Identifying Cointegration by Eigenanalysis*

Rongmao Zhang⋆ Peter Robinson† Qiwei Yao‡

⋆School of Mathematics, Zhejiang University, Hangzhou, 310058, China

†Department of Economics, London School of Economics
London, WC2A 2AE, U.K.

‡Department of Statistics, London School of Economics
London, WC2A 2AE, U.K.

rmzhang@zju.edu.cn p.m.robinson@lse.ac.uk q.yao@lse.ac.uk

6 January 2016

Abstract

We propose a new and easy-to-use method for identifying cointegrated components of nonstationary time series, consisting of an eigenalysis 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. A Monte Carlo study of finite-sample performance, and a small empirical illustration, are reported. Asymptotic justification of the method is also established in a fractional setting.

Key words and phrases. Cointegration, eigenanalysis, $I(d)$, nonstationary processes, vector time series.

MSC2010 subject classifications. Primary 62M10, 91B84, secondary 37M10.

*Partially supported by NSFC grants 11371318 (RZ), ESRC grant ES/J007242/1 (PR), and EPSRC grant EP/L01226X/1 (QY).

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), 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 is fixed as the length of our time series, n, 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 order of integration is developed in Section 3. Simulations and a small real data are reported in Section 4. In Section 5, both the proposed method and part of the asymptotic theory are extended to the fractional case. 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 following 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.

Note that $\mathbf{A}$ and $\mathbf{x}_t$ in (2.1) are not uniquely defined, as $(\mathbf{A}, \mathbf{x}_t)$ 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.

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 a $I(d)$ 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, where $U = O_e(V)$ indicates that $P(C_1 \leq |U/V| < C_2) \to 1$ as $n \to \infty$, where $C_2 > C_1 > 0$ are two finite constants. 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. 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 $\mathbf{a} \in \mathcal{M}(\mathbf{A}_1)$. Hence intuitively $\mathcal{M}(\mathbf{A}_2)$ should be the linear space spanned by the r eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the r smallest eigenvalues, and $\mathcal{M}(\mathbf{A}_1)$ is that spanned by the $(p-r)$ eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the $(p-r)$ largest eigenvalues. This point can be further elucidated as follows. Let $(\widehat{\boldsymbol{\gamma}}_1, \cdots, \widehat{\boldsymbol{\gamma}}_p)$ be the orthonormal eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the eigenvalues arranged in descending order and

$$\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2) = (\widehat{\boldsymbol{\gamma}}_1, \cdots, \widehat{\boldsymbol{\gamma}}_p),$$

then

$$\widehat{\mathbf{A}}'\widehat{\mathbf{W}}\widehat{\mathbf{A}} = \sum_{j=0}^{j_0}(\widehat{\mathbf{A}}'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}})(\widehat{\mathbf{A}}'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}) = \begin{pmatrix} \widehat{\mathbf{A}}_1'\widehat{\mathbf{W}}\widehat{\mathbf{A}}_1 & 0 \\ 0 & \widehat{\mathbf{A}}_2'\widehat{\mathbf{W}}\widehat{\mathbf{A}}_2 \end{pmatrix} \tag{2.4}$$

$$= \sum_{j=0}^{j_0} \begin{pmatrix} \widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_1 + \widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_1 & 0 \\ 0 & \widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_2 + \widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_2 \end{pmatrix}.$$

The $(1,1)$-th block on the RHS is dominated by $\sum_{j=0}^{j_0}\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_1$. The $(2,2)$-th block consists of two lower order terms, and is dominated by $\sum_{j=0}^{j_0}\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_2$ as $\sum_{j=0}^{j_0}\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_2 = o_p(1)$ (since j_0 is fixed). Consequently, we estimate $\mathbf{A}$ and $\mathbf{x}_t$ by

$$\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2), \quad \text{and} \quad \widehat{\mathbf{x}}_t = (\widehat{\mathbf{A}}_1'\mathbf{y}_t, \widehat{\mathbf{A}}_2'\mathbf{y}_t). \tag{2.5}$$

The idea 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. 2014). 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 4 in Section 4 below, and also Lam and Yao (2012) and Chang et al. (2014).

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 (2.5), is arranged according to the descending order of the eigenvalues of $\widehat{\mathbf{W}}$. Therefore, the order of the components reflects reversely the likeness of the stationarity of those component series, with $\{\widehat{x}_t^p\}$ most likely being a stationary cointegrating error series. Hence the unit-root tests (Phillips and Ouliaris, 1988) can be applied to each of the component series $\{\widehat{x}_t^p\}$, $\{\widehat{x}_t^{p-1}\}, \cdots$ to determine the cointegration rank r. Below we propose some alternative criteria to determine r.

Let $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_p \geq 0$ be the eigenvalues of $\widehat{\mathbf{W}}$. By (2.4) and (2.2), $\widehat{\lambda}_i$ is at least of the order of n^2 for all $1 \leq i \leq p-r$, and $\widehat{\lambda}_i = O_p(1)$ for all $p-r < i \leq p$. Hence as long as $1 \leq r < p$, $\widehat{\lambda}_i/(n\widehat{\lambda}_p) \to \infty$ in probability for all $1 \leq i \leq p-r$, and $\widehat{\lambda}_i/(n\widehat{\lambda}_p) = o_p(1)$ for all $p-r < i \leq p$. This leads to estimating r by

$$\widehat{r} = \max\{j : \widehat{\lambda}_{p+1-j}/(n\widehat{\lambda}_p) \leq 1,\ 1 \leq j \leq p\}. \tag{2.6}$$

See also Lam and Yao (2012) and Ahn and Horenstein (2013) for procedures based on eigenvalue ratios for factor models.

Alternatively we may define a simple information criterion as follows

$$IC(l) = \sum_{j=1}^{l} \widehat{\lambda}_{p+1-j} + (p-l)\omega_n,$$

where $\omega_n \to \infty$, $\omega_n/n^{4d_{\min}-2} \to 0$ in probability (as we allow ω_n to be data-dependent), and $d_{\min}$ is the smallest integration order among all the components of $\mathbf{y}_{t1}$. Then r can be estimated as

$$\widetilde{r} = \arg\min_{1 \leq l \leq p} IC(l). \tag{2.7}$$

Note that when $\omega_n = n\widehat{\lambda}_p$, it holds that $\widetilde{r} = \widehat{r}$.

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 $\mathcal{M}(\widehat{\mathbf{A}}_2)$ is a consistent estimator for 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 both the estimators $\widehat{r}$ and $\widetilde{r}$, defined respectively in (2.6) and (2.7), are consistent for the cointegration rank r. We consider two asymptotic regimes: (i) p is fixed while $n \to \infty$, and (ii) $p \to \infty$ more slowly than n. In this section we always assume that j_0 in (2.3) is a fixed positive integer.

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

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

where $\|\mathbf{H}\|_2 = \sup_{\|\mathbf{a}\|=1} \|\mathbf{H}\mathbf{a}\|$ 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.

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).

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

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

where $d_{\min} = \min_{1\leq i\leq p-r} d_i$, $I_0 = \{i : X_i \sim I(d_{\min}),\, 1 \leq i \leq p-r\}$ and $|I_0|$ denotes the number of elements in I_0.

Theorem 2. *Let $1 \leq r < p$ and Condition 1 hold.*

(i) For $\widehat{r}$ defined in (2.6), $\lim_{n\to\infty} P(\,\widehat{r} = r\,) = 1$.

(ii) For $\widetilde{r}$ defined in (2.7), $\lim_{n\to\infty} P(\,\widetilde{r} = r\,) = 1$ provided $1/\omega_n + \omega_n/n^{4d_{\min}-2} = o_p(1)$.

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

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 the components of $\mathbf{z}_t$ are independent and $\mathrm{E}\mathbf{z}_t = 0$. For each component (z_t^i) of $\mathbf{z}_t$, there exists an independent and standard normal sequence $\{\nu_t^i\}$ for which as $n \to \infty$,

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

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

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

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

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

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

Remark 1. *The inequalities in the line below (3.2) will hold if the z_t's are I(0) with spectral density continuous at zero frequency, because the 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 2. *When integration orders of all nonstationary components are the same, the independence assumption in Condition 2(i) can be relaxed and replaced by $\mathbf{z}_t = \mathbf{B}\boldsymbol{e}_t$, where $\mathbf{B}$ is a $p\times m$ constant matrix, $m \ge p-r$, all the components of $\boldsymbol{e}_t = (e_t^1,\cdots,e_t^m)'$ are independent, and $\{e_t^i\}$ satisfies (3.2) for $1\le i\le m$.*

Remark 3. *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 $\sup_{1\le i\le p}\mathrm{E}|\varepsilon_t^i|^q<\infty$ for some $q>2$. Furthermore, for some $2<q^*\le\min\{4,q\}$,

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

(ii) The components of $\boldsymbol{\varepsilon}_t$ are independent, $\mathrm{E}\boldsymbol{\varepsilon}_t=0$, and $\max_{1\le i\le p}\mathrm{E}|\varepsilon_t^i|^\kappa<\infty$ 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, it holds that*

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

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 4. *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 $\widetilde{r}$, defined in (2.7), converges to r in probability, provided $\lim_{n\to\infty} P\{\log n < \omega_n < (\lambda^* n^{2d_{min}-1})^2/\log n\} = 1$.*

4 Numerical properties

We illustrate the proposed method with both simulated and real data examples below. Note that the comparison with Johanson's (1991) likelihood method is carried out for Example 1 only, as Examples 2 & 3 consider the settings with $d_1, \cdots, d_m > 1$ for which his method is not applicable.

Example 1. Let in model (2.1) all components of $\mathbf{x}_{t2}$ be stationary AR(1) and all components of $\mathbf{x}_{t1}$ be ARIMA(1,1,1) processes. The AR(1) coefficients are generated independently from $U(-0.8, 0.8)$, where the AR and MA coefficients in the ARIMA(1,1,1) are generated independently from $U(0.3, 0.8)$ and $U(0, 0.95)$. The innovations in these processes are independent $N(0,1)$. The elements of $\mathbf{A}$ are generated independently from $U(-3,3)$. We consider various combinations for p, $r = p/4$ and $r = 3$, and sample size n between 500 and 2500; see Table 1. For each setting, generate 500 replicates. We estimate the cointegration rank r using both the ratio method (2.6) and the information criterion (2.7). Since the estimated cointegration rank 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:

$$D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2)) = \Big\{1 - \frac{1}{\max(r, r^*)}\mathrm{tr}\big(\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\mathbf{B}_2(\mathbf{B}_2'\mathbf{B}_2)^{-1}\mathbf{B}_2'\big)\Big\}^{1/2}, \tag{4.1}$$

where $r^* = \mathrm{rank}(\widehat{\mathbf{A}}_2)$, and $\mathbf{B}_2$ 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 orthogonal with each other, and 0 if and only if the two subspaces are the same. When $r^* = r$ 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). We use $j_0 = 5$ in the definition of $\widehat{\mathbf{W}}$ in (2.3).

We compare the performance of our procedure with Johansen's (1991) trace test with significance level $\alpha = 0.05$. Since the limiting distribution, i.e. the distribution of $(\int_0^1 (d\mathbf{W})\mathbf{F}'$ $[\int_0^1 \mathbf{F}\mathbf{F}'\,dt]^{-1}\int_0^1 \mathbf{F}(d\mathbf{W})')$, in his test is nonstandard, we approximate it by the distribution of

$$\Big[\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{X}_{t-1} - \bar{X}_{-1})'\Big]\Big[\sum_{t=1}^{T}(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}}_{-1})(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}}_{-1})'\Big]^{-1}\Big[\sum_{t=1}^{T}(\boldsymbol{X}_{t-1} - \bar{X}_{-1})\boldsymbol{\varepsilon}_t'\Big], \tag{4.2}$$

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 $N(0,1)$ variables. Setting $T = 1000$, the critical values were calculated in a simulation with 6000 repetitions of the trace of (4.2) for $p > 5$. This is the procedure used in Johansen and Juselius (1990) for calculating the critical values with $p \leq 5$.

Table 1 reports the relative frequencies (Freq) of the events $\{\widehat{r} = r\}$ and $\{\widetilde{r} = r\}$, and the average of the distance (Dist) $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$ (see (4.1)) for $r = p/4$ in a simulation with 500 replications, where the penalty ω_n in the information criterion $IC(l)$ is taken as either $\omega_n^1 = n^{5/4}\widehat{\lambda}_p$ or $\omega_n^2 = n^{3/2}\widehat{\lambda}_p$, and $\widehat{\lambda}_p$ is the smallest eigenvalue of $\widehat{\mathbf{W}}$. Also included in Table 1 are the results resulted from applying the Johansen likelihood test for the transformed component series. From Table 1, we see that our procedure always has higher relative frequencies and smaller distances, which indicates that our procedure outperforms Johansen's likelihood method when r is relatively small. Similar pattern are observed in Table 2 with $r = 3$.

Table 1: Relative frequencies for $\widetilde{r} = r$, $r = p/4$ and average distance in simulation with 500 replications in Example 1.

p	n	500		1000		1500		2000		2500	
		Freq	Dist	Freq	Dist	Freq	Dist	Freq	Dist	Freq	Dist
8	Johansen	0.390	0.371	0.452	0.326	0.490	0.302	0.480	0.307	0.514	0.289
	ratio	0.748	0.174	0.848	0.105	0.884	0.081	0.886	0.079	0.890	0.074
	IC(ω_n^1)	0.654	0.217	0.780	0.136	0.802	0.123	0.818	0.112	0.852	.091
	IC(ω_n^2)	0.448	0.338	0.572	0.136	0.628	0.222	0.656	0.206	0.690	0.185
12	Johansen	0.210	0.449	0.344	0.355	0.380	0.336	0.400	0.322	0.464	0.287
	ratio	0.658	0.236	0.794	0.138	0.770	0.151	0.844	0.102	0.840	0.107
	IC(ω_n^1)	0.556	0.261	0.708	0.168	0.748	0.145	0.796	0.114	0.824	0.101
	IC(ω_n^2)	0.366	0.358	0.444	0.299	0.518	0.258	0.536	0.247	0.610	0.206
20	Johansen	0.008	0.604	0.050	0.503	0.080	0.456	0.134	0.425	0.164	0.406
	ratio	0.404	0.390	0.544	0.299	0.620	0.243	0.704	0.184	0.730	0.169
	IC(ω_n^1)	0.390	0.342	0.554	0.245	0.670	0.183	0.686	0.154	0.768	0.121
	IC(ω_n^2)	0.232	0.417	0.346	0.331	0.400	0.294	0.456	0.256	0.472	0.245
28	Johansen	0	0.696	0	0.595	0.002	0.549	0.004	0.522	0.010	0.501
	ratio	0.234	0.489	0.386	0.372	0.462	0.332	0.558	0.280	0.582	0.250
	IC(ω_n^1)	0.252	0.407	0.454	0.274	0.546	0.228	0.610	0.195	0.700	0.144
	IC(ω_n^2)	0.176	0.442	0.270	0.350	0.334	0.304	0.358	0.284	0.436	0.240

Example 2. Let in model (2.1) all components of $\mathbf{x}_{t2}$ be stationary AR(1) and all components of $\mathbf{x}_{t1}$ be ARIMA(1,2,1) processes. The AR(1) coefficients are generated independently from $U(-0.8, 0.8)$, the AR and MA coefficients in those ARIMA(1,2,1) are generated independently from $U(0.3, 0.8)$ and $U(0, 0.95)$. The innovations in these processes are independent $N(0,1)$. The elements of $\mathbf{A}$ are generated independently from $U(-3, 3)$. We consider various combinations for p and r, and sample size n between 300 and 2500; see Table 3. For each setting, we replicate the simulation 1000 times.

Fig.1 presents the sample ACF/CCF of $\mathbf{y}_t$ for one instance with $n = 1000$, $p = 6$ and $r = 2$.

Table 2: Relative frequencies for $\widetilde{r} = r$, $r = 3$ and average distance in simulation with 500 replications in Example 1.

p	n	500		1000		1500		2000		2500	
		Freq	Dist	Freq	Dist	Freq	Dist	Freq	Dist	Freq	Dist
5	Johansen	0.834	0.086	0.838	0.086	0.844	0.081	0.856	0.074	0.842	0.080
	ratio	0.620	0.252	0.704	0.197	0.734	0.176	0.802	0.131	0.816	0.122
	IC(ω_n^1)	0.802	0.130	0.868	0.087	0.884	0.077	0.916	0.058	0.922	0.050
	IC(ω_n^2)	0.902	0.065	0.922	0.051	0.940	0.038	0.958	0.0284	0.962	0.025
10	Johansen	0.522	0.277	0.562	0.247	0.540	0.255	0.606	0.217	0.550	0.248
	ratio	0.636	0.242	0.788	0.145	0.790	0.144	0.860	0.095	0.874	0.084
	IC(ω_n^1)	0.654	0.210	0.812	0.118	0.856	0.089	0.882	0.072	0.896	0.062
	IC(ω_n^2)	0.504	0.275	0.628	0.201	0.672	0.175	0.746	0.134	0.740	0.137
15	Johansen	0.078	0.594	0.184	0.485	0.276	0.423	0.320	0.392	0.346	0.371
	ratio	0.526	0.311	0.692	0.192	0.742	0.163	0.808	0.119	0.810	0.116
	IC(ω_n^1)	0.356	0.381	0.494	0.285	0.578	0.235	0.644	0.196	0.684	0.170
	IC(ω_n^2)	0.174	0.498	0.262	0.428	0.316	0.390	0.336	0.372	0.388	0.337
20	Johansen	0	0.749	0.018	0.656	0.022	0.618	0.053	0.579	0.060	0.572
	ratio	0.316	0.443	0.464	0.315	0.552	0.268	0.606	0.233	0.645	0.207
	IC(ω_n^1)	0.176	0.527	0.262	0.440	0.316	0.396	0.361	0.362	0.396	0.337
	IC(ω_n^2)	0.100	0.617	0.114	0.577	0.140	0.544	0.176	0.516	0.178	0.509
25	Johansen	0	0.847	0	0.766	0.002	0.713	0	0.695	0.013	0.672
	ratio	0.242	0.497	0.315	0.423	0.374	0.371	0.413	0.343	0.530	0.275
	IC(ω_n^1)	0.132	0.596	0.166	0.527	0.170	0.518	0.213	0.472	0.246	0.461
	IC(ω_n^2)	0.064	0.690	0.071	0.660	0.078	0.644	0.096	0.612	0.103	0.603
30	Johansen	0	0.905	0	0.831	0	0.796	0	0.769	0.003	0.747
	ratio	0.186	0.559	0.270	0.456	0.295	0.423	0.313	0.405	0.363	0.368
	IC(ω_n^1)	0.103	0.656	0.106	0.616	0.105	0.593	0.146	0.551	0.140	0.548
	IC(ω_n^2)	0.056	0.744	0.030	0.730	0.060	0.698	0.063	0.683	0.045	0.693

By applying our proposed method to this sample, the sample ACF/CCF of the transformed $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ are plotted in Fig.2. Those figures show clearly that all the components of $\mathbf{y}_t$ are nonstationary while the last two components of $\widehat{\mathbf{x}}_t$ are stationary.

(Put Figures 1 & 2 about here.)

Table 3 reports the relative frequencies of the events $\{\widehat{r} = r\}$ and $\{\widetilde{r} = r\}$ in a simulation with 1000 replications, where the penalty ω_n in the information criterion $IC(l)$ is taken as either $\omega_n^1 = n^{5/4}\widehat{\lambda}_p$ or $\omega_n^2 = n^{3/2}\widehat{\lambda}_p$, and $\widehat{\lambda}_p$ is the smallest eigenvalue of $\widehat{\mathbf{W}}$. Also included in Table 3 are the results resulted from applying the Phillips-Perron unit-root test for the transformed component series. We choose the Phillips-Perron method among other unit-root tests as it is applicable with different integration orders. Note that when $\omega_n = n\widehat{\lambda}_p$, $\widehat{r} = \widetilde{r}$. While the numerical results in Table 3 lend further evidence for the consistency of both estimators, finite sample performance depends on choice of the penalty parameter ω_n: large ω_n should be used when r is relatively large. But the unit-root tests lead to very accurate estimates for the cointegration ranks for this example.

The boxplots of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$ are presented in Figs. 3 – 5 for $(p, r) = (6, 2)$, $(10, 4)$ and

Table 3: Relative frequencies for $\widehat{r} = r$ and $\widetilde{r} = r$ in simulation with 1000 replications in Example 2.

(p, r)	n	300	500	1000	1500	2000	2500
p=6, r=2	Ratio	0.835	0.887	0.979	0.993	1.000	0.999
	IC(ω_n^1)	0.923	0.953	0.994	1.000	1.000	1.000
	IC(ω_n^2)	0.967	0.985	0.998	0.998	1.000	0.998
	Unit-root test	0.999	1.000	0.998	0.989	0.995	0.992
p=6, r=4	Ratio	0.278	0.343	0.715	0.921	0.970	0.988
	IC(ω_n^1)	0.543	0.644	0.920	0.987	0.994	0.998
	IC(ω_n^2)	0.762	0.852	0.986	0.999	1.000	0.988
	Unit-root test	1.000	1.000	1.000	0.995	0.998	0.993
p=10, r=2	Ratio	0.799	0.906	0.993	0.992	0.995	0.991
	IC(ω_n^1)	0.822	0.95	0.988	0.984	0.978	1.000
	IC(ω_n^2)	0.736	0.904	0.953	0.922	0.928	1.000
	Unit-root test	0.978	0.991	0.997	0.991	0.990	0.986
p=10, r=4	Ratio	0.333	0.459	0.880	0.978	0.988	0.997
	IC(ω_n^1)	0.594	0.774	0.982	0.998	0.999	0.999
	IC(ω_n^2)	0.802	0.937	0.996	0.998	0.999	0.999
	Unit-root test	0.999	1.000	0.998	0.989	0.995	0.992
p=20, r=6	Ratio	0.994	0.998	0.996	0.996	0.994	0.991
	IC(ω_n^1)	0.075	0.565	0.948	0.940	0.896	0.882
	IC(ω_n^2)	0.330	0.791	0.798	0.691	0.616	0.558
	Unit-root test	0.815	0.289	0.503	0.805	0.901	0.858
p=20, r=10	Ratio	0.000	0.005	0.479	0.873	0.946	0.951
	IC(ω_n^1)	0.000	0.050	0.857	0.974	0.991	0.993
	IC(ω_n^2)	0.003	0.410	0.972	0.996	0.995	0.999
	Unit-root test	0.994	0.961	0.976	0.999	0.994	0.989
p=20, r=14	Ratio	0.000	0.000	0.026	0.356	0.753	0.874
	IC(ω_n^1)	0.000	0.000	0.254	0.791	0.949	0.983
	IC(ω_n^2)	0.000	0.015	0.717	0.958	0.993	0.996
	Unit-root test	0.987	1.000	0.998	0.999	0.993	0.996

$(20, 14)$ respectively. For all the settings reported, $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$ decreases as the sample size n increases.

(Put Figures 3 – 5 about here.)

To illustrate the impact of j_0 used in defining $\widehat{\mathbf{W}}$ in (2.3), we ran the simulation with $n = 500$ and j_0 taking 7 different values between 5 and 100. Each setting is repeated 500 times. The results are reported in Table 4. The different values of j_0 lead to about the same performance in terms of the relative frequencies for $\widetilde{r} = r$ and the means and the standard deviations for distance $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$. For example, when $p = 10$ and $r = 2$, the estimation for r improves slightly when j_0 increases, while the estimation for the cointegration space becomes slightly worse. Overall Table 4 suggests that the proposed method may be insensitive to the choice of j_0 in (2.3).

Example 3. Now we consider an example in which some components of $\mathbf{y}_{t1}$ are $I(1)$ and some are $I(2)$. More precisely in model (2.1) all components of $\mathbf{x}_{t2}$ are stationary AR(1), s components of $\mathbf{x}_{t1}$ are ARIMA(1,1,1), and the other $p - r - s$ components are ARIMA(0,2,1). The AR(1) coefficients are taken as $-0.8 + 1.6i/r$, $i = 1, 2, \cdots, r$. The ARIMA(1,1,1) coefficients are taken as $0.3 + 0.5i/s$ and $0.2 + 0.6i/s$, $i = 1, 2, \cdots, s$ respectively. The ARIMA(0,2,1) coefficients are

Table 4: Relative frequencies for $\widetilde{r} = r$ with ω_n^2, and means and standard deviations (STD) of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$ in simulation with 500 replications in Example 2.

(p, r)	j_0		5	10	20	30	40	50	100
p=6, r=2	Relative frequency		0.984	0.984	0.988	0.988	0.982	0.986	0.986
	D_1	mean	0.010	0.012	0.009	0.007	0.013	0.010	0.010
		STD	0.083	0.088	0.077	0.070	0.094	0.083	0.083
p=6, r=4	Relative frequency		0.862	0.862	0.870	0.874	0.878	0.866	0.878
	D_1	mean	0.074	0.078	0.075	0.073	0.068	0.077	0.070
		STD	0.187	0.201	0.199	0.197	0.187	0.199	0.192
p=10, r=2	Relative frequency		0.882	0.884	0.916	0.900	0.934	0.936	0.942
	D_1	mean	0.008	0.006	0.006	0.009	0.007	0.008	0.009
		STD	0.055	0.045	0.045	0.063	0.055	0.062	0.070

generated independently from $U(-0.95, 0.95)$. The innovations in these processes are independent $N(0, 1)$. Let the elements of $\mathbf{A}$ be generated independently from $U(-3, 3)$. We consider various combinations for p, r, s, and the sample size n. For each setting, we replicate the simulation 1000 times and estimate the cointegration rank r using both the ratio method (2.6) and the information criterion (2.7) with ω_n equal to either $\omega_n^1 = n^{5/4}\widehat{\lambda}_p$ or $\omega_n^3 = n^{2/3}\widehat{\lambda}_p$.

Fig.6 plots the sample ACF/CCF of $\mathbf{y}_t$ for one instance with $n = 1000$, $p = 6$, $r = 2$ and $s = 2$. The sample ACF/CCF of the transformed $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ are plotted in Fig.7. Those figures show clearly that all the components of $\mathbf{y}_t$ are nonstationary while the last two components of $\widehat{\mathbf{x}}_t$ are stationary.

(Put Figures 6 – 7 about here.)

Table 5 reports the relative frequencies of the events $\{\widehat{r} = r\}$ and $\{\widetilde{r} = r\}$ in a simulation with 1000 replications, where the penalty ω_n in the information criterion $IC(l)$ is taken as either $\omega_n^1 = n^{5/4}\widehat{\lambda}_p$ or $\omega_n^3 = n^{2/3}\widehat{\lambda}_p$, and $\widehat{\lambda}_p$ is the smallest eigenvalue of $\widehat{\mathbf{W}}$. The estimates for the cointegration ranks by the Phillips-Perron test are more accurate. Comparing to Table 3, the estimation for the cointegration rank r is less accurate than that for Example 1. This is due to the existence of different integration orders for the different components of $\mathbf{y}_t$, which implies that the eigenvalues of $\widehat{\mathbf{W}}$ are more diverse; see (2.2). However the estimation for the cointegrated variables themselves is hardly affected. We plot the boxplots of the distances between the true cointegrated space $\mathcal{M}(\mathbf{B}_2)$ and its estimator $\mathcal{M}(\widehat{\mathbf{A}}_2)$, defined as in (4.1), in Figs 8–9 for the four different settings for (p, r, s), where the cointegration rank r is either estimated by the ratio method or simply set at its true value. The distances for the estimation with true r are significantly smaller than those for the estimation with estimated r. These results indicate clearly that while the performance of the estimators for the cointegration rank r is not entirely satisfactory when the components of $\mathbf{y}_t$ have different cointegration ranks, the transformed series $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ contain the well estimated cointegrated variables.

Table 5: Relative frequencies for $\widehat{r} = r$ and $\widetilde{r} = r$ in simulation with 1000 replications in Example 3.

(p, r, s)	n	300	500	1000	1500	2000	2500
(6, 2, 2)	Ratio	0.711	0.778	0.873	0.918	0.909	0.873
	IC(ω_n^1)	0.788	0.841	0.877	0.857	0.783	0.687
	IC(ω_n^3)	0.476	0.522	0.623	0.749	0.846	0.893
	Unit-root	0.968	0.969	0.961	0.951	0.959	0.947
(6, 4, 1)	Ratio	0.186	0.205	0.290	0.535	0.698	0.822
	IC(ω_n^1)	0.406	0.430	0.600	0.836	0.925	0.962
	IC(ω_n^3)	0.020	0.035	0.039	0.114	0.235	0.346
	Unit-root	1.000	1.000	0.997	0.999	0.997	0.998
(10, 4, 1)	Ratio	0.056	0.084	0.384	0.786	0.936	0.968
	IC(ω_n^1)	0.232	0.306	0.752	0.940	0.964	0.937
	IC(ω_n^3)	0.002	0.007	0.049	0.303	0.607	0.801
	Unit-root	0.961	0.978	0.970	0.946	0.907	0.904
(10, 6, 2)	Ratio	0.018	0.035	0.105	0.372	0.637	0.801
	IC(ω_n^1)	0.096	0.122	0.448	0.744	0.872	0.914
	IC(ω_n^3)	0.000	0.000	0.002	0.046	0.192	0.369
	Unit-root	0.986	0.984	0.986	0.960	0.954	0.952
(20, 10, 1)	Ratio	0.000	0.002	0.043	0.493	0.802	0.925
	IC(ω_n^1)	0.001	0.003	0.354	0.831	0.909	0.898
	IC(ω_n^3)	0.000	0.000	0.000	0.033	0.250	0.522
	Unit-root	0.577	0.720	0.772	0.734	0.654	0.633
(20, 14, 2)	Ratio	0.000	0.000	0.000	0.060	0.295	0.560
	IC(ω_n^1)	0.000	0.000	0.021	0.409	0.698	0.845
	IC(ω_n^3)	0.000	.000	0.000	0.001	0.008	0.046
	Unit-root	0.962	0.939	0.879	0.873	0.854	0.827

Example 4. For an empirical example, we consider the 8 monthly US Industrial Production indices in 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.10. Applying the proposed method to these data, the transformed series $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ are plotted in Fig.11 together with their sample ACF. The ratio method (2.6) claims $\widehat{r} = 3$ cointegrated variables. The IC method (2.7) leads to $\widetilde{r} = 3$ with $\omega_n = n^{5/4}\widehat{\lambda}_8$, and $\widetilde{r} = 4$ with $\omega_n = n^{3/2}\widehat{\lambda}_8$, where $\widehat{\lambda}_8$ is the minimum eigenvalue of $\widehat{\mathbf{W}}$ defined as in (2.3). Indeed the last 3 or 4 series in Fig.11 certainly look stationary.

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.12.

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.12, 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.

5 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 the cases when the components of $\mathbf{y}_t$ may be fractionally integrated. For simplicity, we now assume p is fixed.

We first present a gentle introduction for fractionally integrated processes and the concept of fractional cointegration.

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 $0 \leq c < 1/2$. Thus let $d_{\min} > 1/2$ be the smallest integration order of elements of $\mathbf{x}_{t1}$ and let $\delta \in [0, 1/2)$ be the largest 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, though now a large j_0 should be used in (2.3).

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

Condition 3.

(i) $\mathrm{E}||\boldsymbol{\varepsilon}_t||_2^q < \infty$ for some $q > \max(4, 2/(2d_{\min} - 1))$ 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 $n \to \infty$,

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

Remark 5. *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:*

$$\boldsymbol{\varepsilon}_t = \mathbf{C}_t \boldsymbol{\varepsilon}_{t-1} + \mathbf{e}_t \tag{5.1}$$

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,*

$$D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_p(n^{-2d_{\min}+1}).$$

Let $\widehat{r}^* = \max\{j : \widehat{\lambda}_{p+1-j}/(n^{d_{\min}+\delta-1}\widehat{\lambda}_p) \le 1,\ 1 \le j \le p\}$ and $\widetilde{r}$ be defined as in (2.7).

Theorem 6. *Let Condition 3 hold.*

(i) $\lim_{n\to\infty} P(\,\widehat{r}^* = r\,) = 1$ *provided* $1 \le r < p$ *and*

(ii) $\lim_{n\to\infty} P(\,\widetilde{r} = r\,) = 1$ *provided* $\lim_{n\to\infty}(1/\omega_n + \omega_n n^{-4d_{\min}+2}) = 0$.

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 the ascending order according to the "degree" of stationarity; reflected by the magnitude of the eigenvalues of $\widehat{\mathbf{W}}$. Then in addition to the proposed information criterion for determining the cointegration order, unit-root tests may be applied to determine the number of stationary components of $\widehat{\mathbf{x}}_t$.

In this paper we only focus on inference on the cointegration rank r and cointegration space $\mathcal{M}(\mathbf{A}_2)$. One practically relevant open problem is to identify the subspaces of $\mathcal{M}(\mathbf{A}_1)$ according to the different integration orders of the components of $\widehat{\mathbf{x}}_{t1}$. Further, it would be interesting to consider letting j_0 in (2.3) diverge together with the sample size n.

7 Appendix: Technical proofs

7.1 Proof for Section 3.1

Let

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

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

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

where $\boldsymbol{\Lambda}_x$ 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 $\boldsymbol{\Gamma}_x$ must be block-diagonal. Define $\mathbf{W}^y = \mathbf{A}\mathbf{W}^x\mathbf{A}'$, then

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

This implies that the columns of $\mathbf{A}\boldsymbol{\Gamma}_x$ are just the orthogonal eigenvector of $\mathbf{W}^y$. Since $\boldsymbol{\Gamma}_x$ is block-diagonal, it follows that $\mathcal{M}(\mathbf{A}_2)$ is 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 the 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

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

where $\lambda(\mathbf{A})$ denotes the set of eigenvalues of a matrix $\mathbf{A}$. When $||\Delta\mathbf{W}^y|| = o_p(\operatorname{sep}(\mathbf{D}_1^x,\mathbf{D}_2^x))$, 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 (2014). However, in our setting $\operatorname{sep}(\mathbf{D}_1^x,\mathbf{D}_2^x)$ can be of smaller order than $||\Delta\mathbf{W}^y||$, i.e., $\operatorname{sep}(\mathbf{D}_1^x,\mathbf{D}_2^x)/||\Delta\mathbf{W}^y|| \xrightarrow{p} 0$ 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 $\mathbf{W}^x$ $(=\sum_{j=1}^{j_0}\boldsymbol{\Sigma}_j^x(\boldsymbol{\Sigma}_j^x)')$, 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 $\mathbf{C}_n\mathbf{W}^y\mathbf{C}_n = O_e(1)$, as a result, the argument of Chen and Hurvich (2006) based on the perturbation bound of Barlow and Slapnicar (2002) cannot be used. To our end, we first establish some lemmas and put their proofs in the supplementary material.

For $1 \le i \le p-r$, set $f_0^i(t) = W^i(t)$, $f_{d_i}^i(t) = \int_0^t f_{d_i-1}^i(s)\,ds$, $\mu_i = \mathrm{E}z_t^i$ 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 s_1 components of $\mathbf{x}_{t1}$ are $I(a_1)$, the next s_2 components are $I(a_2)$ and the last s_l components of $\mathbf{x}_{t1}$ are $I(a_l)$, that is,

$$\boldsymbol{x}_{t1} = (\overbrace{x_t^1, \cdots, x_t^{s_1}}^{I(a_1)}, \overbrace{x_t^{s_1+1}, \cdots, x_t^{s_1+s_2}}^{I(a_2)}, \cdots, \overbrace{x_t^{\sum_{j=1}^{l-1}s_j+1}, \cdots, x_t^{\sum_{j=1}^{l}s_j}}^{I(a_l)})',$$

where $a_1 > a_2 > \cdots > a_l = d_{\min}$ are positive integers and $\sum_{i=1}^l s_i = p-r$. For $1 \le i \le l$, define $\nu_i = \sum_{j=1}^{i-1} s_j$. Then for any $\mathbf{x}_t(s_i) := (x_t^{\nu_i+1}, \cdots, x_t^{\nu_i+s_i})'$, if $\boldsymbol{\mu}_i := (\mu_{\nu_i+1}, \cdots, \mu_{\nu_i+s_i})' \neq 0$, there must exist a $s_i \times (s_i-1)$ matrix $\mathbf{P}_i$ and $s_1 \times 1$ vector $\bar{\boldsymbol{\mu}}_i$ such that $\mathbf{P}_i'\mathbf{P}_i = \mathbf{I}_{(s_i-1)}$, $(\mathbf{P}_i, \boldsymbol{\mu}_i)$ has full rank s_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 $\mathbf{B}_i = \mathbf{I}_{s_i}$ if $\boldsymbol{\mu}_i = 0$, and $\mathbf{\Theta}_n = \mathrm{diag}(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_l, \mathbf{I}_r)$. Define

$$\mathbf{D}_{n1} = \mathrm{diag}\Big(\overbrace{n^{a_1-1/2}, \cdots, n^{a_1-1/2}}^{s_1}, \cdots, \overbrace{n^{a_l-1/2}, \cdots, n^{a_l-1/2}}^{s_l}\Big), \;\; \mathbf{D}_{n2} = (\overbrace{1, \cdots, 1}^{r}),$$

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, $\mathbf{F}_i(t) = (F^{\nu_i+1}(t), \cdots, F^{\nu_i+s_i}(t))'$, $\mathbf{M}_i(t) = (\mathbf{F}_i'(t)\mathbf{P}_i, H^{a_i}(t))'I(\boldsymbol{\mu}_i \neq 0) + \mathbf{F}_i(t)I(\boldsymbol{\mu}_i = 0)$, and $\mathbf{M}(t) = (\mathbf{M}_1'(t), \mathbf{M}_2'(t), \cdots, \mathbf{M}_l'(t))'$. By Lemma 7 and continuous mapping theorem, we have

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

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

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 $\mathbf{F}(t) = (F^1(t), F^2(t), \cdots, F^{p-r}(t))'$. We have

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

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

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*

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

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

Proof of Theorem 1. Since

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

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

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

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

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

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 can be transformed to show that

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

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 remaining proof for Theorem 1 is to show that there exist two positive constants c_1, c_2 such that in probability $\eta \geq c_1 n^{2d_{min}-1}/\sqrt{j_0}$ provided $|I_0| \geq 2$ or $|I_0| = 1$ and $\text{E}z_t^{I_0} = 0$ and $\eta \geq c_2 n^{2d_{min}}/\sqrt{j_0}$ provided $|I_0| = 1$ and $\text{E}z_t^{I_0} \neq 0$.

Define $\lambda_i(\mathbf{A})$ 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 $\text{E}z_t^{I_0} = 0$, $\lambda_{p-r}(\mathbf{\Sigma}_j^x) = O_e(n^{2d_{min}-1})$ and $\lambda_{p-r+1}(\widehat{\mathbf{\Sigma}}_j^x) = O_e(1)$. Thus, there exists two positive constants c_3, c_4 such that in probability

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

and

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

Hence, in probability

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

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

$$\eta \geq c' n^{2d_{min}}/j_0. \tag{7.16}$$

Since j_0 is fixed, combining (7.9), (7.16) and (7.16), we complete the proof of Theorem 1. □

Proof of Theorem 2. First, we prove the consistency of $\widehat{r}$. Similar to (7.15), there exist two positive constants c_5, c_6 such that

$$c_5 \leq \lambda_p(\widehat{\mathbf{\Sigma}}_0^x(\widehat{\mathbf{\Sigma}}_0^x)') \leq \lambda_p(\widehat{\mathbf{W}}^x) \leq c_6 j_0^2. \tag{7.17}$$

Since $\widehat{\lambda}_i = \lambda_i(\widehat{\mathbf{W}}) = \lambda_i(\mathbf{A}\widehat{\mathbf{W}}^x\mathbf{A}') = \lambda_i(\widehat{\mathbf{W}}^x)$, equations (7.14), (7.15) and (7.17) imply when $|I_0| \geq 2$ or $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$,

$$\frac{\widehat{\lambda}_i}{n\widehat{\lambda}_p} \geq \frac{c_3 n^{4d_{min}-3}}{c_6 j_0^2} \text{ for } i \leq p-r \quad \text{and} \quad \frac{\widehat{\lambda}_i}{n\widehat{\lambda}_p} \leq \frac{c_4 j_0^2}{n c_5} \text{ for } i = p-r+1, \cdots, p$$

hold uniformly in probability. As a result, we have $\lim_{n\to\infty} P\{\widehat{r} = r\} = 1$ provided $|I_0| \geq 2$ or $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$. The consistency of $\widehat{r}$ under the setting that $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} \neq 0$, can be proved similarly.

As for the consistency of $\widetilde{r}$, it follows from its definition that

$$\sum_{j=1}^{\widetilde{r}} \widehat{\lambda}_{p+1-j} + (p-\widetilde{r})\omega_n \leq \sum_{j=1}^{r} \widehat{\lambda}_{p+1-j} + (p-r)\omega_n. \tag{7.18}$$

Suppose that $\widetilde{r} < r$, it follows from (7.18) that

$$(r-\widetilde{r})\omega_n \leq \sum_{j=\widetilde{r}+1}^{r} \widehat{\lambda}_{p+1-j} \leq (r-\widetilde{r})\widehat{\lambda}_{p+1-r}. \tag{7.19}$$

However equation (7.15) implies that in probability,

$$\widehat{\lambda}_{p+1-r} = \lambda_{p+1-r}(\mathbf{A}\widehat{\mathbf{W}}^x\mathbf{A}') = \lambda_{p+1-r}(\widehat{\mathbf{W}}^x) \leq c_4 j_0^2.$$

Since $\omega_n / j_0^2 \to \infty$, it follows that equation (7.19) holds with probability zero. This gives that

$$\lim_{n\to\infty} P\{\widetilde{r} < r\} = 0. \tag{7.20}$$

On the other hand, if $\widetilde{r} > r$, equation (7.18) yields

$$(\widetilde{r}-r)\widehat{\lambda}_{p-r} \leq \sum_{j=r+1}^{\widetilde{r}} \widehat{\lambda}_{p+1-j} \leq (\widetilde{r}-r)\omega_n. \tag{7.21}$$

By (7.14), we have when $|I_0| \geq 2$ or $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$,

$$\widehat{\lambda}_{p-r} = \lambda_{p+1-r}(\widehat{\mathbf{W}}^x) \geq c_3 n^{2(2d_{min}-1)}. \tag{7.22}$$

A similar argument to (7.14) deduces when $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} \neq 0$,

$$\widehat{\lambda}_{p-r} = \lambda_{p+1-r}(\widehat{\mathbf{W}}^x) \geq c_3' n^{4d_{min}}. \tag{7.23}$$

Since $\omega_n / n^{2(2d-1)} \to 0$ as $n \to \infty$, equations (7.21)–(7.23) imply

$$\lim_{n\to\infty} P\{\widetilde{r} > r\} = 0. \tag{7.24}$$

Equation (7.20) together with (7.24) give the conclusion (ii) 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

$$||(\mathbf{W}^y)^{-1/2}(\widehat{\mathbf{W}})^{1/2}||_F = O_p(p^{1/2}) \quad \text{and} \quad ||(\mathbf{W}^x)^{-1/2}(\widehat{\mathbf{W}}^x)^{1/2}||_F = O_p(p^{1/2}). \tag{7.25}$$

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

7.3 Proofs for Section 5

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 $\mathbf{K}_n = \text{diag}(k_{n1}, \cdots, k_{np})$, $\mathbf{\Lambda}(t,s) = \text{diag}(\lambda_1(t-s), \cdots, \lambda_p(t-s))$ 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.

Lemma 11. *Let* $I_1^c = \{i : d_i > 1/2\}$, $\mathbf{x}_{t,I} = (x_t^i,\ i \in I)'$ *and* $\mathbf{Z}_n(t) = (\mathbf{x}'_{[nt],I_1^c}, \sum_{j=1}^{[nt]}\mathbf{x}'_{j,I_1})'$. *Under (ii) of Condition 3, we have*

$$\mathbf{K}_n^{-1}\mathbf{Z}_n(t) \xRightarrow{J_1} \mathbf{B}_t,\ \textit{on } D[0,1]^p. \tag{7.26}$$

Proof. Let $d_{I_1} = \{d_i :\ i \in I_1\}$, then $\sum_{j=1}^{[nt]}\mathbf{x}_{j,I_1}$ is a integrated fractional process with order $d_{I_1}+1$, 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. □

Lemma 12. *Under Condition 3, for any* $0 \le j \le j_0$, *we have*

(i) If $I_1 = \emptyset$, *then*

$$\mathbf{L}_n^{-1}\widehat{\mathbf{\Sigma}}_j^x\mathbf{L}_n^{-1} \xrightarrow{d} \int_0^1 \mathbf{U}_t\mathbf{U}_t'\, dt \ \textit{and}\ \mathbf{L}_n^{-1}\mathbf{\Sigma}_j^x\mathbf{L}_n^{-1} \xrightarrow{d} \text{diag}\Big(\int_0^1 \mathbf{U}_{t,I_1^c}\mathbf{U}'_{t,I_1^c}\, dt, \int_0^1 \mathbf{U}_{t,I_1}\mathbf{U}'_{t,I_1}\, dt\Big).$$

(ii) If $I_1 \neq \emptyset$, *then*

$$\mathbf{L}_n^{-1}\widehat{\mathbf{\Sigma}}_j^x\mathbf{L}_n^{-1} \xrightarrow{d} \text{diag}\left(\int_0^1 \mathbf{U}_{t,I_1^c}\mathbf{U}'_{t,I_1^c}\, dt,\ \text{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1})\right) \quad \textit{and} \tag{7.27}$$

$$\mathbf{L}_n^{-1}\mathbf{\Sigma}_j^x\mathbf{L}_n^{-1} \xrightarrow{d} \text{diag}\left(\int_0^1 \mathbf{U}_{t,I_1^c}\mathbf{U}'_{t,I_1^c}\, dt,\ \text{Cov}(\mathbf{x}_{t+j,I_1}\mathbf{x}_{t,I_1})\right), \tag{7.28}$$

where $\mathbf{L}_n = \text{diag}(l_{n1}, \cdots, l_{np}),\ l_{ni} = n^{d_i-1/2}I(d_i > 1/2) + I(d_i < 1/2)$.

By Lemma 12, Theorems 5 and 6 can be established in a similar manner as to Theorems 1 and 2. Therefore we omit the detailed proofs.

References

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

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.

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

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

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

Chang, J. Y., Guo, B. and Yao, Q. (2014). Segmenting multiple time series by contemporaneous linear transformation. *A manuscript*.

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

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.

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

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.

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

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

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.

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

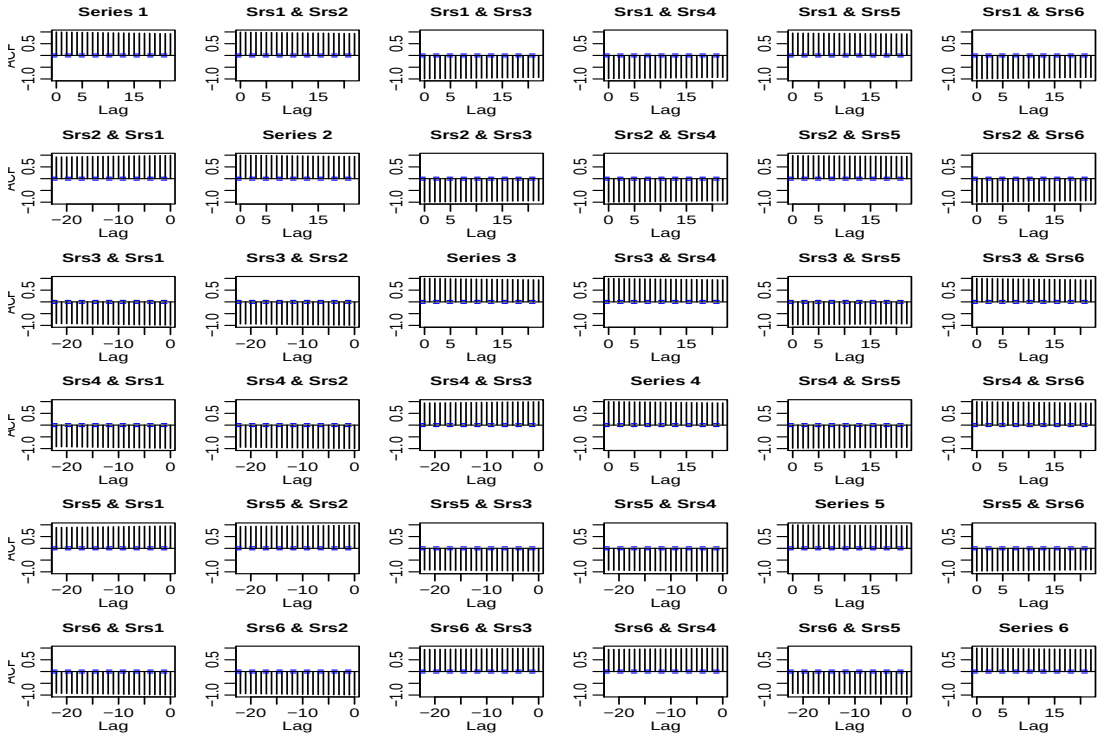

Figure 1: Sample ACF/CCF of $\mathbf{y}_t$ with $n = 1000$, $p = 6$ and $r = 2$ in Example 1.

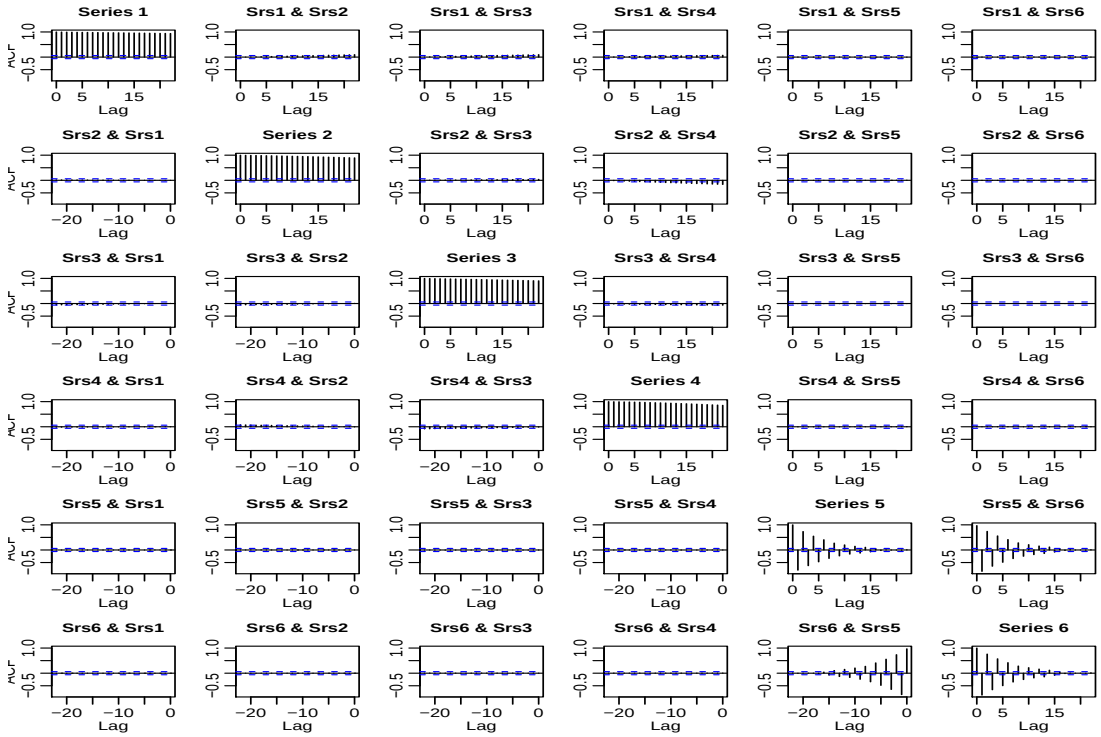

Figure 2: Sample ACF/CCF of $\widehat{\mathbf{x}}_t$ with $n = 1000$, $p = 6$ and $r = 2$ in Example 1.

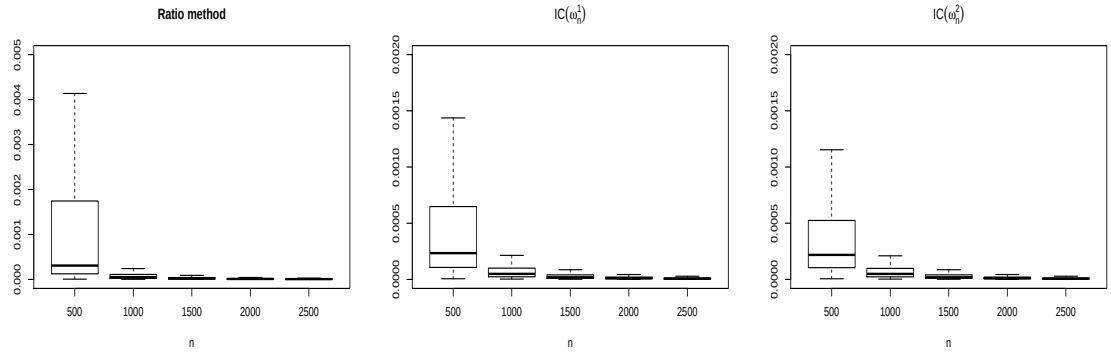

Figure 3: Boxplots of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$ for $p = 6$, $r = 2$ and $500 \leq n \leq 2500$ in Example 1.

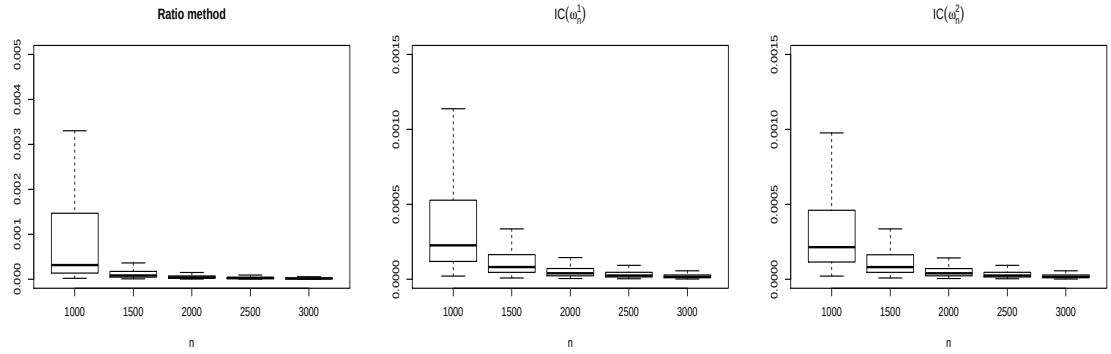

Figure 4: Boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ for $p = 10$, $r = 4$ and $1000 \leq n \leq 3000$ in Example 1.

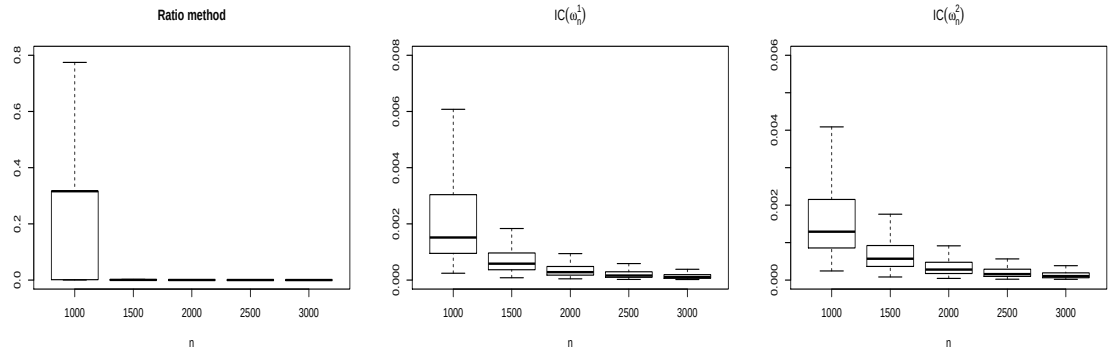

Figure 5: Boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ for $p = 20$, $r = 10$ and $1000 \leq n \leq 3000$ in Example 1.

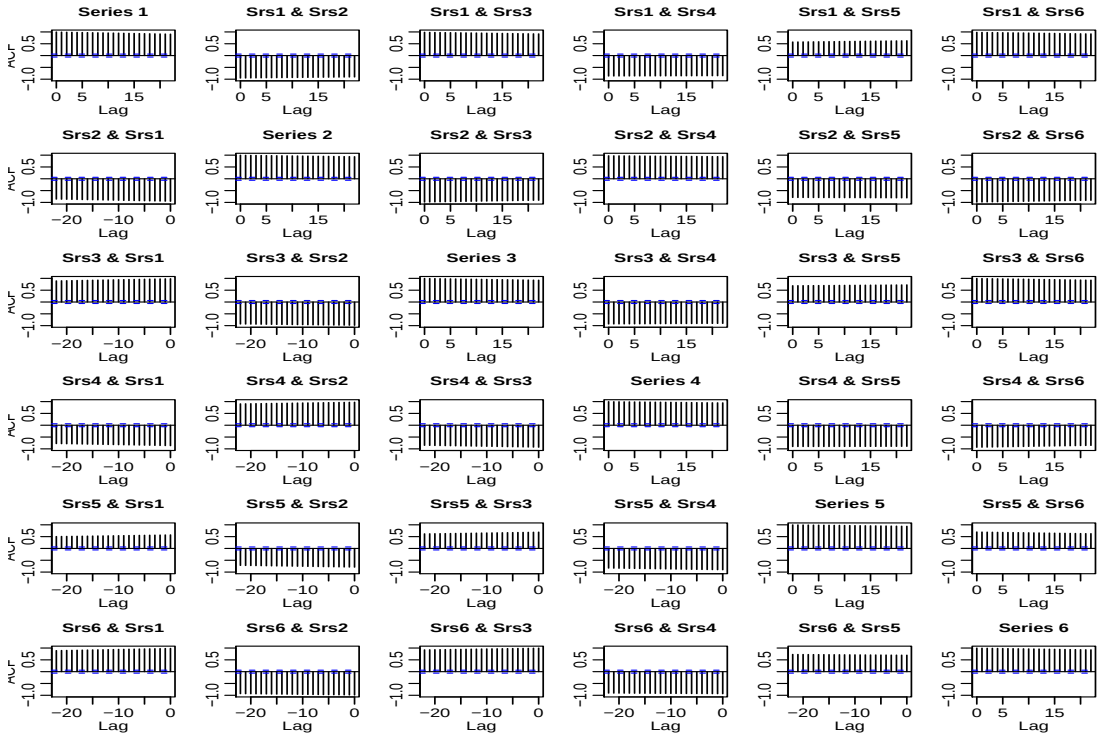

Figure 6: Sample ACF/CCF of $\mathbf{y}_t$ with $n = 1000$, $p = 6$, $r = 2$ and $s = 2$ in Example 2.

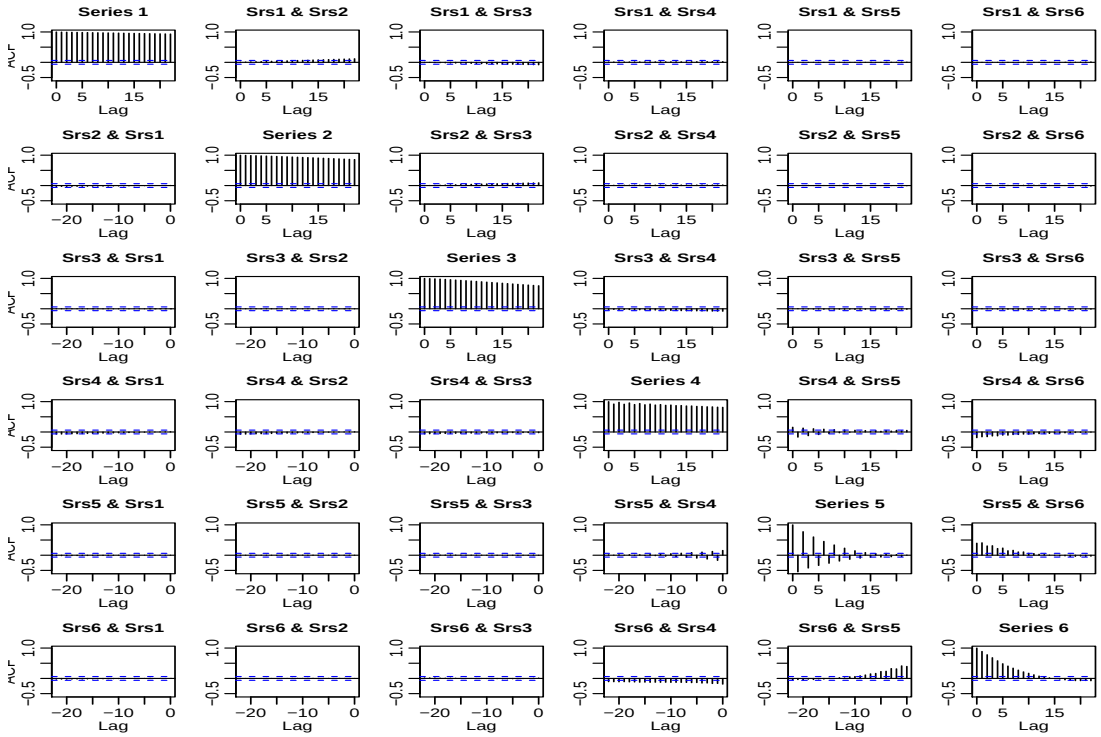

Figure 7: Sample ACF/CCF of $\widehat{\mathbf{x}}_t$ with $n = 1000$, $p = 6$, $r = 2$ and $s = 2$ in Example 2.

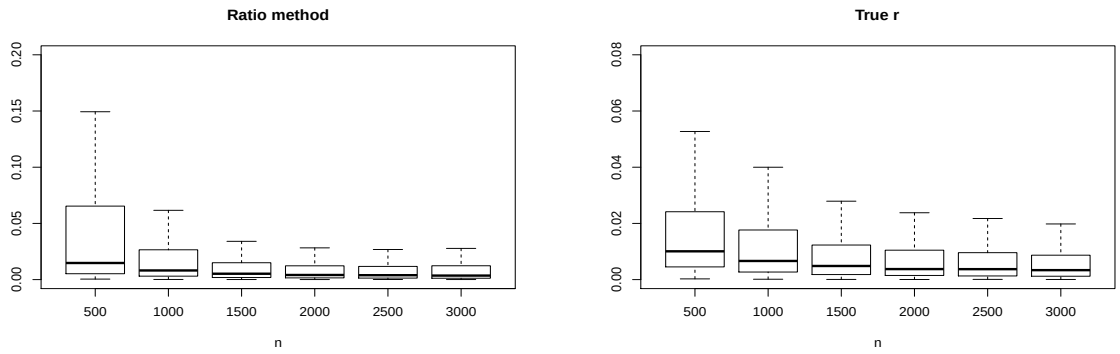

Figure 8: Boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ with the estimated $\widehat{r}$ (left panel) and the true r (right panel) for Example 2, while $p = 6, r = 2, s = 2$ and $500 \leq n \leq 2500$.

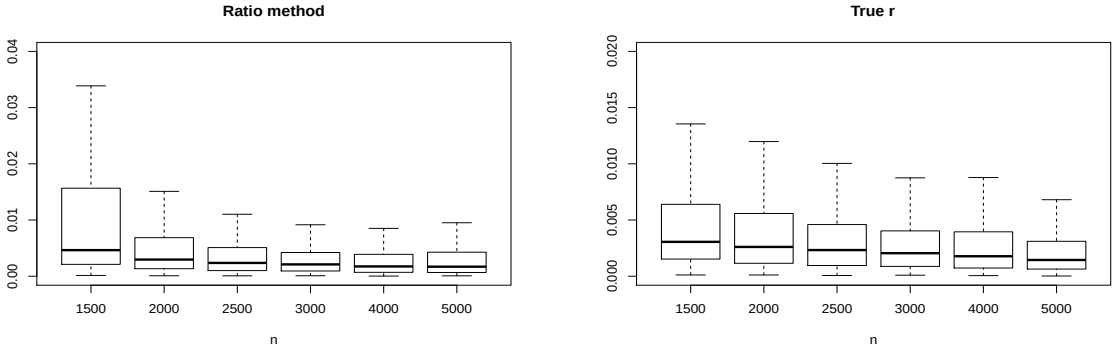

Figure 9: Boxplots of $D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2))$ with the estimated $\widehat{r}$ (left panel) and the true r (right panel) for Example 2, while $p = 10, r = 4, s = 1$ and $1500 \leq n \leq 5000$.

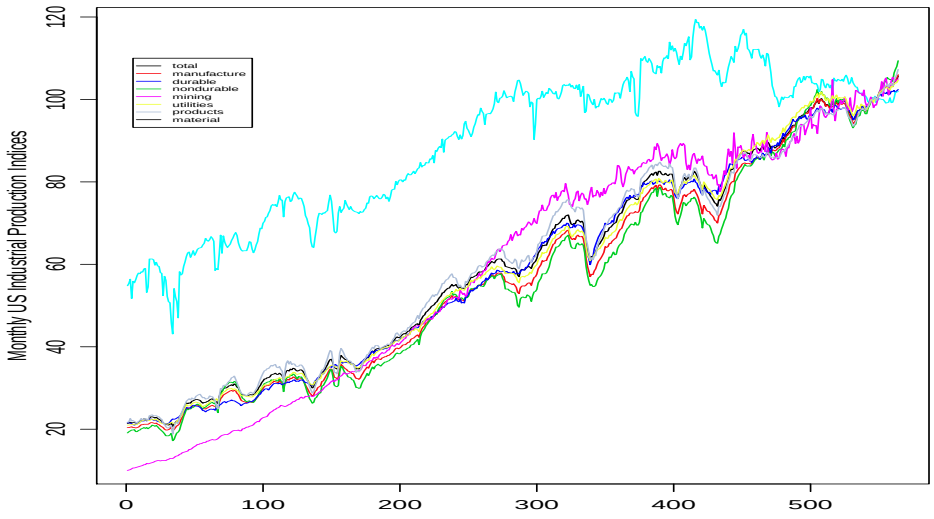

Figure 10: Time series plots of the 8 monthly U.S. Industrial Production indices in January 1947 - December 1993.

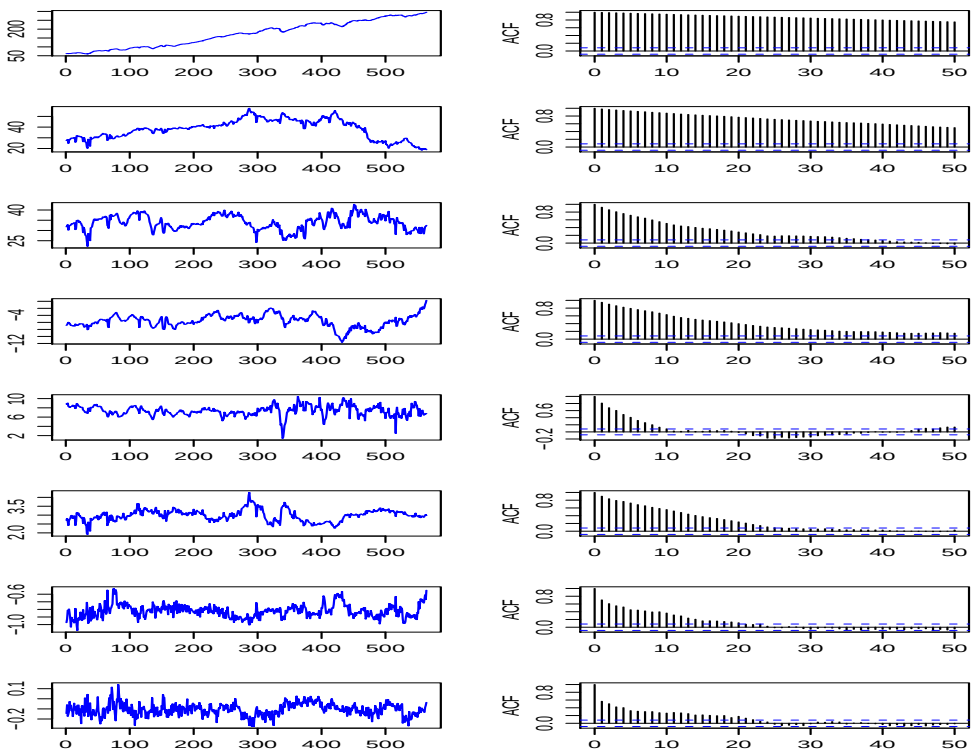

Figure 11: Time series plots of the estimated $\widehat{\mathbf{x}}_t$ by the proposed method and their sample ACF for the 8 monthly U.S. Industrial Production indices.

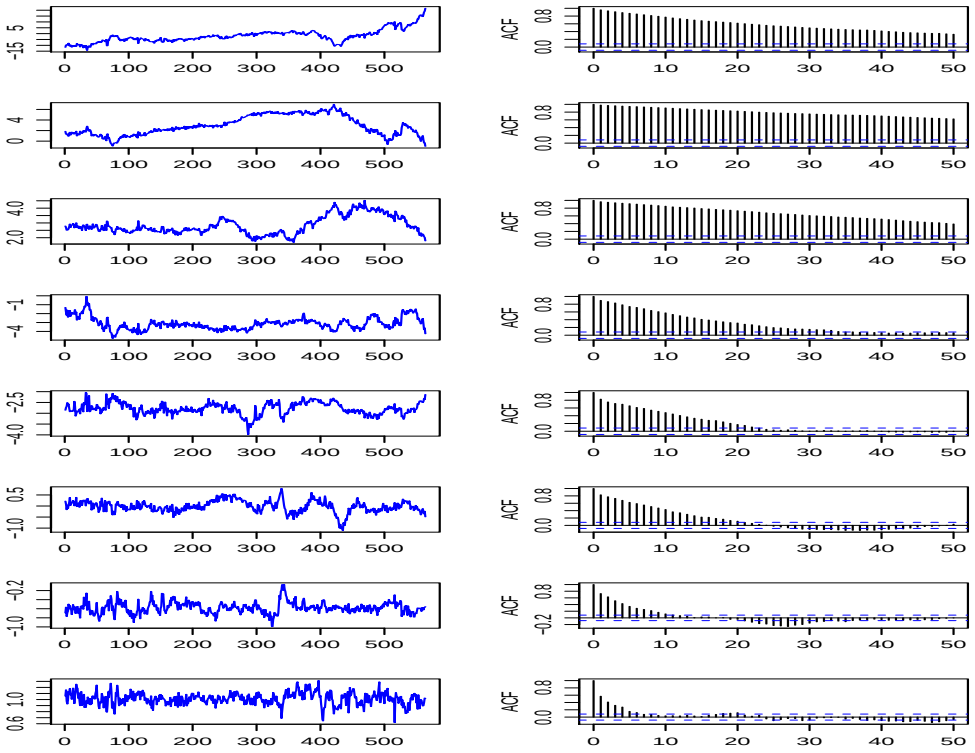

Figure 12: Time series plots of the estimated $\widehat{\mathbf{x}}_t$ by Johansen's method and their sample ACF for the 8 monthly U.S. Industrial Production indices.