2$ and $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2)) = D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ defined in (3.1). The relative frequencies (RF) for the occurrence of the event $\{\widehat{r} = r\}$ and the average value of $D1 = D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ over 500 replications are listed in Table 1 under the name new method (New).

Also included in Table 1 are the results of Johansen's likelihood estimation with cointegration rank r estimated by the trace test; see Johansen (1991). We apply the method twice with testing level set at 0.05 and 0.01, respectively, marked as Jo(0.05) and Jo(0.01) in Table 1. The null-distribution of the trace test statistic is approximated by that of

$$\Big[\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_{t-1} - \bar{X})'\Big]\Big[\sum_{t=1}^{T}(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}})(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}})'\Big]^{-1}\Big[\sum_{t=1}^{T}(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}})\boldsymbol{\varepsilon}_t'\Big],$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, \cdots, \varepsilon_{t,p-r})'$, $\boldsymbol{X}_0 = 0$ and $\boldsymbol{X}_t = \sum_{j=1}^{t}\boldsymbol{\varepsilon}_t$, and $\{\varepsilon_{t,i}\}$ are independent $N(0,1)$. See Johansen and Juselius (1990). This approximate distribution is calculated by simulation with $T = 1000$ and 6000 replications.

Table 1 indicates clearly that the newly proposed method always outperforms Johansen's method. More precisely the estimator $\widehat{r}$ defined in (2.5) achieves higher relatively frequencies for hitting the true value r than those achieved by the trace test with significance level at either 0.05 or 0.01. Note that the first part of Table 1 with $p = 3$ and $r = 2$ corresponds to the same setting of Example 3 of Johansen (1995). The inference is more challenging when p and r increase. When $p = 30, r = 10$, our new method works reasonably well when the sample size $n = 1000$ and it works almost perfectly when $n \geq 1500$. On the other hand, Johansen's method, which is not designed for large p, fails to perform even when $n = 2000$ or 2500.

Example 2. Now in model (2.1) let $\mathbf{x}_{t2}$ consist of r stationary AR(1) processes with coefficients $-0.4+i/r$ $(i = 1, \cdots, r)$, and let s components of $\mathbf{x}_{t1}$ be ARIMA(1,1,1) with coefficients $0.3+0.5i/s$

Table 1: Relative frequencies (RF) of correct estimation of r and average distance D_1 defined in (5.1) in simulation with 500 replications for Example 1.

		n=200		n=300		n=500		n=1000		n=1500		n=2000	
	Method	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
p=3	Jo(0.05)	.930	.051	.964	.028	.944	.036	.954	.028	.942	.034	.966	.020
r=2	Jo(0.01)	.980	.026	.996	.011	.990	.011	.992	.007	.994	.005	.984	.010
	New	.968	.032	.998	.009	1.00	.005	1.00	.002	1.00	.002	1.00	.001
p=6	Jo(0.05)	.558	.276	.636	.226	.644	.217	.640	.215	.702	.177	.640	.214
r=2	Jo(0.01)	.760	.184	.856	.117	.802	.123	.862	.083	.852	.088	.866	.079
	New	.388	.375	.838	.117	.982	.027	.994	.013	1.00	.004	1.00	.006
p=9	Jo(0.05)	.200	.445	.216	.422	.312	.367	.344	.345	.352	.337	.380	.323
r=3	Jo(0.01)	.558	.290	.598	.251	.666	.185	.708	.154	.742	.135	.752	.129
	New	.016	.605	.384	.341	.922	.066	.998	.018	1.00	.010	1.00	.006
p=12	Jo(0.05)	.064	.539	.144	.466	.198	.416	.254	.375	.270	.362	.288	.352
r=4	Jo(0.01)	.226	.426	.318	.354	.432	.282	.490	.243	.520	.225	.544	.212
	New	0	.681	.054	.534	.794	.120	.996	.021	1.00	.011	.998	.009
p=18	Jo(0.05)	0	.653	.006	.586	.016	.535	.056	.478	.090	.448	.092	.443
r=6	Jo(0.01)	.006	.595	.020	.522	.046	.468	.158	.379	.226	.349	.236	.333
	New	0	.737	0	.675	.092	.429	.986	.032	1.00	.016	1.00	.011
p=24	Jo(0.05)	0	.742	0	.664	0	.580	.008	.507	.002	.488	.006	.480
r=8	Jo(0.01)	0	.703	0	.613	0	.532	.006	.468	.026	.438	.020	.435
	New	0	.759	0	.719	0	.593	.898	.064	1.00	.022	1.00	.014
p=30	Jo(0.05)	0	.790	0	.732	0	.628	0	.556	0	.527	.002	.512
r=10	Jo(0.01)	0	.772	0	.691	0	.591	.004	.514	.004	.480	.004	.466
	New	0	.771	0	.742	0	.662	.482	.186	.984	.030	1.00	.018

and $0.2 + 0.6i/s\,(i = 1, \cdots, s)$, and the other $p - r - s$ components be ARIMA(0,2,1) with coefficients generated independently from $U(-0.95, 0.95)$. Hence $\mathbf{x}_{t1}$ consists of a mixture of $I(1)$ and $I(2)$ processes. All innovations involved are independent $N(0, 1)$. Let the elements of $\mathbf{A}$ be generated independently from $U(-3, 3)$. We estimate the cointegration rank r by (2.5), and apply the same method to the differenced $\widehat{\mathbf{x}}_{t1}$ to estimate s; see Section 2.4 above. For each setting, we replicate the exercise 500 times. The relative frequencies for the occurrence of events $\{\widehat{r} = r\}$ and $\{\widehat{s} = s\}$ are listed in Table 2.

Also included in Table 2 are the results from applying the Phillips-Perron unit-root test (PP.test), with significance level set at 0.01, for estimating r; see (2.7). By applying the same procedure to the differenced $\widehat{\mathbf{x}}_{t1}$, we also obtain the estimated s. When p is small, the PP.test estimates r slightly better than (2.5) though both methods perform well. For estimating s, the PP.test is much worse than (2.5). When p is large, (2.5) performs substantially better than the PP.test. Also noticeable in Table 2 is the fact that the larger r/p is, the more accurate are the estimates for r, and the larger $s/(p - r)$ is, the more accurate are the estimates for s. Overall (2.5) provides more a stable performance than PP.test.

Figs 1–2 present the boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ and $D_1(\mathcal{M}(\widehat{\mathbf{A}}_{11}),$

Table 2: Relative frequencies of correct estimation of r and s by the Phillips-Perron test (PP.test) and method (2.5) in simulation with 500 replications for Example 2.

(p, r, s)	n	200		300		500		1000		1500		2000	
	Method	r	s	r	s	r	s	r	s	r	s	r	s
(6, 2, 2)	PP.test	.964	.412	.970	.440	.978	.420	.982	.416	.970	.448	.960	.460
	(2.5)	.614	.486	.908	.766	.962	.814	.944	.876	.942	.892	.924	.898
(6, 3, 1)	PP.test	.996	.288	1.00	.336	.996	.342	.992	.408	.998	.416	.998	.430
	(2.5)	.904	.604	.992	.782	.998	.896	.986	.924	.992	.940	.988	.958
(10, 4, 4)	PP.test	.840	.348	.874	.392	.854	.392	.852	.446	.842	.430	.824	.454
	(2.5)	.078	.162	.538	.480	.924	.798	.940	.866	.896	.858	.880	.870
(10, 6, 2)	PP.test	.984	.262	.986	.276	.978	.330	.984	.322	.978	.404	.974	.406
	(2.5)	.566	.488	.932	.740	.954	.826	.942	.874	.920	.876	.910	.884
(15, 8, 4)	PP.test	.780	.192	.792	.174	.812	.218	.750	.232	.726	.260	.658	.310
	(2.5)	.006	.110	.326	.372	.868	.684	.836	.708	.858	.770	.830	.768

$\mathcal{M}(\mathbf{B}_{11}))$ for $(p, r, s) = (6, 2, 2)$ and $(10, 4, 4)$ respectively, where $\mathcal{M}(\mathbf{B}_{11})$ is the true cointegration space specified by the $I(1)$ components of $\mathbf{x}_{t1}$. As expected, the estimation errors decrease as sample size n increases, and the errors with $(p, r, s) = (10, 4, 4)$ are greater than those with $(p, r, s) = (6, 2, 2)$.

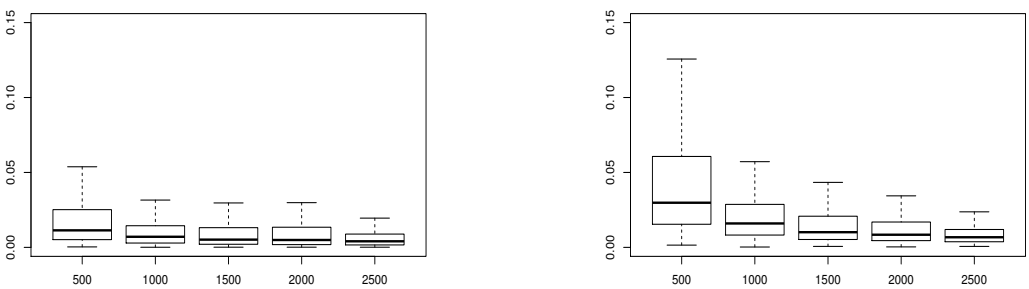

Figure 1: Example 2: Boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ (left panel) and $D_1(\mathcal{M}(\widehat{\mathbf{A}}_{11}), \mathcal{M}(\mathbf{B}_{11}))$ (right panel) when $(p, r, s) = (6, 2, 2)$. The labels on the horizontal axis are sample size n.

Example 3. Now we consider an example in which the components of $\mathbf{y}_t$ are $I(1)$ with linear trend, i.e.,

$$\mathbf{y}_t = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 t + \boldsymbol{Z}_t = \mathbf{A}\mathbf{x}_t^* \tag{5.2}$$

for some $(\mathbf{x}_t^*)' = (\mathbf{x}_{t1}^*, \mathbf{x}_{t2})$, where $\mathbf{x}_{t1}^* = \boldsymbol{\mu}_1^* + \boldsymbol{\mu}_2^* t + \mathbf{x}_{t1}$, $\mathbf{x}_{t1}$ is nonstationary process and $\mathbf{x}_{t2}$ is stationary process. In our simulation, all component of $\boldsymbol{\mu}_1^*$ and $\boldsymbol{\mu}_2^*$ are taken as 0.3 and 0.5 respectively, all components of $\mathbf{x}_{t2}$ are AR(1) with coefficients generated from $U(-0.8, 0.8)$, all components of $\mathbf{x}_{t1}$ are ARIMA(1,1,1) with AR coefficients generated from $U(0, 0.8)$ and MA coefficients generated from $U(0, 0.95)$, and all innovations are independent $N(0, 1)$. Table 3 reports

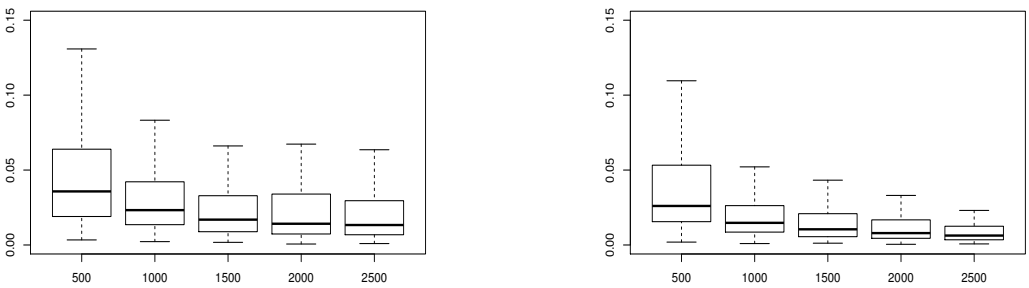

Figure 2: Example 2: Boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ (left panel) and $D_1(\mathcal{M}(\widehat{\mathbf{A}}_{11}), \mathcal{M}(\mathbf{B}_{11}))$ (right panel) when $(p, r, s) = (10, 4, 4)$. The labels on the horizontal axis are sample size n.

the relative frequencies of the occurrence of the event $\{\widehat{r} = r\}$ and the average distance (5.1) in a simulation with 500 replications, where the cointegration rank is estimated by (2.5) with $c_0 = 0.3$. Also included in Table 3 are the results obtained from applying the Phillips-Perron unit-root test to estimate r, see (2.7). Table 3 indicates that (2.5) works well even in the presence of a deterministic linear trend, where our theoretical setting exclude. However the Phillips-Perron test performs poorly for large p and small r/p.

Table 3: Relative frequencies of correct estimation of r and average distance in simulation with 500 replications for Example 3.

(p,r)	Method	n=200		n=300		n=500		n=1000		n=1500		n=2000	
		RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
(6, 2)	PP.test	.882	.087	.780	.143	.664	.200	.950	.030	.838	.096	.746	.150
	New	.452	.331	.858	.107	.982	.026	1.00	.002	1.00	.004	.998	.008
(6, 4)	PP.test	.988	.010	.988	.007	1.00	.002	.998	.002	.996	.002	.990	.005
	New	.974	.016	1.00	.002	1.00	.002	1.00	.001	1.00	.002	1.00	5e-4
(10, 4)	PP.test	.842	.092	.398	.293	.624	.182	.330	.324	.334	.319	.488	.244
	New	.066	.485	.328	.327	.966	.042	1.00	.021	1.00	.007	1.00	.012
(10, 6)	PP.test	.766	.107	.316	.279	.664	.132	.846	.062	.806	.076	.876	.048
	New	.432	.231	.796	.103	.998	.010	1.00	.006	1.00	.003	1.00	.002
(15, 6)	PP.test	.082	.454	.166	.377	.094	.424	.142	.388	.046	.468	.096	.436
	New	0	.651	.004	.521	.506	.221	.996	.021	.998	.021	1.00	.005
(15, 10)	PP.test	.290	.240	.592	.137	.336	.217	.484	.157	.798	.064	.446	.177
	New	.066	.332	.646	.124	.964	.034	1.00	.003	1.00	.004	1.00	.007
(30, 10)	PP.test	0	.628	0	.667	0	.671	0	.686	0	.696	0	.703
	New	0	.769	0	.737	0	.655	.364	.234	.974	.062	.994	.040
(30, 20)	PP.test	0	.346	.004	.329	.010	.324	.010	.314	.034	.294	.006	.329
	New	0	.456	.002	.368	.344	.168	1.00	.019	1.00	.010	1.00	.010

Example 4. We consider fractional cointegration cases now. Let the components of $\mathbf{x}_{t1}$ be

$I(d)$ processes with a fractional order $d = 4/5$ or $3/4$, the components of $\mathbf{x}_{t2}$ be AR(1) with autoregressive coefficients $0.2i$ $(i = 1, \cdots, r)$, the elements of $\mathbf{A}$ be generated independently from $U(-3, 3)$, and all innovations be independent and $N(0, 1)$. We consider various combinations for p, r, s, and the sample size n. For each setting, we replicate the simulation 500 times and estimate the cointegration rank r using (2.5) with $c_0 = 0.3$. The relative frequencies for the occurrence of the event $\{\widehat{r} = r\}$ and the mean of distance (5.1) over 500 replications are listed in Table 4. While the proposed methodology works well, the accuracy is slightly lower than that integer cointegration orders. See the examples above. We also notice that the estimation errors with $d = 3/4$ are greater than those with $d = 4/5$.

To illustrate the impact of the choice of j_0 on the estimation, we consider the above fractional cointegration with $p = 6$, $r = 4$ and order $d = 4/5$, $3/4$ and $2/3$. By setting sample size $n = 1000$ and j_0 between 5 and 100, the relative frequencies for the occurrence of the event $\{\widehat{r} = r\}$ and the mean of the distance (5.1) are reported in Table 5. As mentioned in Section 2, using different values of j_0 hardly changes the results.

Table 4: Relative frequencies (RF) of the occurrence of event $\{\widehat{r} = r\}$ and average distance D_1 defined in (5.1) in simulation with 500 replications for Example 4.

d	$(p,\ r)$	n=200		n=300		n=500		n=1000		n=1500		n=2000	
		RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
4/5	(3, 2)	.828	.134	.948	.068	.978	.040	1.00	.017	.998	.014	1.00	.010
	(6, 2)	.020	.664	.240	.507	.664	.294	.946	.119	.966	.101	.986	.070
	(9, 3)	0	.721	.004	.656	.188	.488	.766	.250	.868	.181	.920	.156
	(12, 4)	0	.743	0	0.701	.014	.596	.528	.380	.716	.307	.788	.275
3/4	(3, 2)	.770	.174	.902	.098	.964	.058	.984	.033	1.00	.019	.998	.017
	(6, 2)	.018	.685	.132	.578	.488	.380	.866	.193	.916	.151	.942	.118
	(9, 3)	0	.733	0	.680	.104	.549	.604	.336	.800	.240	.864	.205
	(12, 4)	0	.754	0	.719	.006	.629	.328	.450	.606	.378	.696	.344

Table 5: Relative frequencies (RF) of the occurrence of event $\{\widehat{r} = r\}$ and average distance D_1 defined in (5.1) with $n = 1000$ in simulation with 500 replications for Example 4.

d	j_0=5		j_0=10		j_0=15		j_0=20		j_0=50		j_0=100	
	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
4/5	.964	.086	.984	.069	.982	.062	.982	.064	.982	.054	.980	.057
3/4	.934	.125	.950	.107	.952	.101	.956	.091	.954	.082	.960	.084
2/3	.788	.226	.788	.209	.788	.199	.804	.195	.814	.171	.806	.179

Example 5. We consider the 8 monthly US Industrial Production indices for January 1947 – December 1993 published by the US Federal Reserve, namely *the total index, manufacturing index, durable manufacturing, nondurable manufacturing, mining, utilities, products* and *materials*. The original 8 time series are plotted in Fig.3. Applying the proposed method to these data, the transformed series $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ are plotted in Fig.4 together with their sample ACF. The proposed

method (2.5) leads to $\widehat{r} = 4$ with $m = 40$, $c = 0.3$ and $j_0 = 50$ or 100.

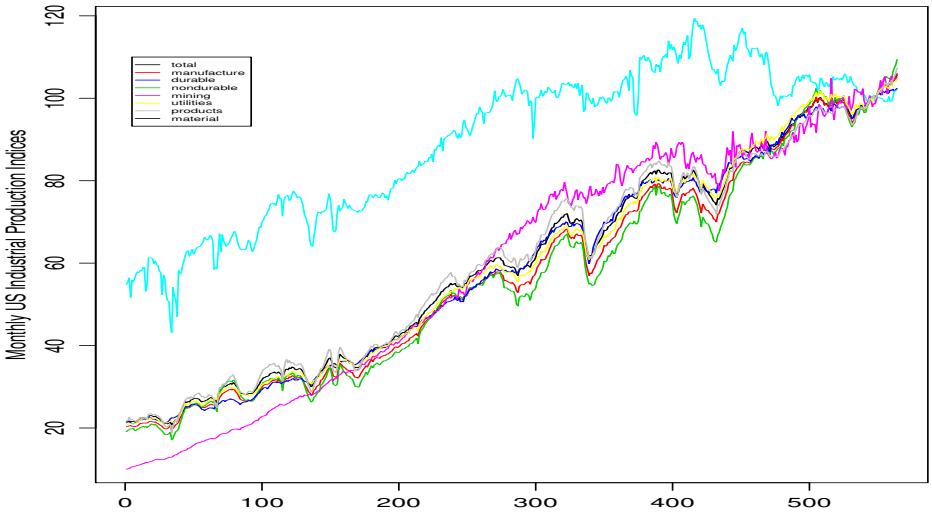

Figure 3: Time series plots of the 8 monthly U.S. Industrial Production indices in January 1947 - December 1993.

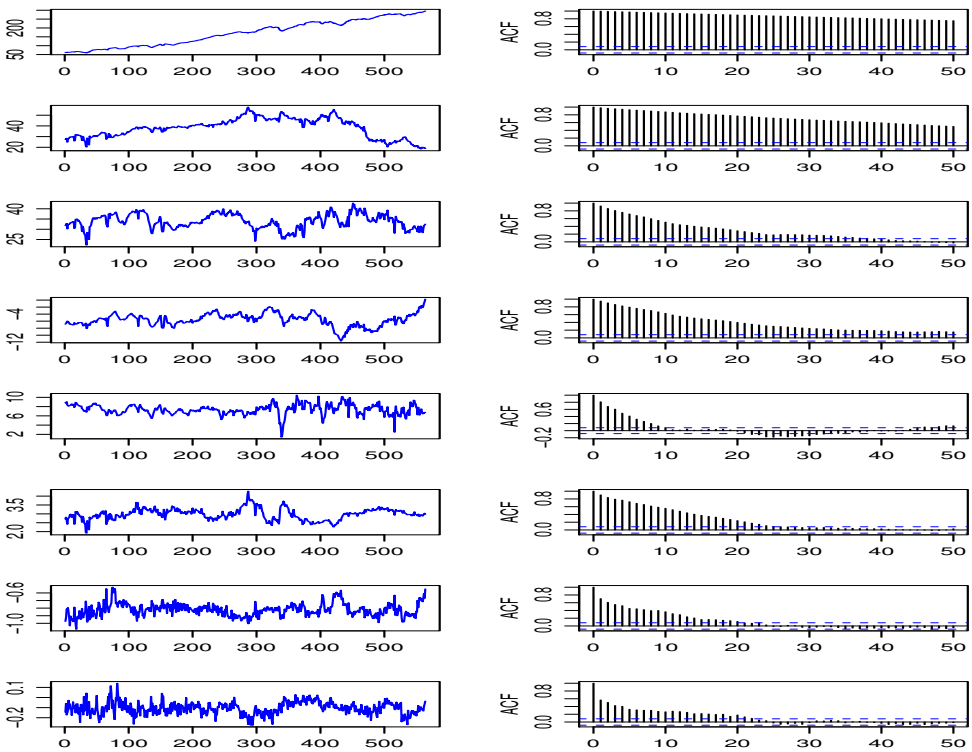

Figure 4: Time series plots of the estimated $\widehat{\mathbf{x}}_t$ by the proposed method and their sample ACF for the 8 monthly U.S. Industrial Production indices.

We also apply Johansen's (1991) likelihood method to this data set. Both the trace and the maximum tests indicate $r = 4$. The corresponding transformed series together with their sample

ACF are plotted in Fig.5.

Let $\widehat{\mathbf{A}}_2$ denote the last 4 columns of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}_2$ consist of the loadings for the last 4 component series displayed in Fig.5, i.e., the columns of $\widehat{\mathbf{A}}_2$ are the loadings of the 4 cointegrated variables identified by the proposed method in this paper, and the columns of $\widehat{\mathbf{B}}_2$ are the loadings of the 4 cointegrated variables identified by Johansen's likelihood method. Then

$$D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\widehat{\mathbf{B}}_2))^2 = 1 - \frac{1}{4}\mathrm{tr}\{\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{B}}_2(\widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2)^{-1}\widehat{\mathbf{B}}_2'\} = 1 - 0.9816 = 0.0184.$$

This indicates that the two sets of cointegrated variables identified by the two methods are effectively equivalent.

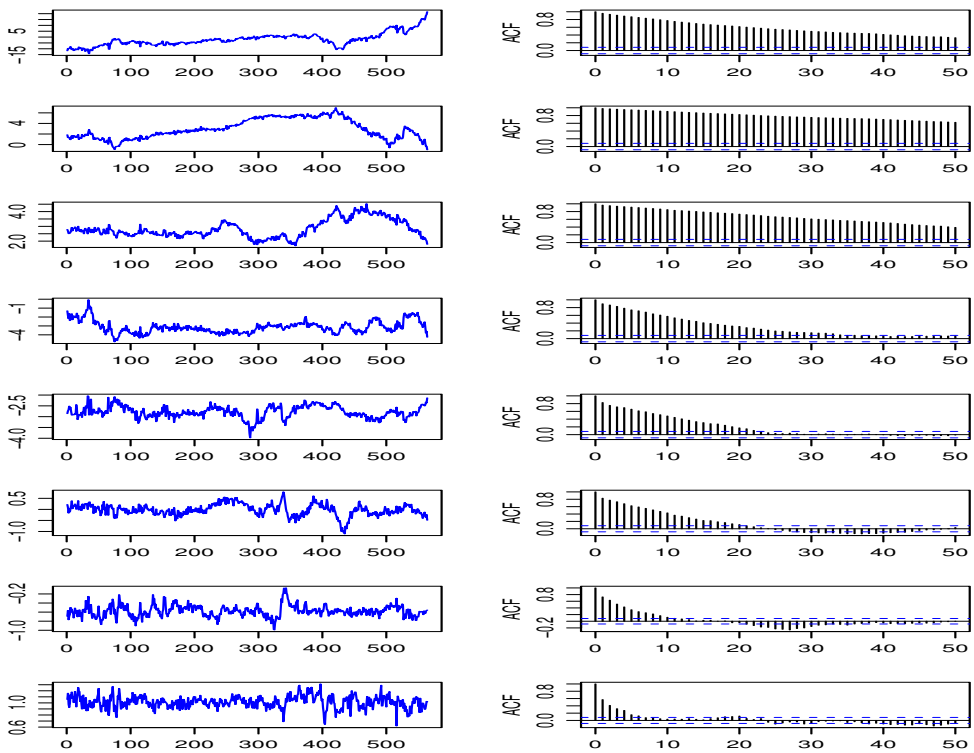

Figure 5: Time series plots of the estimated $\widehat{\mathbf{x}}_t$ by Johansen's method and their sample ACF for the 8 monthly U.S. Industrial Production indices.

To illustrate the impact of the choice of c_0 on the estimation, we consider model (2.1) with $p = 2$ and the following three specifications for $\mathbf{x}_t$:

(i) $r = 0$, both components of $\mathbf{x}_t$ are ARIMA(1, 1, 1) processes with coefficient $(0.6, 0.4)$ and $(0.8, 0)$,

(ii) $r = 1$, x_{t1} is ARIMA(1, 1, 1) with $(0.6, 0.4)$ and x_{t2} is AR(1) with coefficient 0.6,

(iii) $r = 2$, x_{t1} is AR(1) with coefficient 0.6 and x_{t2} is ARMA(1, 1) with coefficient $(0.6, 0.4)$.

The elements of $\mathbf{A}$ are generated independently from $U(-3,3)$ and c_0 is taken from 0.05 to 0.95. In each setting, we replicate the simulation 500 times with sample size $n = 200, 300, 500$ and 1000. The relative frequencies for the occurrence of the event $\{\widehat{r} = r\}$ are reported in Table 6. When $r = 0$, smaller c_0 would lead to better performance, however when $r = 2$, larger c_0 may result in better performance. It is because that as $r = 0$, both the components are I(1), smaller c_0 tends to estimate r as 0, while as $r = 2$, both the components are I(0), larger c_0 tends to estimate r as 2, see Remark 1. Further, it is shown that when c_0 is taken away from the endpoints, say $c_0 \in (0.2, 0.5)$, then the proposed procedure works well for all cases, especially when n is large.

Table 7 reports the simulation results with $p = 3$, $\mathbf{A}$ generated in the same manner as the above, and three settings for $\mathbf{x}_t$:

(i) $r = 0$, the components of $\mathbf{x}_t$ are all ARIMA(1, 1, 1) with coefficients $(0.6, 0)$, $(0.3, 0.7)$ and $(0.8, 0.4)$,

(ii) $r = 1$, x_{t1} and x_{t2} are both ARIMA(1, 1, 1) with coefficients $(0.5, 0)$, $(0.8, 0.4)$, and x_{t3} is AR(1) with coefficient 0.6,

(iii) $r = 2$, x_{t1} is ARIMA(1, 1, 1) with coefficient $(0.8, 0.4)$, x_{t2} is AR(1) with coefficient 0.6 and x_{t3} is ARMA(1, 1) with coefficient $(0.5, 0.5)$.

The pattern of Table 7 is very similar to that of Table 6, i.e. the estimation is stable for $c_0 \in (0.2, 0.5)$.

Table 6: Relative frequencies (RF) of the occurrence of event $\{\widehat{r} = r\}$ for $p = 2$ with different c_0 and 500 replications.

r	n	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
0	200	1.00	1.00	1.00	.996	.988	.976	.946	.926	.872	.814	.722	.616	.480	.350	.218	.112	.036	.002	0
	300	1.00	1.00	1.00	1.00	1.00	1.00	.998	.996	.984	.958	.932	.880	.796	.672	.522	.364	.180	.034	0
	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.990	.980	.946	.868	.754	.546	.232	.016
	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.988	.922	.696	.226
1	200	.648	.878	.966	.992	.992	.994	.988	.976	.968	.952	.936	.902	.838	.766	.642	.538	.416	.264	.074
	300	.514	.864	.984	.998	1.00	1.00	1.00	1.00	.998	.992	.984	.974	.952	.930	.858	.752	.620	.398	.152
	500	.442	.882	.990	.998	.998	1.00	.998	.998	.998	.996	.994	.994	.988	.982	.952	.910	.844	.636	.336
	1000	.210	.900	.992	.998	.998	.998	.998	.998	.998	.998	.998	.996	.996	.996	.994	.990	.980	.910	.634
2	200	.230	.616	.866	.970	.994	.998	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	300	.150	.568	.884	.980	.990	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	500	.088	.502	.910	.992	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	.018	.558	.978	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

6 Conclusions

We propose in this paper a simple, direct and model-free method for identifying cointegration relationships among multiple time series of which different components series may have different integration orders. The method boils down to an eigenanalysis for a non-negative definite matrix. One may view that the components of the transformed series $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ are arranged in ascending

Table 7: Relative frequencies (RF) of the occurrence of event $\{\widehat{r} = r\}$ for $p = 3$ with different c_0 and 500 replications.

r	n	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
0	200	.986	.982	.962	.938	.896	.826	.744	.608	.490	.370	.276	.178	.096	.036	.008	.002	0	0	0
	300	1.00	1.00	.998	.998	.994	.982	.962	.926	.866	.806	.702	.562	.390	.238	.118	.040	.006	0	0
	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.986	.972	.940	.864	.738	.526	.316	.112	.008	0
	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.978	.876	.692	.282	.004
1	200	.732	.932	.958	.968	.954	.932	.908	.860	.788	.718	.632	.530	.430	.308	.188	.096	.032	0	0
	300	.584	.900	.986	1.00	1.00	.996	.988	.974	.956	.924	.880	.816	.724	.600	.418	.268	.114	.022	0
	500	.456	.896	.984	.994	.998	1.00	1.00	.994	.990	.988	.988	.972	.944	.908	.824	.676	.446	.198	.032
	1000	.258	.900	.996	.996	.998	.998	1.00	.998	.998	.996	.996	.996	.994	.990	.990	.962	.884	.626	.194
2	200	.288	.780	.964	.998	1.00	1.00	1.00	1.00	1.00	1.00	.998	.990	.962	.942	.886	.828	.724	.522	.252
	300	.448	.814	.944	.990	.998	.998	.998	.994	.992	.982	.962	.934	.878	.820	.756	.666	.500	.322	.126
	500	.210	.786	.982	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.978	.950	.892	.726	.420
	1000	.096	.848	.996	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.996	.960	.714

order according to the "degree" of stationarity; reflected by the magnitude of the eigenvalues of $\widehat{\mathbf{W}}$.

7 Appendix: Technical proofs

7.1 Proof for Section 3.1

Let

$$\begin{aligned}\mathbf{\Sigma}_j^x &= \operatorname{diag}\left[\left(\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\overline{\mathbf{x}}_1)(\mathbf{x}_{t1}-\overline{\mathbf{x}}_1)'\right),\left(\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2}-\overline{\mathbf{x}}_2)(\mathbf{x}_{t2}-\overline{\mathbf{x}}_2)'\right)\right]\\ &\equiv \operatorname{diag}(\mathbf{\Sigma}_{j1}^x,\mathbf{\Sigma}_{j2}^x),\end{aligned}$$

$\mathbf{W}^x = \sum_{j=0}^{j_0}\mathbf{\Sigma}_j^x(\mathbf{\Sigma}_j^x)' =: \operatorname{diag}(\mathbf{D}_1^x,\mathbf{D}_2^x)$ and $\mathbf{\Gamma}_x$ be the $p\times p$ orthogonal matrix such that

$$\mathbf{W}^x\mathbf{\Gamma}_x = \mathbf{\Gamma}_x\mathbf{\Lambda}_x,$$

where $\mathbf{\Lambda}_x$ is the diagonal matrix of eigenvalues of $\mathbf{W}^x$. Since $\mathbf{x}_{t1}$ is nonstationary and $\mathbf{x}_{t2}$ is stationary, intuitively $\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,1}-\overline{\mathbf{x}}_1)\ (\mathbf{x}_{t1}-\overline{\mathbf{x}}_1)'$ and $\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2}-\overline{\mathbf{x}}_2)(\mathbf{x}_{t2}-\overline{\mathbf{x}}_2)'$ do not share the same eigenvalues, so $\mathbf{\Gamma}_x$ must be block-diagonal. Define $\mathbf{W}^y = \mathbf{A}\mathbf{W}^x\mathbf{A}'$, then

$$\mathbf{W}^y = \mathbf{A}\mathbf{W}^x\mathbf{A}' = \mathbf{A}\mathbf{\Gamma}_x\mathbf{\Lambda}_x\mathbf{\Gamma}_x'\mathbf{A}'.$$

This implies that the columns of $\mathbf{A}\mathbf{\Gamma}_x$ are just the orthogonal eigenvectors of $\mathbf{W}^y$. Since $\mathbf{\Gamma}_x$ is block-diagonal, it follows that $\mathcal{M}(\mathbf{A}_2)$ is the same as the space spanned by the eigenvectors corresponding to the smallest r eigenvalues of $\mathbf{W}^y$. As a result, to show the distance between the cointegration space and its estimate is small, we only need to show that the space spanned by the eigenvectors of $\mathbf{W}^y$ can be approximated by that of $\widehat{\mathbf{W}}$. This question is usually solved by perturbation matrix theory. In particular, let

$$\widehat{\mathbf{W}} = \mathbf{W}^y + \Delta\mathbf{W}^y,\ \ \Delta\mathbf{W}^y = \widehat{\mathbf{W}} - \mathbf{W}^y,$$

and

$$\mathrm{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x) = \min_{\lambda\in\lambda(\mathbf{D}_1^x),\, \mu\in\lambda(\mathbf{D}_2^x)} |\lambda - \mu|,$$

where $\lambda(\mathbf{A})$ denotes the set of eigenvalues of a matrix $\mathbf{A}$. When $||\Delta\mathbf{W}^y|| = o_p(\mathrm{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x))$, one can use the perturbation results of Golub and Loan (1996) to establish the bound of Theorems 1, 3 and 5, see also Lam and Yao (2012) or Chang, Guo and Yao (2017). However, in our setting $\mathrm{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x)$ can be of smaller order than $||\Delta\mathbf{W}^y||$, i.e., $\mathrm{sep}(\mathbf{D}_1^x, \mathbf{D}_2^x)/||\Delta\mathbf{W}^y|| \xrightarrow{p} 0$ as $n \to \infty$ and the above method will not work.

To fix this problem, we adopt the perturbation results of Dopico, Moro and Molera (2000) instead. A similar idea was used by Chen and Hurvich (2006) to recover their fractional cointegration spaces via the periodogram matrix, using a random diagonal block matrix instead. However, because of the quadratic form of $\mathbf{W}^x$ $(= \sum_{j=1}^{j_0} \mathbf{\Sigma}_j^x (\mathbf{\Sigma}_j^x)')$, we cannot find a normalizing constant matrix $\mathbf{C}_n$ such that $\mathbf{C}_n\mathbf{W}^x\mathbf{C}_n = O_e(1)$ or $\mathbf{C}_n\mathbf{W}^y\mathbf{C}_n = O_e(1)$, so as a result, the argument of Chen and Hurvich (2006) based on the perturbation bound of Barlow and Slapnicar (2002) cannot be used. To this end, we first establish some lemmas (i.e. Lemmas 7-10 below) and we legate their proofs to supplementary material.

For $1 \le i \le p-r$, set $f_0^i(t) = W^i(t)$, $f_{d_i}^i(t) = \int_0^t f_{d_i-1}^i(s)\,dt$, $\mu_i = \mathrm{E}z_t^i$ and define

$$F^i(t) \;=\; f_{d_i}^i(t) - \int_0^1 f_{d_i}^i(t)\,dt, \quad G_d(t) = \frac{\prod_{j=0}^{d-1}(t+j)}{d!}, \quad \bar{G}_d = \frac{1}{n}\sum_{t=1}^n G_d(t).$$

Then, we have the following weak convergence result for the sample autocovariance.

Lemma 7. *Let $L_d(t) = G_d(t) - \bar{G}_d$. Suppose $x_t^i \sim I(d_i)$, $1 \le i \le p-r$, then under Condition 1,*

$$\Big(\frac{x_t^i - \bar{x}^i - \mu_i L_{d_i}(t)}{n^{d_i-1/2}},\, 1 \le i \le p-r\Big) \xrightarrow{d} \Big(F^i(t),\, 1 \le i \le p-r\Big) \text{ and} \tag{7.1}$$

$$\Big(\frac{1}{n^{d_i+1/2}}\sum_{t=1}^n (x_t^i - \bar{x}^i - \mu_i L_{d_i}(t))(x_t^j - \mathrm{E}x_t^j), i \le p-r, p-r+1 \le j \le p\Big) \xrightarrow{p} \mathbf{0}. \tag{7.2}$$

Next, we establish a bound for the eigenvalues of $\mathbf{\Sigma}_j^x$ and $\mathbf{A}'\widehat{\mathbf{\Sigma}}_j\mathbf{A} =: \widehat{\mathbf{\Sigma}}_j^x$.

Without loss of generality, we assume the first r_1 components of $\mathbf{x}_{t1}$ are $I(a_1)$, the next r_2 components are $I(a_2)$ and the last r_q components of $\mathbf{x}_{t1}$ are $I(a_q)$, that is,

$$\boldsymbol{x}_{t1} = (\overbrace{x_t^1, \cdots, x_t^{r_q}}^{I(a_q)}, \overbrace{x_t^{r_q+1}, \cdots, x_t^{r_q+r_{q-1}}}^{I(a_{q-1})}, \cdots, \overbrace{x_t^{\sum_{j=2}^q r_j+1}, \cdots, x_t^{\sum_{j=1}^q r_j}}^{I(a_1)})',$$

where $a_1 < a_2 < \cdots < a_q$ are positive integers and $\sum_{i=1}^q r_i = p-r$. For $1 \le i \le q$, define $\nu_q = 0$ and $\nu_i = \sum_{j=i+1}^q r_j$. Then for any $\mathbf{x}_t(r_i) := (x_t^{\nu_i+1}, \cdots, x_t^{\nu_i+r_i})'$, if $\boldsymbol{\mu}_i := (\mu_{\nu_i+1}, \cdots, \mu_{\nu_i+r_i})' \neq 0$, there must exist a $r_i \times (r_i-1)$ matrix $\mathbf{P}_i$ and $r_i \times 1$ vector $\bar{\boldsymbol{\mu}}_i$ such that $\mathbf{P}_i'\mathbf{P}_i = \mathbf{I}_{(r_i-1)}$, $(\mathbf{P}_i, \boldsymbol{\mu}_i)$

has full rank r_i, $\mathbf{P}_i'\boldsymbol{\mu}_i = 0$ and $\bar{\boldsymbol{\mu}}_i'\boldsymbol{\mu}_i = 1$, where $\mathbf{I}_a$ denotes $a \times a$ matrix. Let $\mathbf{B}_i = (\mathbf{P}_i, n^{-1/2}\bar{\boldsymbol{\mu}}_i)'$ if $\boldsymbol{\mu}_i \neq 0$ and $\mathbf{B}_i = \mathbf{I}_{r_i}$ if $\boldsymbol{\mu}_i = 0$, and $\boldsymbol{\Theta}_n = \mathrm{diag}(\mathbf{B}_q, \cdots, \mathbf{B}_2, \mathbf{B}_1, \mathbf{I}_r)$. Define

$$\mathbf{D}_{n1} = \mathrm{diag}\Big(\overbrace{n^{a_q-1/2}, \cdots, n^{a_q-1/2}}^{r_q}, \cdots, \overbrace{n^{a_1-1/2}, \cdots, n^{a_1-1/2}}^{r_1}\Big),\ \mathbf{D}_{n2} = (\overbrace{1, \cdots, 1}^{r}),$$

and $\mathbf{D}_n =: \mathrm{diag}(\mathbf{D}_{n1}, \mathbf{D}_{n2})$. Let $H^d(t) = t^d/d! - 1/(d+1)!$, $F^i(t)$ be given as in Lemma 7, $\mathbf{F}_i(t) = (F^{\nu_i+1}(t), \cdots, F^{\nu_i+r_i}(t))'$, $\mathbf{M}_i(t) = (\mathbf{F}_i'(t)\mathbf{P}_i, H^{a_i}(t))' I(\boldsymbol{\mu}_i \neq 0) + \mathbf{F}_i(t) I(\boldsymbol{\mu}_i = 0)$, and $\mathbf{M}(t) = (\mathbf{M}_q'(t), \mathbf{M}_{q-1}'(t), \cdots, \mathbf{M}_1'(t))'$. Then Lemma 8 below follows from Lemma 7 and the continuous mapping theorem.

Lemma 8. *Let* $\boldsymbol{\Gamma}_j(x) = diag\Big(\frac{1}{n}\sum_{t=1}^n (\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)(\mathbf{x}_{t1} - \bar{\mathbf{x}}_1)',\ \mathrm{Cov}(\mathbf{x}_{1+j,2},\ \mathbf{x}_{1,2})\Big)$. *Under Condition 1, we have*

$$\mathbf{D}_n^{-1}\boldsymbol{\Theta}_n\boldsymbol{\Gamma}_j^x\boldsymbol{\Theta}_n'\mathbf{D}_n^{-1} \xrightarrow{d} \mathrm{diag}\Big(\int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt,\ \mathrm{Cov}(\mathbf{x}_{1+j,2},\ \mathbf{x}_{1,2})\Big).$$

Let $F^i(t)$, $1 \le i \le p-r$ be defined in Lemma 7, where $W^i(t) = \sigma_{ii}B^i(t)$ and $B^i(t)$, $1 \le i \le p-r$ are independent Brownian motions. Let $\mathbf{F}(t) = (F^1(t), F^2(t), \cdots, F^{p-r}(t))'$.

Lemma 9. *Under condition 2 and* $p = o(n^{1/2-\tau})$ *with* $0 < \tau < 1/2$,

$$\Big\|\mathbf{D}_n^{-1}\boldsymbol{\Gamma}_j^x\mathbf{D}_n^{-1} - \mathrm{diag}\Big(\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\,dt,\ \mathrm{Cov}(\mathbf{x}_{1+j,2},\ \mathbf{x}_{1,2})\Big)\Big\|_2 = o_p(1). \tag{7.3}$$

Further, $\int_0^1 \mathbf{F}(t)\mathbf{F}'(t)\,dt$ *is positive definite.*

Lemma 10. *Under Condition 1, or Condition 2 and* $p = o(n^{1/2-\tau})$, *we have*

$$\max_{0\le j\le j_0} \|\mathbf{D}_n^{-1}\boldsymbol{\Theta}_n(\boldsymbol{\Sigma}_j^x - \boldsymbol{\Gamma}_j^x)\boldsymbol{\Theta}_n'\mathbf{D}_n^{-1}\|_2 \xrightarrow{p} 0 \text{ and} \tag{7.4}$$

$$\max_{0\le j\le j_0} \|\mathbf{D}_n^{-1}\boldsymbol{\Theta}_n(\widehat{\boldsymbol{\Sigma}}_j^x - \boldsymbol{\Gamma}_j^x)\boldsymbol{\Theta}_n'\mathbf{D}_n^{-1}\|_2 \xrightarrow{p} 0. \tag{7.5}$$

Proof of Theorem 1. Since

$$\begin{aligned}\{D(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{A}_2))\}^2 &= \frac{1}{r}\{\mathrm{tr}[\mathbf{A}_2'(I_p - \widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2')\mathbf{A}_2]\} \\ &\le ||\mathbf{A}_2'(\mathbf{A}_2\mathbf{A}_2' - \widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2')\mathbf{A}_2||_2 \le 2||\widehat{\mathbf{A}}_2 - \mathbf{A}_2||_2^2,\end{aligned}$$

it follows from Theorem I.5.5 of Stewart and Sun (1990) (see also Proposition 2.1 of Vu and Lei (2013)) that

$$D(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{A}_2)) \le \sqrt{2}||\widehat{\mathbf{A}}_2 - \mathbf{A}_2||_2 \le \sqrt{2}||\widehat{\mathbf{A}}_2 - \mathbf{A}_2||_F \le 2\sqrt{2}||\sin\Theta(\widehat{\mathbf{A}}_2, \mathbf{A}_2)||_F, \tag{7.6}$$

where $\Theta(\widehat{\mathbf{A}}_2, \mathbf{A}_2) = \arccos[(\mathbf{A}_2'\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2\mathbf{A}_2)^{1/2}]$ is the canonical angle between the column spaces of $\widehat{\mathbf{A}}_2$ and $\mathbf{A}_2$. Let $\eta = \min_{\lambda\in\lambda(\mathbf{D}_1^x),\,\mu\in\lambda(\widetilde{\mathbf{D}}_2^x)} |\lambda - \mu|/\sqrt{\lambda\mu}$, where $\lambda(\widetilde{\mathbf{D}}_2^x)$ consists of the r smallest eigenvalues of $\mathbf{A}'\widehat{\mathbf{W}}\mathbf{A} =: \widehat{\mathbf{W}}^x$. By Theorem 2.4 of Dopico, Moro and Molera (2000), we have

$$||\sin\Theta(\widehat{\mathbf{A}}_2, \mathbf{A}_2)||_F \le ||(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2}||_F/\eta. \tag{7.7}$$

Note that

$$(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2} = (\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2} - (\mathbf{W}^y)^{1/2}(\widehat{\mathbf{W}})^{-1/2}. \tag{7.8}$$

Thus, by equations (7.6), (7.7) and (7.8), we have

$$D(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{A}_2)) \leq (||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F + ||(\mathbf{W}^y)^{1/2}(\widehat{\mathbf{W}})^{-1/2}||_F)/\eta.$$

Next, we show that$||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F = O_p(1)$, which is equivalent to

$$||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F = O_p(1). \tag{7.9}$$

Note that

$$0 \leq \mathbf{\Sigma}_0^x \leq (\mathbf{W}^x)^{1/2} \leq \sum_{j=0}^{j_0}\{\mathbf{\Sigma}_j^x(\mathbf{\Sigma}_j^x)'\}^{1/2} \text{ and } 0 \leq \widehat{\mathbf{\Sigma}}_0^x \leq (\widehat{\mathbf{W}}^x)^{1/2} \leq \sum_{j=0}^{j_0}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}. \tag{7.10}$$

It follows from (7.10) that

$$||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F \leq \sum_{j=0}^{j_0}||(\mathbf{\Sigma}_0^x)^{-1}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}||_F.$$

Thus, for (7.9), it is enough to show the eigenvalues of $(\mathbf{\Sigma}_0^x)^{-1}\sum_{j=0}^{j_0}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}$ are $O_p(1)$, which is equivalent to

$$\text{the solutions } \lambda \text{ of } |\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2} - \lambda\mathbf{\Sigma}_0^x| = 0 \text{ are } O_p(1). \tag{7.11}$$

Since $\text{diag}\left(\int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt,\ \text{Var}(\mathbf{x}_{1,2})\right) > 0$, by Lemma 10 the solutions (λ) of equation

$$|\mathbf{D}_n^{-1}\mathbf{\Theta}_n\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}\mathbf{\Theta}_n'\mathbf{D}_n^{-1} - \lambda\mathbf{D}_n^{-1}\mathbf{\Theta}_n\mathbf{\Sigma}_0^x\mathbf{\Theta}_n'\mathbf{D}_n^{-1}| = 0 \tag{7.12}$$

are bounded in probability. Thus, we have (7.11) and (7.9) as desired.

Similarly, we can show

$$||(\mathbf{W}^y)^{1/2}(\widehat{\mathbf{W}})^{-1/2}||_F = ||(\mathbf{W}^x)^{1/2}(\widehat{\mathbf{W}}^x)^{-1/2}||_F = O_p(1). \tag{7.13}$$

Using equations (7.10) and (7.13), the remainder of the proof of Theorem 1 consists of showing that there exist two positive constants c_1, c_2 such that in probability $\eta \geq c_1 n^{2a_1-1}/\sqrt{j_0}$ provided $|I_0| \geq 2$ or $|I_0| = 1$ and $\text{E}z_t^{I_0} = 0$ and $\eta \geq c_2 n^{2a_1}/\sqrt{j_0}$ provided $|I_0| = 1$ and $\text{E}z_t^{I_0} \neq 0$.

Define $\lambda_i(\mathbf{A})$ to be the i-th eigenvalue of a matrix $\mathbf{A}$. Note that

$$\text{diag}\left(\int_0^1 \mathbf{M}(t)\mathbf{M}'(t)\,dt,\ \text{Var}(\mathbf{x}_{1,2})\right) > 0.$$

By Lemmas 8 and 10, it follows that when $|I_0| \geq 2$ or $|I_0| = 1$ and $\text{E}z_t^{I_0} = 0$, $\lambda_{p-r}(\mathbf{\Sigma}_j^x) = O_e(n^{2a_1-1})$ and $\lambda_{p-r+1}(\widehat{\mathbf{\Sigma}}_j^x) = O_e(1)$. Thus, there exist two positive constants c_3, c_4 such that in probability

$$\lambda_{p-r}(\mathbf{W}^x) \geq \lambda_{p-r}(\mathbf{\Sigma}_0^x(\mathbf{\Sigma}_0^x)') \geq c_3 n^{2(2a_1-1)} \tag{7.14}$$

and

$$c_3 \le \lambda_{p-r+1}(\widehat{\mathbf{\Sigma}}_0^x(\widehat{\mathbf{\Sigma}}_0^x)') \le \lambda_{p-r+1}(\widehat{\mathbf{W}}^x) \le \Big[\lambda_{p-r+1}\Big(\sum_{j=0}^{j_0}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}\Big)\Big]^2 \le c_4 j_0^2. \tag{7.15}$$

Hence, in probability

$$\eta \ge |c_3 n^{2(2a_1-1)} - c_4 j_0^2|/\sqrt{c_3 n^{2(2a_1-1)} c_4 j_0^2} \ge c' n^{2a_1-1}/j_0.$$

Similarly, we have $|I_0| = 1$ and $\mathrm{E} z_t^{I_0} \neq 0$, then in probability,

$$\eta \ge c' n^{2a_1}/j_0. \tag{7.16}$$

Since j_0 is fixed, combining (7.9), (7.16) and (7.16), we complete the proof of (i) and (ii). Conclusion (iii) can be shown similarly by treating A_{1i} as the role of A_2, see also the proof of Theorem 1 of Chen and Hurvich (2006), we omit the details here. □

Let $\mathbf{A}_{1,0} = \mathbf{A}_2$ and $\widehat{\mathbf{B}}_{1i} = (\widehat{\gamma}_{\nu_i+1}, \cdots, \widehat{\gamma}_{\nu_i+r_i})$ for $i = 1, \cdots, q$ and $\widehat{\mathbf{B}}_{10} = (\widehat{\gamma}_{p-r+1}, \cdots, \widehat{\gamma}_p)$.

Lemma 11. *Under Condition 1, we have*

$$\|\mathbf{B}_{1,l}\mathbf{A}_{1,h}\|_F = O_p(n^{-2|a_h-a_l|}), \text{ for } l \neq h.$$

Proof. Let $\eta(\mathbf{B}_{1,l}, \mathbf{A}_{1,h})$ be defined as η above, i.e.,

$$\eta(\mathbf{B}_{1,l}, \mathbf{A}_{1,h}) = \min_{\lambda\in\{\widehat{\lambda}_{\nu_l+1},\cdots,\widehat{\gamma}_{\nu_l+r_l}\},\, \mu\in\{\lambda_{\nu_h+1},\cdots,\lambda_{\nu_h+r_h}\}} |\lambda-\mu|/\sqrt{\lambda\mu}.$$

By Lemmas 8 and 10, using the same arguments as in Theorem 1, we have

$$\eta(\mathbf{B}_{1,l}, \mathbf{A}_{1,h}) \ge c n^{2|a_h-a_l|} \tag{7.17}$$

for some $c > 0$. It has been shown in Theorem 1 that $||(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2}||_F = O_p(1)$, thus by Theorem 2.4 of Dopico, Moro and Molera (2000) (see also Theorem 4.1 of Barlow and Slapničar (2000)), we have

$$\begin{aligned}\|\mathbf{B}_{1,l}\mathbf{A}_{1,h}\|_F &\le ||(\mathbf{W}^y)^{-1/2}\Delta\mathbf{W}^y(\widehat{\mathbf{W}})^{-1/2}||_F/\eta(\mathbf{B}_l, \mathbf{A}_h)\\ &= O_p(n^{-2|a_h-a_l|}).\end{aligned}$$

This completes the proof of Lemma11. □

Proof of Theorem 2. First, we prove the consistency of $\widehat{r}$. For any $1 \le i \le p$,

$$\widehat{x}_t^i = \widehat{\gamma}_i'\mathbf{y}_t = (\widehat{\gamma}_i'\mathbf{A}_{1q}\mathbf{x}_{t1q}, \cdots, \widehat{\gamma}_i'\mathbf{A}_{11}\mathbf{x}_{t11}, \widehat{\gamma}_i'\mathbf{A}_2\mathbf{x}_{t2}). \tag{7.18}$$

Let ν_i be defined as in Lemma 7 and $r_0 = r$. By Lemma 11, when $\nu_l + 1 \le i \le \nu_l + r_l$, $l \neq h$,

$$\widehat{\gamma}_i'\mathbf{A}_{1h} = O_p(n^{-2|a_h-a_l|}).$$

Thus, by $\sup_{1\le t\le n}|\mathbf{x}_{t1h}| = O_p(n^{a_h-1/2})$ for $h\ge 1$ (see Lemma 7), we have

$$\widehat{\gamma}_i'\mathbf{A}_{1h}\mathbf{x}_{t1h} = O_p(n^{-a_h+2a_l-1/2})I(h>l) + O_p(n^{-2a_l+3a_h-1/2})I(1\le h<l).$$

As a result, by (7.18), it follows that for any $\nu_l+1\le i\le \nu_l+r_l$,

$$\begin{aligned}\widehat{x}_t^i &= \widehat{\gamma}_i'\mathbf{A}_{1l}\mathbf{x}_{t1l} + O_p(\sum_{h=l+1}^{q} n^{-a_h+2a_l-1/2} + \sum_{h=1}^{l} n^{-2a_l+3a_{l-1}-1/2})\\ &= \widehat{\gamma}_i'\mathbf{A}_{1l}\mathbf{x}_{t1l} + O_p(n^{-a_{l+1}+2a_l-1/2} + n^{-2a_l+3a_{l-1}-1/2}),\end{aligned}$$

where $\mathbf{x}_{t10}=\mathbf{x}_{t2}$. Thus, for any given m, we have

$$\begin{aligned}&\sum_{k=1}^{m}\Big(\frac{1}{n-k}\sum_{t=1}^{n-k}(\widehat{x}_{t+k}^i-\overline{\widehat{x}}^i)(\widehat{x}_{t,i}-\overline{\widehat{x}}^i)\Big)\\ =\ &\frac{\widehat{\gamma}_i'\mathbf{A}_{1l}}{n-k}\sum_{k=1}^{m}\sum_{t=1}^{n-k}(\mathbf{x}_{t+k,1l}-\overline{\mathbf{x}}_{1l})(\mathbf{x}_{t1l}-\overline{\mathbf{x}}_{1l})'\mathbf{A}_{1l}'\widehat{\gamma}_i(1+o_p(1)).\end{aligned} \tag{7.19}$$

By (7.19), we have that for any $\nu_l+1\le i\le\nu_l+r_l$, $l=1,\cdots,q$

$$\begin{aligned}&\sum_{k=1}^{m}\Big(\frac{1}{n-k}\sum_{t=1}^{n-k}(\widehat{x}_{t+k}^i-\overline{\widehat{x}}^i)(\widehat{x}_{t,i}-\overline{\widehat{x}}^i)\Big)\\ =\ & m\widehat{\gamma}_i'\mathbf{A}_{1l}\Big(\frac{1}{n}\sum_{t=1}^{n}(\mathbf{x}_{t1l}-\overline{\mathbf{x}}_{1l})(\mathbf{x}_{t1l}-\overline{\mathbf{x}}_{1l})'\Big)\mathbf{A}_{1l}'\widehat{\gamma}_i(1+o_p(1)) = O_e(mn^{2a_l-1}).\end{aligned} \tag{7.20}$$

On the other hand, by (7.19) and $\|\sum_{k=1}^{m}\frac{1}{n-k}\sum_{t=1}^{n-k}(\mathbf{x}_{t+k,2}-\overline{\mathbf{x}}_2)(\mathbf{x}_{t,2}'-\overline{\mathbf{x}}_2)\|\le C$ in probability, it follows that for $p-r+1\le i\le p$,

$$\sum_{k=1}^{m}\Big(\frac{1}{n-k}\sum_{t=1}^{n-k}(\widehat{x}_{t+k}^i-\overline{\widehat{x}}^i)(\widehat{x}_{t,i}-\overline{\widehat{x}}^i)\Big) = O_p(1). \tag{7.21}$$

Equation (7.20) together with (7.21) yields the conclusion of Theorem 2 as desired. □

7.2 Proofs for Section 3.2

Proof of Theorems 3 and 4. Theorem 3 can be shown similarly to Theorem 1 by using Lemma 9 instead of Lemma 8, except that when $p\to\infty$,

$$||(\mathbf{\Sigma}_0^x)^{-1}\{\widehat{\mathbf{\Sigma}}_j^x(\widehat{\mathbf{\Sigma}}_j^x)'\}^{1/2}||_F = O_p\left(\left(\sum_{i=1}^{p}(\widetilde{\lambda}_i)^2\right)^{1/2}\right) = O_p(p^{1/2}),$$

where $\widetilde{\lambda}_i, 1\le i\le p$ are solutions of (7.11). As a result, (7.9) should be replaced by

$$||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F = O_p(p^{1/2}) \text{ and } ||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F = O_p(p^{1/2}). \tag{7.22}$$

Theorem 4 can be shown similarly to Theorem 2. We omit the details. □

7.3 Proofs for Section 4

To prove Theorems 5 and 6, we first introduce some notation. Let $k_{ni} = n^{d_i-1/2}I(d_i > 1/2) + n^{d_i+1/2}I(d_i < 1/2)$ and $\lambda_i(t-s) = (t-s)^{d_i-1}/\Gamma(d_i)I(d_i > 1/2) + (t-s)^{d_i}/\Gamma(d_i+1)I(d_i < 1/2)$. Define $\mathbf{K}_n = \mathrm{diag}(k_{n1}, \cdots, k_{np})$, $\mathbf{\Lambda}(t,s) = \mathrm{diag}(\lambda_1(t-s), \cdots, \lambda_p(t-s))$ and

$$\mathbf{B}_0 = 0,\ \mathbf{B}_t = (B_t^1, \cdots, B_t^p)' = \int_0^t \mathbf{\Lambda}(t,s)\, d\mathbf{W}_s,\ \mathbf{U}_t = \mathbf{B}_t - \int_0^1 \mathbf{B}_t\, dt,$$

where $\mathbf{W}_s$ is given in (ii) of Condition 3. Let $\nabla^{d_l} v_t^l = \mu_l$, $I_1^c = \{i : d_i > 1/2\}$ and $\mathbf{x}_{t,I} = (x_t^i,\ i \in I)'$ and $\mathbf{v}_{t,I} = (v_t^i,\ i \in I)'$

Lemma 12. *Let* $\mathbf{Z}_n(t) = ((\mathbf{x}_{[nt],I_1^c} - \mathbf{v}_{[nt],I_1^c})', \sum_{j=1}^{[nt]}(\mathbf{x}_{j,I_1} - \mathbf{v}_{j,I_1})')'$. *Under (ii) of Condition 3,*

$$\mathbf{K}_n^{-1}\mathbf{Z}_n(t) \overset{J_1}{\Longrightarrow} \mathbf{B}_t,\ \text{on } D[0,1]^p. \tag{7.23}$$

Proof. Let $d_{I_1} = \{d_i : i \in I_1\}$, then $\sum_{j=1}^{[nt]} \mathbf{x}_{j,I_1}$ is an integrated fractional process with order $d_{I_1} + 1$, and each of its components has order larger than 1/2. Using (ii) of Condition 3 instead of Marinucci and Robinson (2000) Lemma 2, we can show this lemma similarly to their Theorem 1. □

Let $\mathbf{\Theta}_n$ and $\mathbf{M}_i(t)$ be defined as that after Lemma 7 by using $H^d(t) = t^d/\Gamma(d_i+1) - 1/\Gamma(d_i+2)$ and $\mathbf{F}_i(t) = (U^{\nu_i+1}(t), \cdots, U^{\nu_i+r_i}(t))'$, where U_t^i be the i-th component of $\mathbf{U}_t$. Let $\mathbf{L}_n = \mathrm{diag}(l_{n1}, \cdots, l_{np})$, $l_{ni} = n^{d_i-1/2}I(d_i > 1/2) + I(d_i < 1/2)$. Similar to Lemma 8, by Lemma 12 and the continuous mapping theorem, we have the following lemma.

Lemma 13. *Let the conditions of Theorem 5 hold. Then the following assertions hold for any* $0 \le j \le j_0$.

(i) If $\delta > 1/2$, *then*

$$\mathbf{L}_n^{-1}\mathbf{\Theta}_n\widehat{\mathbf{\Sigma}}_j^x\mathbf{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \int_0^1 (\mathbf{M}_t', \mathbf{U}_{t2}')(\mathbf{M}_t', \mathbf{U}_{t2}')\, dt, \quad \text{and}$$

$$\mathbf{L}_n^{-1}\mathbf{\Theta}_n\mathbf{\Sigma}_j^x\mathbf{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \mathrm{diag}\Big(\int_0^1 \mathbf{M}_t\mathbf{M}_t'\, dt,\ \int_0^1 \mathbf{U}_{t,2}\mathbf{U}_{t,2}'\, dt\Big),$$

where $\mathbf{U}_{t,2}$ *is corresponding to the last* p *components of* $\mathbf{U}_t$.

(ii) If $\delta < 1/2$, *then*

$$\mathbf{L}_n^{-1}\mathbf{\Theta}_n\widehat{\mathbf{\Sigma}}_j^x\mathbf{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \mathrm{diag}\left(\int_0^1 \mathbf{M}_t\mathbf{M}_t'\, dt,\ \mathrm{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1})\right), \tag{7.24}$$

and

$$\mathbf{L}_n^{-1}\mathbf{\Theta}_n\mathbf{\Sigma}_j^x\mathbf{\Theta}_n'\mathbf{L}_n^{-1} \xrightarrow{d} \mathrm{diag}\left(\int_0^1 \mathbf{M}_t\mathbf{M}_t'\, dt,\ \mathrm{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1})\right). \tag{7.25}$$

By Lemma 13, Theorems 5 and 6 can be established in a similar manner as to Theorems 1 and 2. Therefore we omit the detailed proofs.

References

[1] Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, **81**, 1203–1227.

[2] Aznar, A. and Salvador, M. (2002). Selecting the rank of cointegration space and the form of the intercept using an information criterion. *Econometric Theory*, **18**, 926–947.

[3] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica*, **6**, 311–329.

[4] Barlow, J. L. and Slapniĉar, I. (2000). Optimal perturbation bounds for Hermitian eigenvalue problem. *Linear Algebra and its Applications*, **309**, 19–43.

[5] Bierens, H. J. (1993). Higher-order sample autocorrelations and the unit root hypithesis. *Journal of Econometrics*, **57** 137–160.

[6] Bierens, H. J. (1997). Nonparametric Cointegration Analysis. *Journal of Econometrics*, **77** 379–404.

[7] Billingsley, P. (1999). *Convergence of Probability Measures* (2nd Edition). Wiley, New York.

[8] Box, G. and Tiao, G. (1977). A canonical analysis of multiple time series. *Biometrika*, **64**, 355–365.

[9] Chang, J. Y., Guo, B. and Yao, Q. (2017). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, to appear.

[10] Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, **38**, 808–835.

[11] Chen, W. W. and Hurvich, C. M. (2006). Semiparametric estimation of fractional cointegrating subspaces. *The Annals of Statistics*, **34**, 2939–2979.

[12] Dopico, F. M., Moro, J. and Molera, J. M. (2000). Weyl-type relative perturbation bounds for eigensystems of Hermitian matrices. *Linear Algebra and its Applications*, **309**, 3–18.

[13] Engle, R. and Granger, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica*, **55**, 251–276.

[14] Fakhre-Zakeria, I. and Lee, S. (2000). On functional limit theorems for multivariate linear processes with applications to sequential estimation. *Journal of Statistical Planning and Inference*, **83**, 11–23.

[15] Golub, G. and Van Loan, C. (1996). *Matrix Computations (3rd edition)*. John Hopkins University Press.

[16] Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* , **16**, 121–130.

[17] Granger, C. W. J. and Weiss, A. A. (1983). Time series analysis of error-correcting models. In: S. Karlln et al. (eds.), *Studies in Econometrics, Time series and Multivariate Analysis,* Academic Press, New York, 255–278.

[18] Hualde, J. and Robinson, P. (2010). Semiparametric inference in multivariate fractionally cointegrated system. *Journal of Econometrics* , **157**, 492–511.

[19] Harris, D. and Poskitt, D. S. (2004). Determination of cointegrating rank in partially non-stationary processes via a generalised von-Neumann criterion. *Econometrics Journal*, **7**, 191–217.

[20] Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive model. *Econometrica*, **59**, 1551–1580.

[21] Johansen, S. (1995). *Likelihood-Based inference in Cointegrated Vector in Gaussian Vector Autoregressive Model*. Oxford University Press, Oxford.

[22] Johansen, S. and Juselius, K.(1990). Maximum Likelihood Estimation and Inference on Cointegration C with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

[23] Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, **40**, 694–726.

[24] Li, Q., Pan, J and Yao, Q. (2009). On determination of cointegration ranks. *Statistics and Its Interface*, **2**, 45–56.

[25] Lin, Z. and Lu, C. (1997). *Limit Theory on Mixing Dependent Random Variables*. Kluwer Academic Publishers, 1997.

[26] Marnucci, D. and Robinson, P. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications*, **86**, 103–120.

[27] Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, **136**, 1237–1257.

[28] Phillips, P. C. B. (1991). Optimal inference for cointegrated systems. *Econometrica*, **59**, 283–306.

[29] Phillips, P. C. B. and Ouliaris, S. (1988). Testing for cointegration using principal components methods. *Journal of Economic Dynamics and Control*, **12**, 205–230.

[30] Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic properties of residual-based cointegration tests. *Econometrica*, **58**, 165–193.

[31] Robinson, P. M. (2008). Diagnostic testing for cointegration. *Journal of Econometrics* , **143**, 206–225.

[32] Robinson, P. M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration order. *Econometrica*, **71**, 1727–1766.

[33] Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, **106**, 217–241.

[34] Saikkonen, P. and Lütkepohl, H. (2000). Testing for the cointegrating rank of a VAR process with an intercept. *Econometric Theory*, **16**, 373–406.

[35] Shintani, M. (2001). A simple cointegrating rank test without vector autoregression. *Journal of Econometrics*, **105**, 337–362.

[36] Stewart, G. W. and Sun, J. (1990). *Matrix perturbation Theory*. Academic Press.

[37] Stock. J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, **55**, 1035-1056.

[38] Stock, J. H. (1999). A class of tests for integration and cointegration, in Engle, R. F. and H. White (eds.), *Cointegration, Causality and Forecasting. A Festchrift in Honour of Clive W. F. Granger*. Oxford University Press.

[39] Vu, V. Q. and Lei, J. (2013). Minimax sparse principle subspace estimation in high dimensions. *The Annals of Statistics*, **41**, 2095–2947.