COINTEGRATION RANK ESTIMATION FOR HIGH-DIMENSIONAL TIME SERIES WITH BREAKS

Chan, Ngai Hang; Zhang, Rongmao

Published in:
Statistica Sinica

Published: 01/05/2023

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Published version (DOI):
10.5705/ss.202021.0117

Publication details:
Chan, N. H., & Zhang, R. (2023). COINTEGRATION RANK ESTIMATION FOR HIGH-DIMENSIONAL TIME SERIES WITH BREAKS. *Statistica Sinica*, *33*(Online Special Issue), 1193-1217. https://doi.org/10.5705/ss.202021.0117

Chan, N. H., & Zhang, R. (2023). COINTEGRATION RANK ESTIMATION FOR HIGH-DIMENSIONAL TIME SERIES WITH BREAKS. *Statistica Sinica*, 33 (Online Special Issue), 1193-1217. https://doi.org/10.5705/ss.202021.0117

Statistica Sinica **33** (2023), 1193-1217
doi:https://doi.org/10.5705/ss.202021.0117

COINTEGRATION RANK ESTIMATION FOR HIGH-DIMENSIONAL TIME SERIES WITH BREAKS

Ngai Hang Chan[1] and Rongmao Zhang[2,3]

[1]*City University of Hong Kong,*
[2]*Minnan Normal University and* [3]*Zhejiang University*

Abstract: We propose an intuitive and simple-to-use procedure for estimating the cointegration rank of a high-dimensional time series system with possible breaks. Based on a similar idea to a principal component analysis, the cointegration rank can be estimated by the number of eigenvalues of a certain nonnegative definite matrix. There are several advantages to the new method: (a) the dimension of the cointegrated time series is allowed to vary with the sample size; (b) it is model free; and (c) it is simple to use and robust against possible breaks in trend. The cointegration rank can be estimated without the need for a priori testing and estimating of the break points. The asymptotic properties of the proposed methods are investigated when the dimension of the time series increases with the sample size, which offers a new alternative to deal with high-dimensional time series. Finally, the proposed precedure is demonstrated by means of simulations.

Key words and phrases: Cointegration, eigenanalysis, high-dimensional time series, nonstationary processes, structural break.

1. Introduction

Cointegration was introduced to visualize the long-run equilibrium between several nonstationary economic series. During the past thirty years, cointegration analysis has attracted increasing attention from both theoretical and empirical researchers in economics and statistics alike. An excellent survey on the early developments in cointegration can be found in Johansen (1995).

Engle and Granger (1987) derived a representation for cointegration in the form of an error correction model (ECM), which reflects the correction of the long-run relationship with short-run dynamics. One of the remarkable features of the ECM is that it clearly identifies the gain in prediction achieved using the cointegrated variables, rather than the standard ARIMA approach. However, the associated inference method, now known as the Engle–Granger method, is designed for bivariate series only. The likelihood inference-based ECM is sys-

Corresponding author: Rongmao Zhang, School of Mathematics, Minnan Normal University, Fujian, China. E-mail: rmzhang@zju.edu.cn.

tematically developed in Johansen (1995). This likelihood approach is regarded as the most effective and efficient inference method for an ECM. It provides the maximum likelihood estimator for cointegration spaces and likelihood ratio tests for cointegration ranks. However, the likelihood method is complex, requiring a correct specification of a finite order of the vector autoregression (VAR) for the short-run dynamic effect. In many applications, using different orders for the VAR would result in different conclusions on the cointegration. If the VAR order is under-specified or if the process lies outside the VAR class, then an optimal inference on the unknown cointegration relations will lose validity; see Hualde and Robinson (2010). Related methods, including the triangular system of Phillips (1991) and the frequency domain approach of Robinson and Yajima (2002) and Müller and Watson, (2013), suffer similar shortcomings. Furthermore, most of these methods focus mainly on fixed-dimensional cointegrated time series, and become impractical when dealing with high-dimensional series; see Ho and Sorensen (1996) and Gonzalo and Pitarakis (1999) for related discussions.

With the advancement of data acquisition technology, high-dimensional time series data are often encountered in finance and economics, for example, high-risk assets in a large portfolio, or large panel data in economics. A natural question is how to detect a cointegration relationship if there exists a long-run equilibrium among the high-dimensional nonstationary economic series. Recently, Onatski and Wang (2018) derived the asymptotic theory for a high-dimensional setting, and Zhang, Robinson and Yao (ZRY, 2019) proposed a method for estimating the cointegration rank based on the eigenvectors of a certain positive-definite matrix consisting of the sample covariance of the observed series $\{\mathbf{y}_t\}$, given by $\hat{\mathbf{W}} = \sum_{j=0}^{j_0} \hat{\mathbf{\Sigma}}_j \hat{\mathbf{\Sigma}}_j'$, where $\hat{\mathbf{\Sigma}}_j = (1/n)\sum_{t=1}^{n-j}(\mathbf{y}_{t+j} - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})'$ and $\bar{\mathbf{y}} = (1/n)\sum_{t=1}^{n} \mathbf{y}_t$ are the sample autocovariance and sample mean of the observations $\{\mathbf{y}_t : t = 1, \ldots, n\}$, respectively. The symbol H' denotes the transpose of the matrix/vector H. Because ZRY employed a strong approximation technique to establish the asymptotic properties, they have to assume that the dimension of the time series cannot exceed $O(n^{1/4})$. In addition, ZRY did not consider the case with structural breaks in the trend, which can induce an apparent unit root phenomenon; see Zivot and Andrews (1992); Perron (1997); Cavaliere and Georgiev (2007), and the references therein. It has been shown that an unaccounted trend break can lead to substantial over-size and power loss in many standard tests for cointegration; see Campos, Ericsson and Hendry (1996) and Harris, Leybourne and Taylor (2016). Compared with the univariate case, relatively less work has been conducted on cointegration with structural breaks in the trend. Sporadic examples are given in Johansen, Mosconi and Nielsen (2000);

Saikkonen and Lütkepohl (2000); Trenkler, Saikkonen and Lütkepohl (2008). As pointed out by Harris, Leybourne and Taylor (2016), these studies all assume both that a break exists and that the lag order of the VAR model is known a priori. To overcome these difficulties, Harris, Leybourne and Taylor (2016) applied Johansen's likelihood ratio to test for the cointegration rank. To determine the cointegration rank, they further required a test at each step (an estimate for the break point and an order selection using the Schwarz information criterion (SC)), which becomes difficult to implement for high-dimensional series.

The purpose of this study is to propose an approach that is simple to use and robust to possible breaks to identify the cointegration rank in a high-dimensional time series setting. To a certain extent, the proposed procedure is similar to a principal component analysis. In a classical principal component analysis, one looks for linear combinations of factors that give rise to large variation, which can be found by eigenvectors corresponding to large eigenvalues of the sample covariance matrix (see Anderson (1984)). Here, we search for linear combinations of integrated series that become stationary (cointegrated). The eigenvalues of the long-run sample covariance matrix (see $\widehat{\mathbf{W}}_x$ below) of the difference of the stationary components are of smaller order than those of the nonstationary components. Thus, we identify the cointegration space by means of the eigenvectors corresponding to small eigenvalues of the long-run sample covariance matrix of the time series. A similar idea was considered in Phillips and Ouliaris (1988); Harris (1997); Robinson and Yajima (2002) for fixed-dimensional cases, where they used the original series to recover the cointegration space.

Specifically, consider differencing the original series $\{\mathbf{y}_t\}$ to eliminate possible breaks. Note that for a p-dimensional observed series $\mathbf{y}_t$ with a linear trend, the differenced series $\nabla\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ would only have a break at the change point. To elucidate this idea, consider a simple example with one possible break in level (cases with multiple breaks can be handled similarly). Let $\mathbf{y}_t = \mathbf{c}_1 + \mathbf{c}_2 I(t \geq t_0) + \mathbf{d}t + \boldsymbol{X}_t$, where $\mathbf{c}_i$, for $i = 1, 2$, and $\mathbf{d}$ are p-dimensional constant vectors, and $\boldsymbol{X}_t$ is a p-dimensional cointegrated process with no trend. Then,

$$\nabla\mathbf{y}_t = \mathbf{d} + \mathbf{c}_2 I(t = t_0) + \nabla\boldsymbol{X}_t.$$

The trend of $\nabla\mathbf{y}_t$ reduces to a constant, except at the change point $t = t_0$. We can therefore remove the trend by considering $\nabla\mathbf{y}_t - \nabla\bar{\mathbf{y}}$, to recover the cointegration relationship, where $\nabla\bar{\mathbf{y}} = (1/n)\sum_{t=1}^{n}\nabla\mathbf{y}_t$. Because the number of break points is small, using $\nabla\mathbf{y}_t - \nabla\bar{\mathbf{y}}$ to estimate the cointegration rank will have little effect on the performance. The distinct advantage of this method is that it avoids

estimating and testing for the trend with possible breaks.

Furthermore, we propose recovering the cointegration space based on a weighted matrix function of the sample autocovariance of the differenced process. Note that if $\boldsymbol{X}_t$ is cointegrated with rank r, then there exist a $p \times p$ matrix $\mathbf{A}$ and a p-dimensional series $\mathbf{z}_t$ with r stationary components and $p - r$ nonstationary components, such that

$$\boldsymbol{X}_t = \mathbf{A}\mathbf{z}_t, \tag{1.1}$$

(see Zhang, Robinson and Yao (2019)). It follows from (1.1) that the differenced process $\{\nabla \boldsymbol{X}_t\}$ satisfies

$$\nabla \boldsymbol{X}_t = \mathbf{A}\nabla \mathbf{z}_t. \tag{1.2}$$

We adopt a similar approach to that of ZRY to estimate the cointegration space, but based on a different sample covariance matrix. Let $\widehat{\boldsymbol{\Sigma}}_{j,x}$ be the sample autocovariances between $\nabla \boldsymbol{X}_t$ and $\nabla \boldsymbol{X}_{t+j}$, for $j \geq 0$. That is,

$$\widehat{\boldsymbol{\Sigma}}_{j,x} = \frac{1}{n-j} \sum_{i=1}^{n-j} (\nabla \boldsymbol{X}_{t+j} - \nabla \bar{\boldsymbol{X}})(\nabla \boldsymbol{X}_t - \nabla \bar{\boldsymbol{X}})'$$

and

$$\widehat{\mathbf{W}}_x = \sum_{j=0}^{M} w(j)(\widehat{\boldsymbol{\Sigma}}_{j,x} + \widehat{\boldsymbol{\Sigma}}'_{j,x}), \tag{1.3}$$

with $w(0) = 1/2$ and $w(j) = (1 - j/(M+1))$, for $j \geq 1$.

Then, $\widehat{\mathbf{W}}_x$ is a symmetric and positive-definite matrix; see Newey and West (1987). Because the convergent rates of the eigenvalues of $\widehat{\mathbf{W}}_x$ are different for cointegrated and non-cointegrated spaces (see (2.3) and (2.4) below), we can recover the cointegration space using the eigenvectors corresponding to the r smallest eigenvalues of $\widehat{\mathbf{W}}_x$. We also propose two methods for estimating the cointegration rank r: one based on an information criterion, and the other based on the ratio of the eigenvalues of $\widehat{\mathbf{W}}_x$. Moreover, we establish the consistency of the estimated cointegration space and the rank for both fixed and diverging p.

Compared with other procedures based on Engle and Granger (1987) ECM, the proposed method has several advantages. First, it avoids a misspecification or lag order selection for the VAR model of the short-run effect. Second, unlike other results, our method works regardless of the existence of breaks in the trend. Third, we allow the dimension p to diverge with n at a rate faster than that of ZRY.

Throughout the paper, $\|\cdot\|$ denotes the spectral norm of a matrix. In particular, for a matrix $\mathbf{H}$, $\|\mathbf{H}\| = \sqrt{\lambda_{max}(\mathbf{H}'\mathbf{H})}$, where $\lambda_{max}(\cdot)$ denotes the largest

eigenvalue of a matrix. The rest of the paper is organized as follows. The proposed methodology is presented in Section 2. The asymptotic theory is developed in Section 3. Numerical studies are reported in Section 4, and Section 5 concludes the paper. Technical proofs are relegated to the Supplementary Material.

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, that is, $\nabla\mathbf{u}_t = \mathbf{u}_t - \mathbf{u}_{t-1}$. Furthermore, if $\mathbf{u}_t$ has a spectral density matrix that is finite and definitely positive at zero frequency, then we say $\mathbf{u}_t$ is an $I(0)$ process. An example of an $I(0)$ process is a stationary and invertible vector ARMA. We say $\mathbf{x}_t$ is a p-dimensional integrated process of order one (i.e., I(1)) if all its components follow a unit-root process, that is, $\nabla\mathbf{x}_t = \mathbf{u}_t$. Now, assume $\mathbf{y}_t$ is a p-dimensional observable time series with a linear trend and admits the following form: $\mathbf{y}_0 = 0$ and

$$\mathbf{y}_t = \sum_{i=1}^{m+1} \mathbf{a}_i I(t_{i-1} < t \le t_i) + \mathbf{b}t + \mathbf{x}_t, \quad \text{and} \quad \mathbf{x}_t = \mathbf{A}\mathbf{z}_t, \tag{2.1}$$

where $0 = t_0 < t_1 < \cdots < t_{m+1} = n$, $\mathbf{b}$ and $\mathbf{a}_i$ are p-dimensional constant vectors, $\mathbf{A}$ is an unknown and invertible constant matrix, $\mathbf{x}_t$ is a latent p-dimensional $I(1)$ process, $\mathbf{z}_t = (\mathbf{z}_{t1}', \mathbf{z}_{t2}')'$ is a p-dimensional process, $\mathbf{z}_{t2}$ is an r-dimensional $I(0)$ process, and $\mathbf{z}_{t1}$ is a $(p-r)$-dimensional $I(1)$ process. Furthermore, no linear combination of $\mathbf{z}_{t1}$ is $I(0)$, because such a stationary variable can be subsumed into $\mathbf{z}_{t2}$. Each component of $\mathbf{z}_{t2}$ is a cointegrating error of $\mathbf{x}_t$, and $r \ge 0$ is the cointegration rank. In the event that there is no cointegration among the components of $\mathbf{x}_t$, then $r = 0$. When $\mathbf{x}_t$ itself is an $I(0)$ process, then $r = p$. For many economic and financial applications, there exists a small number of cointegrated variables, that is, $r \ge 1$ is a small integer. Expression (2.1) was considered in ZRY for the case without a trend, that is, all the coefficients $\mathbf{a}_i$ and $\mathbf{b}$ are zero.

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

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

where $\mathbf{H}_{11}$ and $\mathbf{H}_{22}$ are square matrices of size $(p-r)$ and r, respectively, and $\mathbf{0}$ denotes a matrix with all entries equal to zero. Therefore, there is no loss of generality in assuming $\mathbf{A}$ to be orthogonal, because any non-orthogonal $\mathbf{A}$ admits the decomposition $\mathbf{A} = \mathbf{QU}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{U}$ is upper-triangular, and we may then replace $(\mathbf{A}, \mathbf{z}_t)$ in (2.1) by $(\mathbf{Q}, \mathbf{U}\mathbf{z}_t)$. In the following, we always assume that $\mathbf{A}$ in (2.1) is orthogonal, that is, $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$, 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 $p \times (p-r)$ and $p \times r$ matrices, respectively. Because $\mathbf{z}_{t2} = \mathbf{A}_2'\mathbf{x}_t$, 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).

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 $\mathbf{z}_{it} = \mathbf{A}_i'\mathbf{x}_t$, for $i = 1, 2$. Our estimation method is motivated by the following observation. To remove the trend effect, we consider the first difference of $\mathbf{y}_t$. From (2.1), it follows that

$$\nabla \mathbf{y}_t = \mathbf{b} + \sum_{i=1}^{m} (\mathbf{a}_{i+1} - \mathbf{a}_i) I(t = t_i + 1) + \nabla \mathbf{x}_t. \tag{2.2}$$

From (2.2), we know that the trend of $\nabla \mathbf{y}_t$ changes only at the break points $t_i + 1$, for $i = 1, 2, \ldots, m$. If we ignore these break points and consider $\nabla \mathbf{y}_t$ as a process with a constant mean, then it can be detrended by subtracting the sample mean, $\nabla \bar{\mathbf{y}} = (1/n) \sum_{t=1}^{n} \nabla \mathbf{y}_t$. That is, consider $\nabla \hat{\mathbf{x}}_t = \nabla \mathbf{y}_t - \nabla \bar{\mathbf{y}}$, and recover the cointegration space based on the sample covariance of $\nabla \hat{\mathbf{x}}_t$. In particular, for $j \geq 0$, we define

$$\hat{\mathbf{\Sigma}}_j = \frac{1}{n-j} \sum_{t=1}^{n-j} \nabla \hat{\mathbf{x}}_{t+j} \nabla \hat{\mathbf{x}}_t',$$

and define $\mathbf{W}$ as in (1.3), that is, let $w(0) = 1/2$ and $w(j) = (1 - j/(M+1))$, for $j \geq 1$, and

$$\hat{\mathbf{W}} = \sum_{j=0}^{M} w(j) (\hat{\mathbf{\Sigma}}_j + \hat{\mathbf{\Sigma}}_j').$$

Owing to possible breaks, $\{\nabla\hat{\mathbf{x}}_t\}$ is not stationary and $\hat{\mathbf{W}}$ is not an exact estimator of the spectral density of $\nabla\mathbf{x}_t$ at frequency zero. However, $\{\nabla\hat{\mathbf{x}}_t\}$ is asymptotically stationary and, consequently, $\hat{\mathbf{W}}$ can still estimate the spectral density consistently at frequency zero. It can be shown that under some mild conditions, for any $\mathbf{c} \in \mathcal{M}(\mathbf{A}_2)$,

$$\mathbf{c}'\hat{\mathbf{W}}\mathbf{c} = O_p\left(\frac{1}{M} + B_{nMp}\right) = o_p(1), \tag{2.3}$$

as $n, M \to \infty$, where B_{nMp} is given in (3.3) below. However, for any $\mathbf{c} \notin \mathcal{M}(\mathbf{A}_2)$,

$$\mathbf{c}'\hat{\mathbf{W}}\mathbf{c} = O_e(1), \tag{2.4}$$

where $U = O_e(V)$ indicates that $P(0 < |U/V| < \infty) \to 1$. Intuitively, the r directions of the cointegration space $\mathcal{M}(\mathbf{A}_2)$ make $|\mathbf{c}'\hat{\mathbf{W}}\mathbf{c}|$ as small as possible. Consequently, $\mathcal{M}(\mathbf{A}_2)$ can be estimated by the linear space spanned by the r eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the r smallest eigenvalues. In addition, $\mathcal{M}(\mathbf{A}_1)$ can be estimated by the linear space spanned by the $(p-r)$ eigenvectors of $\widehat{\mathbf{W}}$ corresponding to the $(p-r)$ largest eigenvalues.

Let $(\hat{\gamma}_1, \ldots, \hat{\gamma}_p)$ be the orthogonal eigenvectors of $\hat{\mathbf{W}}$ that correspond to the eigenvalues arranged in descending order. Define

$$\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2),\ \hat{\mathbf{A}}_1 = (\hat{\gamma}_1, \ldots, \hat{\gamma}_{p-r}),\ \hat{\mathbf{A}}_2 = (\hat{\gamma}_{p-r+1}, \ldots, \hat{\gamma}_p). \tag{2.5}$$

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 sample autocovariance matrices has been used for factor modeling for dimension reduction (see Lam and Yao (2012), and the references therein), and for identifying a cointegration space (ZRY). These papers are based on a quadratic form of the sample autocovariance of the observed series $\mathbf{y}_t$, which requires estimating the trend and testing for possible break points. From (2.2) to (2.5), these difficulties can be circumvented by means of $\nabla\mathbf{y}_t$.

2.3. Determining cointegration ranks

We introduce two criteria to estimate the cointegration rank r based on the estimated eigenvalues of $\widehat{\mathbf{W}}$. Let $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_p \geq 0$ be the eigenvalues of $\widehat{\mathbf{W}}$ and $\nu_n = M^{-1} + B_{nMp}$. From (2.3) and (2.4), $\widehat{\lambda}_i$ converges to some positive constant for all $1 \leq i \leq p-r$, and $\hat{\lambda}_i = O_p(\nu_n)$ for all $p-r < i \leq p$. Hence, for

$1 \le r < p$, $\widehat{\lambda}_i/\nu_n^{1/2} \to \infty$ in probability for all $1 \le i \le p-r$. For $p-r < i \le p$, $\hat{\lambda}_i/\nu_n^{1/2} = o_p(1)$. This leads to estimating r by

$$\hat{r} = \max\left\{j : \frac{\hat{\lambda}_{p+1-j}}{\nu_n^{1/2}} \le 1,\ 1 \le j \le p\right\}, \tag{2.6}$$

which can be viewed as a revised ratio estimate of Lam and Yao (2012). However, Lam and Yao's estimator is not consistent, whereas $\widehat{r}$ is; see Theorem 2.

Alternatively, we may define a so-called information criterion as follows:

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

where the first term is used to fit the cointegrated components, and the second term is the penalty for the lack of fit and $\omega_n \to 0$. Then, r can be estimated by

$$\widetilde{r} = \underset{1 \le l \le p}{\operatorname{argmin}}\, IC(l). \tag{2.7}$$

Note that when $\omega_n = \nu_n^{1/2}$, then $\widetilde{r} = \hat{r}$. The consistency of $\widetilde{r}$ is established in Theorem 2.

3. Asymptotic Properties

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

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

Then, $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) \in [0,1]$, being zero if and only if $\mathcal{M}(\widehat{\mathbf{A}}_2) = \mathcal{M}(\mathbf{A}_2)$, and one if and only if $\mathcal{M}(\widehat{\mathbf{A}}_2)$ and $\mathcal{M}(\mathbf{A}_2)$ are orthogonal. Furthermore, we show that both estimators $\widehat{r}$ and $\widetilde{r}$, defined in (2.6) and (2.7), respectively, are consistent. We consider two asymptotic regimes: (i) p is fixed, while $n \to \infty$, and (ii) $p \to \infty$ as $n \to \infty$.

Put $\mathbf{z}_{t1} = (z_{t,1}, \ldots, z_{t,p-r})'$ and $\mathbf{z}_{t2} = (z_{t,p-r+1}, \ldots, z_{t,p})'$. Under (2.1), $z_{t,j}$ is $I(1)$ for $1 \le j \le p-r$, and $e_{t,j} \equiv \nabla z_{t,j}$ is weakly stationary, that is, an $I(0)$ process. Write $\boldsymbol{e}_{t1} = (e_{t,1}, \ldots, e_{t,p-r})'$ and $\boldsymbol{e}_t = (\boldsymbol{e}_{t1}', \mathbf{z}_{t2}')'$. We introduce the following regularity conditions.

Regularity Conditions

(i) There exists a sequence of independent and identically distributed (i.i.d.) random vectors $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \varepsilon_{t2}, \ldots, \varepsilon_{tl})'$, such that

$$\boldsymbol{e}_t = \sum_{j=-\infty}^{t} \mathbf{C}_{t-j}\boldsymbol{\varepsilon}_j, \tag{3.2}$$

where ε_{ti}, for $1 \le i \le l$, are independent random variables with mean zero and $\max_{1\le i\le l} \mathrm{E}|\varepsilon_{ti}|^4 < \infty$, and $\mathbf{C}_j$, $j \ge 0$ are $p \times l$ matrices satisfying

$$\sum_{j=0}^{\infty} \|\mathbf{C}_j\| < \infty \quad \text{and} \quad \lambda_{min}\left(\sum_{j=0}^{\infty} \mathbf{C}_j\mathbf{C}_j^T\right) > \delta,$$

for some $\delta > 0$, where λ_{min} denotes the smallest eigenvalue of a matrix.

(ii) The bandwidth M defined in (1.3) and the dimensions p satisfy that $M \to \infty$ and

$$B_{nMp} = \frac{M\sqrt{p}\sum_{i=1}^{m} \|\mathbf{a}_{i+1} - \mathbf{a}_i\|}{n} + \sqrt{\frac{M\min(p^2, l^2)}{n}} \to 0 \tag{3.3}$$

as $n \to \infty$.

Remark 1. Model (3.2) includes a large class of linear processes, including the vector ARMA process. The process when $l = p$ with fixed p, (3.2), has been studied by several authors (see, for example, Marinucci and Robinson (2000)). However, little is known about the large p case. Recently, Chen and Wu (2019) considered the process (3.2) for the large p case under the assumption that $\varepsilon_{ti}, t, i \in \mathcal{Z}$ are i.i.d. random variables, and assume that $\sum_{j=0}^{\infty} \|\mathbf{C}_j\|_F < \infty$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. Note that $\sum_{j=0}^{\infty} \|\mathbf{C}_j\| < \infty$ is weaker than $\sum_{j=0}^{\infty} \|\mathbf{C}_j\|_F < \infty$, and is sufficient to ensure that $\sum_{k=0}^{\infty} \|\mathrm{E}(\boldsymbol{e}_k\boldsymbol{e}_0^T)\| < \infty$, which implies that the important measure of the stability (the spectral density) of $\boldsymbol{e}_t$ exists and is bounded and continuous (Basu and Michailidis (2015)). The assumption of the independence of the elements of $\boldsymbol{\varepsilon}_t$ is imposed for simplicity in presenting the proof. When $l \le p$, the independent assumption can be replaced by $\mathrm{E}(\|\varepsilon_t\varepsilon_t^T\|^2) = O(l)$.

Remark 2. When $\sum_{j=0}^{\infty} \|\mathbf{C}_j\|_F < \infty$, we can replace B_{nMp} by

$$B^*_{nMp} = \frac{M\sqrt{p}\sum_{i=1}^{m} \|\mathbf{a}_{i+1} - \mathbf{a}_i\|}{n} + \sqrt{\frac{M}{n}} \to 0.$$

Remark 3. It is easy to see that B_{nMp} in Condition 1(ii) satisfies $B_{nMp} = O(\sqrt{Mp^2/n})$. Furthermore, when l is finite, then Condition (ii) reduces to $M\sqrt{p}\sum_{i=1}^{m}\|\mathbf{a}_{i+1}-\mathbf{a}_i\|/n \to 0$, which depends on the jump size of the common breaks $\sum_{i=1}^{m}\|\mathbf{a}_{i+1}-\mathbf{a}_i\|$. If only a finite number of components of $\mathbf{y}_t$ have breaks in level at each break point, then $\sum_{i=1}^{m}\|\mathbf{a}_{i+1}-\mathbf{a}_i\| = O(1)$ and Condition (ii) is equivalent to $M\sqrt{p}/n \to 0$. As a result, p is allowed to be bigger than the sample size n when $M = o(n^{1/2})$. When $l \geq p$, because the number m of break points is always finite, Condition (ii) reduces to $Mp^2/n \to 0$. In all cases, under the fourth moment condition and $M = o(n^{1/4})$, as in Theorem 2 of Newey and West (1987), p is allowed to be larger than $O(n^{3/8})$, which is much larger than that of ZRY.

Theorem 1. *Let r be given. Under Condition* 1*, we have*

$$\|\hat{\mathbf{A}} - \mathbf{A}\| = O_p(B_{nMp}). \tag{3.4}$$

For the eigenvalues $\widehat{\lambda}_1 \geq \cdots \geq \lambda_p$ of $\widehat{\mathbf{W}}$, we have

$$\lambda_{p-r} = O_e(1) \quad \textit{and} \quad \lambda_{p-r+1} = O_p\left(\frac{1}{M} + B_{nMp}\right). \tag{3.5}$$

For the cointegration space, we have

$$D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_p(B_{nMp}). \tag{3.6}$$

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

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

(ii) For $\widetilde{r}$ defined in (2.7)*, $\lim_{n\to\infty} P(\,\widetilde{r} = r\,) = 1$ provided $\omega_n \to 0$ and $\omega_n/\nu_n \to \infty$.*

Remark 4. Theorem 1 shows that as $B_{nMp} \to 0$ and $M \to \infty$, one can recover the cointegration space consistently without knowing a priori if breaks exist. Theorem 2 shows that when the cointegration rank r is unknown, then one can estimate it consistently based on the ratio or information criterion of the eigenvalues of $\widehat{\mathbf{W}}$, without estimating and testing for possible breaks. In addition, the conclusions of Theorem 1 still hold by replacing r with its estimator. Hence, the proposed method is much simpler than the others, and is applicable to high-dimensional cointegrated series.

4. Numerical Results

We illustrate the proposed method using five simulated examples. A comparison with Johansen (1991) likelihood method is also conducted in Example 1. Examples 2–4 illustrate the method in the presence of a break point, in which we also compare the proposed method with that based on a unit-root test, such as the sample autocorrelation function (ACF) method of ZRY (see also Zhang and Chan (2018)) and the Phillips–Perron test (PP.test). Example 5 applies the method to the case with breaks in a linear trend. We also apply the method to the monthly exchange rates of the U.S. dollar versus the currencies of 19 OECD countries in Example 6. In view of Condition 1(ii), the dimension p is usually no bigger than $n^{1/2}$. Therefore, we consider only $p \leq 80$ in these examples, which is slightly bigger than the square root of the sample size $n = 2500$.

Example 1. To facilitate the computation, in model (2.1), let $\mathbf{a}_1 = \cdots = \mathbf{a}_m$ be a p-dimensional vector with each component independently generated from $U(-1, 1)$. Likewise, let $\mathbf{b}$ be a p-dimensional vector with each component independently generated from $U(1, 2)$, and let $\mathbf{z}_{t2}$ consist of r stationary AR(1) processes with coefficients generated from $U(-0.5, 0.5)$. Let the remaining $p - r$ components of $\mathbf{z}_{t1}$ be ARIMA(1,1,1) models with AR and MA coefficients generated independently from $U(0, 0.6)$ and $U(0, 0.8)$, respectively. All innovations are independent $N(0, 1)$ random variables. Such a setting ensures all regularity conditions are valid. Let $\mathbf{A}$ be an orthogonalized version of a matrix with elements generated independently from $U(-3, 3)$. We estimate the cointegration rank r by (2.7), with $\omega_n = \nu_n^{1/3}$, $\nu_n = M^{-1} + \bar{B}_{nMP}$, and $M = 50$ in $\widehat{\mathbf{W}}$. For each setting, we replicate the exercise 500 times. We then estimate $\widehat{\mathbf{A}}$ by (2.5). Because $\widetilde{r}$ or $\widehat{r}$ is not necessarily equal to r, we extend the definition of the discrepancy measure (3.1) as follows:

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

When $\widehat{r} = r$, $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$, defined in (3.1). The relative frequencies (RFs) for the occurrence of the event $\{\widehat{r} = r\}$ and the average value of $D_1 = D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ for 500 replications are listed in Table 1 under the label IC.

Because the cointegration rank is the number of stationary components of $\mathbf{z}_t$, one can also estimate the rank using a unit-root test (see Engle and Granger (1987)). We also include the results of the unit-root test estimation introduced

by Zhang and Chan (2018) (see also ZRY) in Table 1 under the label ACF. In particular, let $\hat{\mathbf{x}}_t = \mathbf{y}_t - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}t$ and

$$\widehat{\mathbf{z}}_t = (\widehat{z}_{t,1}, \ldots, \widehat{z}_{t,p})' = \widehat{\mathbf{A}}'\hat{\mathbf{x}}_t, \tag{4.2}$$

where $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})$ is given by

$$(\widehat{\mathbf{a}}, \widehat{\mathbf{b}}) = \underset{(\mathbf{a},\mathbf{b})}{\operatorname{argmin}} \sum_{t=1}^{n} (\mathbf{y}_t - \mathbf{a} - \mathbf{b}t)'(\mathbf{y}_t - \mathbf{a} - \mathbf{b}t). \tag{4.3}$$

Note that the order of the components inversely reflects the closeness to stationarity of the component series, with the last component $\{\widehat{z}_{t,p}\}$ most likely to be stationary, and the first component $\{\widehat{z}_{t,1}\}$ most likely to be an $I(1)$ process. Let $\widehat{\rho}_i(\cdot)$ be the ACF of $\widehat{z}_{t,i}$ given by

$$\widehat{\rho}_i(k) = \frac{\widehat{\gamma}_i(k)}{\widehat{\gamma}_i(0)},$$

where $\widehat{\gamma}_i(k) = (1/n)\sum_{t=1}^{n-k} \widehat{z}_{t+k,i}\widehat{z}_{t,i}$, for $i = 1, 2, \ldots, p$, and define

$$T_{n,i} = \frac{n}{q(q+1)} \sum_{k=1}^{q} (\widehat{\rho}_i(k) - 1).$$

We estimate the cointegration rank r by

$$\widehat{r}^* = \sum_{i=1}^{p} I\{T_{n,i} < -a_n\}, \tag{4.4}$$

where $a_n \to \infty$ and R is a prescribed upper bound of r. In our numerical experiments, we set $a_n = (n/q)^{1/2}$, $q = [n^{1/4}]$.

To illustrate the performance of the proposed method, in Table 1, we also report 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 levels 0.05 and 0.01, respectively, written as Jo(0.05) and Jo(0.01) in Table 1. The null distribution of the trace test statistic is approximated by

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

where $\boldsymbol{e}_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,p-r})'$, $\boldsymbol{X}_0 = 0$ and $\boldsymbol{X}_t = \sum_{j=1}^{t} \boldsymbol{e}_t$, and $\{\varepsilon_{t,i}\}$ are independent $N(0,1)$; see Johansen and Juselius (1990). This approximated distribution

is calculated by simulation with $T = 1000$ and 6000 replications. Owing to the heavy computational burden, we consider only the cases for $p \leq 30$.

Table 1 clearly indicates that the newly proposed method outperforms Johansen's method. The estimators $\widetilde{r}$ and $\widehat{r}^*$ defined in (2.7) and (4.4), respectively, achieve higher RFs for attaining the true value r than those achieved by the trace test, with significance level at either 0.05 or 0.01. For small n, the situation becomes more challenging when p and r increase. For all $p \leq 30$, the new method works reasonably well, even when $n = 300$, and almost perfectly when $n \geq 500$. However, Johansen's method, which is not designed for large p, fails for $p \geq 20$, even when $n = 2000$.

Example 2. Consider model (2.1) with a linear trend and one break point, that is,

$$\mathbf{y}_t = \mathbf{a}_1 I(0 < t \leq t_1) + \mathbf{a}_2 I(t_1 < t \leq n) + \mathbf{b}t + \mathbf{x}_t,$$
$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t,\ t = 1, 2, \ldots, n,$$

where the change point t_1 is taken as $[n/2]$, the ith components of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}$ are taken as $a_{1i} = i/p$, $a_{2i} = 2 + i/p$, and $b_i = 0.5$, respectively, and $\mathbf{z}_{t2}$ consists of r stationary AR(1) processes with coefficients $-0.4 + i/r\,(i = 1, \ldots, r)$. Let the $p - r$ components of $\mathbf{z}_{t1}$ be an ARIMA(1,1,1) with autoregressive coefficients generated independently from $U(0, 0.6)$ and moving average coefficients $0.2 + 0.6i/(p - r)\,(i = 1, \ldots, p - r)$, and let $\mathbf{A}$ be generated as in Example 1.

We estimate the cointegration rank r by (2.7) with $\omega_n = \nu_n^{1/3}$ and (4.4) with $a_n = (n/q)^{1/2}$, $q = [n^{1/4}]$. We still take $M = 50$ in $\widehat{\mathbf{W}}$, and replicate the exercise 500 times for each setting. The RFs for the occurrence of the event $\{\widetilde{r} = r\}$ or $\{\widehat{r} = r\}$ and the average value of $D_1 = D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ over 500 replications for $t_1 = [n/2]$ are listed in Table 2 under the labels IC and ACF, respectively. Table 2 shows that both proposed procedures work reasonably well when p is small and the sample size n is relatively large. When p is large, for example $p \geq 50$, the ACF method works poorly, but the proposed method can still estimate the cointegration rank and the cointegration space very well. This confirms that ZRY's method works for smaller p than the proposed method does.

Also included in Table 2 are the results of applying the Phillips–Perron unit-root test (PP.test) to the process $\{\widehat{\mathbf{z}}_t\}$ defined in (4.2), with the significance level set at 0.01, for estimating r. When p is large and n is small, the PP.test estimates r slightly better than the procedures (4.4) do, but not as well as (2.7) does. In all cases, (2.7) always performs better than the other methods. Table 2 also shows that for a given p, a larger r/p yields more accurate estimates for r. In

Table 1. RFs of $\{\hat{r} = r\}$ and the average distance D_1 between the estimated and the true cointegration spaces (see (4.1)) in the simulation with 500 replications for Example 1.

		$n = 300$		$n = 500$		$n = 1000$		$n = 1500$		$n = 2000$	
	Method	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
	Jo(0.05)	0.170	0.656	0.228	0.606	0.424	0.456	0.408	0.472	0.266	0.571
p=6	Jo(0.01)	0.406	0.467	0.460	0.420	0.692	0.258	0.666	0.285	0.536	0.362
r=1	*IC*	0.998	0.071	0.998	0.060	1.000	0.052	0.998	0.064	1.000	0.045
	ACF	0.888	0.141	0.972	0.077	0.990	0.058	0.996	0.065	1.000	0.045
	Jo(0.05)	0.484	0.310	0.448	0.323	0.376	0.365	0.450	0.333	0.534	0.260
p=6	Jo(0.01)	0.730	0.185	0.706	0.191	0.646	0.222	0.706	0.200	0.796	0.126
r=3	*IC*	1.000	0.067	1.000	0.056	1.000	0.063	1.000	0.066	1.000	0.029
	ACF	0.940	0.093	0.982	0.064	0.980	0.074	0.826	0.160	1.000	0.029
	Jo(0.05)	0.188	0.489	0.372	0.370	0.260	0.437	0.216	0.470	0.346	0.383
p=8	Jo(0.01)	0.404	0.352	0.694	0.207	0.536	0.283	0.484	0.310	0.610	0.233
r=3	*IC*	1.000	0.080	1.000	0.066	1.000	0.067	1.000	0.060	1.000	0.047
	ACF	0.788	0.170	0.922	0.101	0.912	0.112	0.924	0.100	0.994	0.050
	Jo(0.05)	0.178	0.371	0.112	0.399	0.102	0.393	0.148	0.373	0.132	0.387
p=12	Jo(0.01)	0.584	0.224	0.466	0.253	0.452	0.247	0.530	0.211	0.448	0.242
r=6	*IC*	0.980	0.104	1.000	0.075	1.000	0.058	1.000	0.040	1.000	0.041
	ACF	0.622	0.210	0.906	0.105	0.942	0.078	0.982	0.046	0.988	0.045
	Jo(0.05)	0.132	0.289	0.102	0.286	0.148	0.263	0.294	0.210	0.232	0.237
p=20	Jo(0.01)	0.474	0.202	0.456	0.187	0.554	0.155	0.748	0.097	0.596	0.134
r=14	*IC*	0.970	0.113	0.996	0.071	1.000	0.054	1.000	0.039	1.000	0.039
	ACF	0.652	0.169	0.928	0.084	0.964	0.061	0.990	0.041	0.992	0.041
	Jo(0.05)	0.134	0.314	0.074	0.297	0.170	0.249	0.066	0.283	0.106	0.265
p=30	Jo(0.01)	0.330	0.272	0.254	0.239	0.436	0.186	0.234	0.221	0.340	0.193
r=20	*IC*	0.694	0.240	0.944	0.145	1.000	0.096	1.000	0.079	1.000	0.067
	ACF	0.072	0.320	0.498	0.203	0.832	0.122	0.854	0.106	0.896	0.087

Herein, Jo(0.05) and Jo(0.01) are based on Johansen's method with levels 0.05 and 0.01, respectively, IC is based on the proposed method, and ACF is based on (4.4) by ZRY; r is the cointegration rank.

general, (2.7) performs more stably than the unit-root test procedures based on the PP.test and (4.4).

To give some insight into the underpinning logic, we present box plots of $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ based on (2.7) and the PP.test for $(p, r) = (6, 2)$ and $(8, 4)$ in Figure 1 and Figure 2, respectively. The two figures show that both the information criterion and the PP.test work well, and the estimation errors decrease as the sample size n increases.

Table 2. RFs of $\{\hat{r} = r\}$ and the average distance D_1 between the estimated and the true cointegration spaces in a simulation with 500 replications for Example 2.

		$n = 300$		$n = 500$		$n = 1000$		$n = 1500$		$n = 2000$		$n = 2500$	
(p, r)	Method	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
	IC	0.998	0.092	0.998	0.084	1.000	0.067	1.000	0.054	1.000	0.051	1.000	0.055
$(6, 2)$	ACF	0.880	0.151	0.936	0.119	0.950	0.098	0.988	0.061	0.968	0.071	0.952	0.085
	PP.test	0.964	0.112	0.936	0.121	0.962	0.091	0.996	0.057	0.984	0.060	0.976	0.068
	IC	0.980	0.088	0.998	0.066	1.000	0.063	1.000	0.038	1.000	0.060	1.000	0.070
$(6, 3)$	ACF	0.892	0.129	0.964	0.083	0.908	0.110	0.996	0.040	0.814	0.161	0.754	0.202
	PP.test	0.916	0.131	0.972	0.084	0.954	0.092	0.992	0.042	0.914	0.117	0.928	0.118
	IC	0.998	0.092	1.000	0.076	1.000	0.066	1.000	0.050	1.000	0.054	1.000	0.053
$(8, 3)$	ACF	0.826	0.168	0.930	0.107	0.948	0.092	0.990	0.055	0.970	0.070	0.842	0.137
	PP.test	0.918	0.138	0.976	0.091	0.958	0.089	0.994	0.053	0.994	0.057	0.944	0.088
	IC	0.996	0.075	1.000	0.060	1.000	0.049	1.000	0.055	1.000	0.041	1.000	0.029
$(8, 4)$	ACF	0.896	0.113	0.970	0.072	0.980	0.057	0.956	0.075	0.996	0.042	1.000	0.029
	PP.test	0.986	0.081	0.998	0.061	0.990	0.053	0.982	0.065	0.992	0.044	0.996	0.031
	IC	1.000	0.101	1.000	0.074	1.000	0.050	1.000	0.051	1.000	0.043	1.000	0.033
$(12, 6)$	ACF	0.762	0.170	0.918	0.100	0.990	0.053	0.960	0.065	0.980	0.050	0.998	0.034
	PP.test	0.966	0.117	0.982	0.087	0.994	0.052	0.992	0.056	0.986	0.049	0.992	0.036
	IC	0.996	0.073	1.000	0.054	1.000	0.037	1.000	0.032	1.000	0.028	1.000	0.031
$(12, 8)$	ACF	0.838	0.115	0.976	0.060	0.998	0.038	0.994	0.034	0.998	0.029	0.986	0.035
	PP.test	0.996	0.073	0.994	0.058	0.994	0.039	0.998	0.033	0.994	0.030	0.992	0.033
	IC	0.942	0.125	0.996	0.080	1.000	0.055	1.000	0.045	1.000	0.043	1.000	0.037
$(20, 14)$	ACF	0.604	0.180	0.894	0.099	0.964	0.064	0.960	0.054	0.940	0.057	0.942	0.052
	PP.test	0.908	0.152	0.990	0.083	0.990	0.059	0.986	0.054	0.978	0.054	0.990	0.045
	IC	0.826	0.226	0.988	0.151	1.000	0.102	1.000	0.082	1.000	0.071	1.000	0.059
$(30, 20)$	ACF	0.114	0.315	0.386	0.237	0.696	0.155	0.796	0.121	0.772	0.121	0.894	0.081
	PP.test	0.532	0.330	0.642	0.300	0.878	0.158	0.934	0.112	0.928	0.106	0.984	0.065
	IC	0.128	0.448	0.520	0.355	0.996	0.233	1.000	0.189	1.000	0.143	1.000	0.134
$(50, 30)$	ACF	0.018	0.481	0.084	0.419	0.048	0.441	0.046	0.435	0.118	0.326	0.030	0.427
	PP.test	0.014	0.766	0.014	0.775	0.048	0.690	0.140	0.587	0.490	0.361	0.310	0.478
	IC	0.160	0.505	0.064	0.459	0	0.371	0.392	0.295	0.930	0.245	0.998	0.223
$(80, 50)$	ACF	0	0.539	0.052	0.474	0.044	0.461	0.002	0.544	0	0.589	0	0.660
	PP.test	0.016	0.675	0	0.799	0.004	0.796	0.002	0.812	0	0.800	0.002	0.823

IC is based on the proposed method, ACF is based on (4.4) by ZRY, the PP.test is based on the Phillips–Perron unit-root test, and r is the cointegration rank.

Example 3. Consider model (2.1) with one change point, that is,

$$\mathbf{y}_t = \mathbf{a}_1 I(0 < t \le t_1) + \mathbf{a}_2 I(t_1 < t \le n) + \mathbf{b}t + \mathbf{A}\mathbf{z}_t,$$

where the change point t_1 is taken as $[n/2]$, and the coefficients $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{b}$, and $\mathbf{z}_{t2}$ are taken as in Example 2. We change only the autoregressive coefficients in the components of $\mathbf{z}_{t1}$, which are taken as $0.3 + (-1)^i 0.5i/(p-r)$, for $i = 1, \ldots, p-r$. The cointegration rank r is estimated as in Example 2. Based on 500 replications, we report the RFs for the occurrence of the event $\{\widetilde{r} = r\}$ or $\{\widehat{r} = r\}$ and the average value of $D_1 = D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ in Table 3. Table 3 shows that the

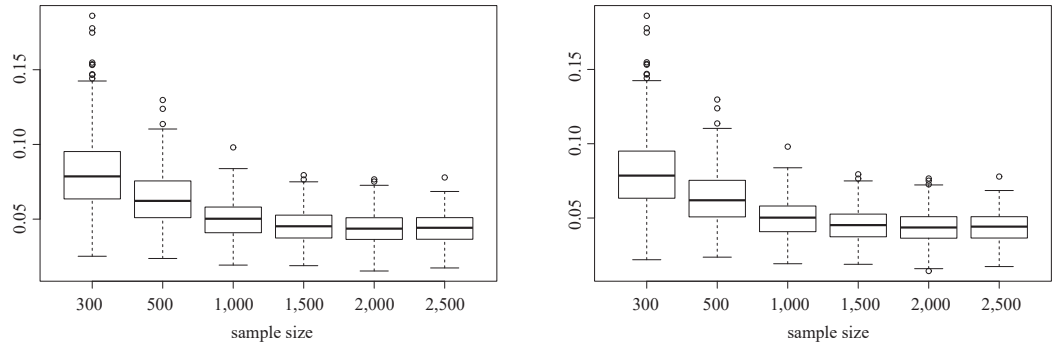

Figure 1. The left panel is the box plot of the distance $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ based on the proposed method, and the right panel is based on the PP.test when $p = 6, r = 2$ in Example 2.

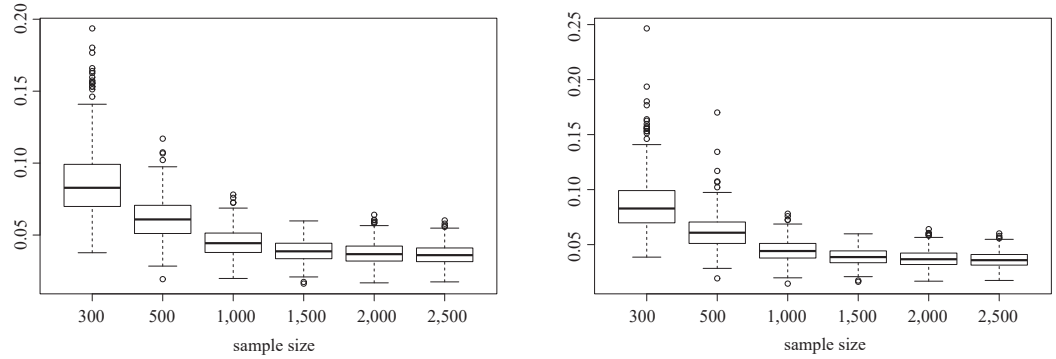

Figure 2. The left panel is the box plot of the distance $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ based on the proposed method, and the right panel is based on the PP.test when $p = 8, r = 4$ in Example 2.

proposed procedure based on (2.7) works reasonably well, especially when p is small or when the sample size n is large. Furthermore, together with Example 2, it can be seen that the unit-root procedures based on the PP.test or (4.4) are not stable and are affected by the coefficients of the nonstationary process. However, (2.7) performs reasonably well, in general.

To demonstrate that the proposed method is more robust than that based on the unit-root test, we also present box plots of $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ based on (2.7) and the PP.test for $(p, r) = (8, 4)$ and $(12, 6)$ in Figure 3 and Figure 4, respectively. From these figures, it can be seen that the estimation errors decrease as the sample size n increases for both procedures. However, the information criterion always performs better than the PP.test.

Table 3. RFs of $\{\hat{r} = r\}$ and the average distance D_1 between the estimated and the true cointegration spaces in the simulation with 500 replications for Example 3.

		$n = 300$		$n = 500$		$n = 1000$		$n = 1500$		$n = 2000$		$n = 2500$	
(p, r)	Method	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
	IC	0.994	0.120	0.998	0.099	1.000	0.088	1.000	0.085	1.000	0.082	1.000	0.083
$(6, 2)$	ACF	0.454	0.430	0.398	0.461	0.282	0.529	0.266	0.539	0.208	0.575	0.158	0.607
	PP.test	0.412	0.461	0.428	0.443	0.370	0.474	0.358	0.480	0.314	0.507	0.264	0.539
	IC	0.990	0.092	1.000	0.071	1.000	0.059	1.000	0.054	1.000	0.052	1.000	0.051
$(6, 3)$	ACF	0.840	0.157	0.896	0.123	0.850	0.137	0.820	0.147	0.786	0.169	0.718	0.207
	PP.test	0.936	0.127	0.946	0.106	0.906	0.123	0.912	0.116	0.898	0.122	0.872	0.139
	IC	0.996	0.116	1.000	0.087	1.000	0.071	1.000	0.063	1.000	0.060	1.000	0.060
$(8, 3)$	ACF	0.692	0.244	0.780	0.197	0.720	0.221	0.668	0.244	0.600	0.282	0.562	0.308
	PP.test	0.848	0.206	0.836	0.188	0.830	0.185	0.828	0.179	0.804	0.192	0.770	0.216
	IC	0.998	0.105	1.000	0.087	1.000	0.079	1.000	0.075	1.000	0.074	1.000	0.072
$(8, 4)$	ACF	0.368	0.415	0.258	0.489	0.148	0.579	0.116	0.602	0.076	0.657	0.080	0.658
	PP.test	0.396	0.537	0.336	0.571	0.264	0.636	0.236	0.652	0.192	0.690	0.214	0.673
	IC	0.998	0.126	1.000	0.097	1.000	0.076	1.000	0.069	1.000	0.066	1.000	0.064
$(12, 6)$	ACF	0.384	0.358	0.330	0.395	0.182	0.485	0.130	0.534	0.080	0.581	0.052	0.622
	PP.test	0.418	0.509	0.418	0.521	0.324	0.588	0.304	0.615	0.250	0.667	0.248	0.665
	IC	0.992	0.095	1.000	0.075	1.000	0.063	1.000	0.058	1.000	0.056	1.000	0.056
$(12, 8)$	ACF	0.398	0.322	0.292	0.381	0.168	0.456	0.120	0.477	0.084	0.513	0.060	0.561
	PP.test	0.468	0.485	0.378	0.570	0.312	0.635	0.254	0.690	0.242	0.700	0.206	0.740
	IC	0.894	0.161	0.996	0.112	1.000	0.079	1.000	0.067	1.000	0.061	1.000	0.057
$(20, 14)$	ACF	0.284	0.331	0.306	0.330	0.212	0.366	0.182	0.396	0.128	0.437	0.088	0.471
	PP.test	0.340	0.490	0.380	0.489	0.382	0.530	0.360	0.572	0.290	0.641	0.270	0.647
	IC	0.706	0.274	0.982	0.196	1.000	0.135	1.000	0.111	1.000	0.096	1.000	0.087
$(30, 20)$	ACF	0.178	0.380	0.132	0.402	0.120	0.405	0.074	0.423	0.068	0.431	0.082	0.438
	PP.test	0.112	0.624	0.160	0.582	0.258	0.513	0.250	0.546	0.250	0.559	0.290	0.539
	IC	0.080	0.466	0.230	0.392	0.988	0.281	0.998	0.231	1.000	0.202	1.000	0.182
$(50, 30)$	ACF	0.104	0.492	0.068	0.520	0.002	0.603	0.002	0.613	0	0.618	0.002	0.620
	PP.test	0.006	0.812	0	0.835	0.002	0.835	0.010	0.787	0.010	0.765	0.014	0.750
	IC	0.312	0.510	0.024	0.470	0.036	0.390	0.662	0.327	0.978	0.288	0.998	0.263
$(80, 50)$	ACF	0.002	0.521	0.092	0.510	0	0.633	0	0.691	0	0.724	0	0.740
	PP.test	0	0.833	0.002	0.828	0	0.890	0	0.892	0	0.881	0	0.876

IC is based on the proposed method, ACF is based on (4.4) by ZRY, the PP.test is based on the Phillips–Perron unit-root test, and r is the cointegration rank.

Example 4. Now, we consider an example in which the components of $\mathbf{y}_t$ are $I(1)$ with a constant trend and one break point, that is,

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{a}_1 I(0 < t \le t_1) + \boldsymbol{a}_2 I(t_1 < t \le n) + \boldsymbol{X}_t \\ &= \boldsymbol{a}_1 I(0 < t \le t_1) + \boldsymbol{a}_2 I(t_1 < t \le n) + \mathbf{A}\mathbf{z}_t,\end{aligned}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{z}_t$ are given as in Example 3. Table 4 reports the RFs of the occurrence of the event $\{\widehat{r} = r\}$ and the average distance (4.1) for the case with a change point $t_1 = [n/2]$ in a simulation with 500 replications, where the cointegration rank is estimated as in Example 2. Also included in Table 4 are the

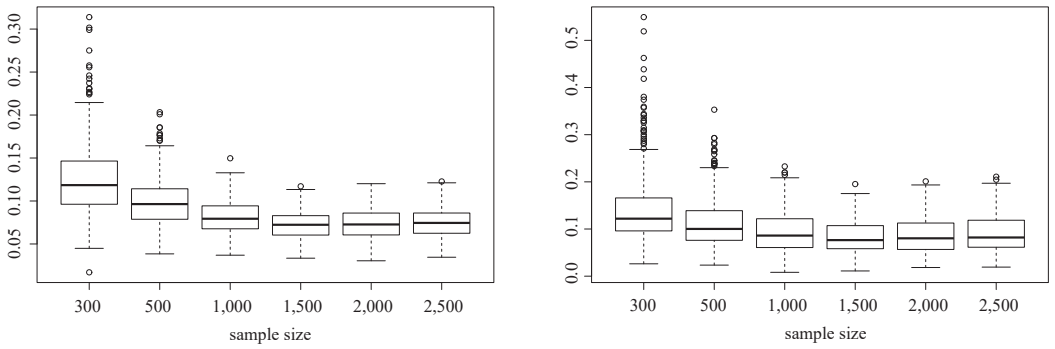

Figure 3. The left panel is the box plot of the distance $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ based on the proposed method, and the right panel is based on the PP.test when $p = 8, r = 4$ in Example 3.

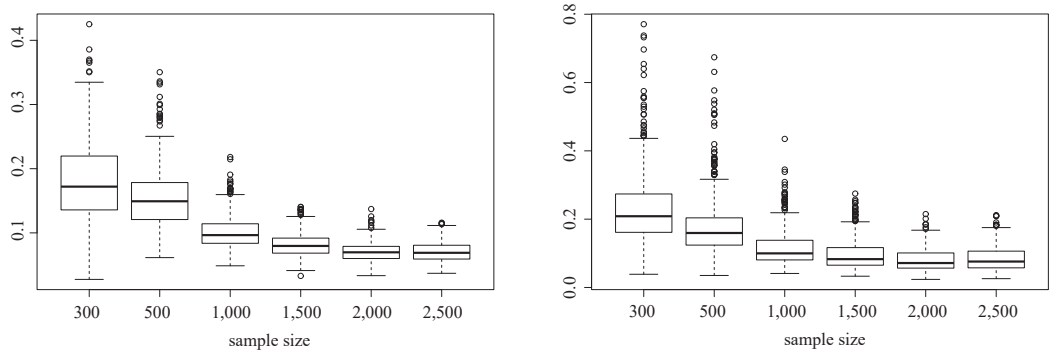

Figure 4. The left panel is the box plot of the distance $D_1^*(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2))$ based on the proposed method, and the right panel is based on the PP.test when $p = 12, r = 6$ in Example 3.

results obtained from applying the Phillips–Perron unit-root test for $\widehat{\mathbf{z}}_t$ defined in (4.2) to estimate r. Table 4 indicates that (2.7) works well, even in the presence of breaks in a constant trend. However, the unit-root procedure performs very poorly if one overlooks the breaks in the data.

Example 5. Next, we consider an example with breaks in level and slope. Let

$$\begin{aligned}\mathbf{y}_t &= (\boldsymbol{a}_1 + \mathbf{b}_1 t) I(0 < t \le t_1) + (\boldsymbol{a}_2 + \mathbf{b}_2 t) I(t_1 < t \le n) + \boldsymbol{X}_t \\ &= (\boldsymbol{a}_1 + \mathbf{b}_1 t) I(0 < t \le t_1) + (\boldsymbol{a}_2 + \mathbf{b}_2 t) I(t_1 < t \le n) + \mathbf{A}\mathbf{z}_t, \qquad (4.5)\end{aligned}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{z}_t$ are given as in Example 3, $t_1 = [n/2]$, and all components of $\mathbf{b}_1$ and $\mathbf{b}_2$ are taken as 0.3 and 0.8, respectively. It follows from (4.5) that

Table 4. RFs of $\{\hat{r} = r\}$ and the average distance D_1 between the estimated and the true cointegration spaces in a simulation with 500 replications for Example 4.

		$n = 300$		$n = 500$		$n = 1000$		$n = 1500$		$n = 2000$		$n = 2500$	
(p, r)	Method	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
	IC	0.996	0.116	1.000	0.100	1.000	0.089	1.000	0.086	1.000	0.083	1.000	0.083
$(6, 2)$	ACF	0.190	0.594	0.176	0.600	0.092	0.651	0.076	0.660	0.066	0.666	0.026	0.692
	PP.test	0.472	0.424	0.428	0.445	0.340	0.494	0.312	0.509	0.260	0.542	0.232	0.560
	IC	0.986	0.095	0.998	0.071	1.000	0.059	1.000	0.053	1.000	0.052	1.000	0.051
$(6, 3)$	ACF	0.590	0.315	0.578	0.313	0.514	0.342	0.488	0.356	0.434	0.391	0.380	0.424
	PP.test	0.922	0.135	0.918	0.126	0.916	0.116	0.910	0.115	0.906	0.118	0.878	0.137
	IC	0.998	0.116	1.000	0.088	1.000	0.071	1.000	0.065	1.000	0.060	1.000	0.057
$(8, 3)$	ACF	0.388	0.451	0.362	0.457	0.340	0.472	0.300	0.493	0.262	0.518	0.222	0.555
	PP.test	0.840	0.210	0.842	0.190	0.814	0.194	0.784	0.210	0.816	0.186	0.796	0.195
	IC	0.994	0.107	1.000	0.088	1.000	0.078	1.000	0.075	1.000	0.073	1.000	0.074
$(8, 4)$	ACF	0.092	0.651	0.052	0.704	0.036	0.725	0.034	0.745	0.016	0.767	0.016	0.789
	PP.test	0.410	0.520	0.332	0.575	0.264	0.629	0.224	0.667	0.214	0.666	0.188	0.696
	IC	0.996	0.130	1.000	0.098	1.000	0.077	1.000	0.069	1.000	0.067	1.000	0.064
$(12, 6)$	ACF	0.058	0.645	0.046	0.663	0.022	0.700	0.018	0.735	0.012	0.769	0.012	0.767
	PP.test	0.394	0.538	0.380	0.552	0.326	0.590	0.318	0.603	0.210	0.699	0.214	0.693
	IC	0.978	0.099	1.000	0.077	0.998	0.063	1.000	0.058	1.000	0.057	1.000	0.055
$(12, 8)$	ACF	0.074	0.576	0.056	0.610	0.040	0.636	0.010	0.667	0.014	0.680	0.010	0.708
	PP.test	0.438	0.507	0.352	0.589	0.322	0.623	0.274	0.669	0.214	0.726	0.214	0.726
	IC	0.780	0.183	0.984	0.116	1.000	0.078	1.000	0.068	1.000	0.061	1.000	0.057
$(20, 14)$	ACF	0.034	0.562	0.030	0.564	0.034	0.575	0.026	0.617	0.018	0.632	0.014	0.650
	PP.test	0.346	0.491	0.350	0.499	0.384	0.524	0.320	0.592	0.302	0.623	0.304	0.630
	IC	0.544	0.291	0.926	0.204	1.000	0.135	1.000	0.111	1.000	0.096	1.000	0.089
$(30, 20)$	ACF	0.004	0.701	0.002	0.661	0	0.643	0	0.644	0.002	0.633	0.002	0.666
	PP.test	0.118	0.617	0.152	0.588	0.232	0.526	0.274	0.511	0.306	0.525	0.282	0.560
	IC	0.092	0.467	0.242	0.391	0.972	0.281	1.000	0.232	1.000	0.202	1.000	0.182
$(50, 30)$	ACF	0	0.854	0.002	0.859	0	0.859	0	0.848	0.002	0.839	0.002	0.836
	PP.test	0.006	0.813	0	0.837	0.002	0.824	0	0.788	0.010	0.775	0.004	0.761
	IC	0.314	0.511	0.030	0.470	0.042	0.390	0.646	0.328	0.976	0.288	1.000	0.263
$(80, 50)$	ACF	0	0.872	0	0.885	0	0.907	0	0.916	0.002	0.916	0	0.919
	PP.test	0.004	0.835	0.002	0.832	0	0.880	0	0.890	0	0.887	0	0.885

IC is based on the proposed method, ACF is based on (4.4) by ZRY, the PP.test is based on the Phillips–Perron unit-root test, and r is the cointegration rank.

$$\begin{aligned}\nabla \mathbf{y}_t = {} & \mathbf{b}_1 I(1 < t \le t_1) + \{\mathbf{b}_2 + (\mathbf{b}_2 - \mathbf{b}_1)t_1\} I(t = t_1 + 1) \\ & + \mathbf{b}_2 I(t > t_1 + 1) + \nabla \mathbf{z}_t. \end{aligned} \tag{4.6}$$

We estimate the change point t_1 by

$$\hat{t}_1 = \underset{1 \le t \le n}{\operatorname{argmax}} |\nabla y_{t,2}|, \tag{4.7}$$

and then apply our procedure to estimate r and the cointegration space by replacing $\nabla \mathbf{y}_{\hat{t}_1+1}$ with zero. The corresponding RFs of the occurrence of the event $\{\widehat{r} = r\}$ and the average distance (4.1) in a simulation with 500 replications are

Table 5. RFs of $\{\hat{r} = r\}$ and the average distance D_1 between the estimated and the true cointegration spaces by the proposed method in a simulation with 500 replications for Example 5.

	$n = 300$		$n = 500$		$n = 1000$		$n = 1500$		$n = 2000$		$n = 2500$	
(p, r)	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1	RF	D_1
$(6, 3)$	0.992	0.090	1.000	0.070	1.000	0.058	1.000	0.054	1.000	0.051	1.000	0.051
$(8, 4)$	0.996	0.105	1.000	0.088	1.000	0.078	1.000	0.076	1.000	0.073	1.000	0.073
$(12, 8)$	0.996	0.091	1.000	0.074	1.000	0.062	1.000	0.058	1.000	0.057	1.000	0.055
$(20, 14)$	0.936	0.158	0.996	0.111	1.000	0.079	1.000	0.067	1.000	0.060	1.000	0.057
$(30, 20)$	0.820	0.272	0.980	0.198	1.000	0.135	1.000	0.112	1.000	0.096	1.000	0.087
$(50, 30)$	0.024	0.473	0.088	0.397	0.966	0.283	1.000	0.233	1.000	0.202	1.000	0.183
$(80, 50)$	0.246	0.512	0.004	0.474	0.010	0.394	0.580	0.330	0.972	0.290	1.000	0.264

reported in Table 5. Table 5 indicates that the proposed procedure (2.7) is simple to carry out, and works well even in the presence of breaks in a linear trend.

In summary, Examples 1–5 show that the proposed procedure (2.7) works reasonably well when p is small and n is large. When p is large, for example when $p \geq 50$, the methods based on the ACF of ZRY, the likelihood ratio of Johansen, and the PP.test all fail to work adequately. The proposed method, however, can still estimate the cointegration rank and the cointegration space very well for large p. That is, (2.7) always performs better than the other methods. It is shown that for a given p, a larger r/p yields more accurate proposed estimates for r. In general, (2.7) is more robust than other existing procedures. Both the figures and the tables show that the estimation errors of the proposed method decrease as the sample size n increases. There are a couple of possible explanations for the superiority of the proposed method. First, the breaks have little effect on estimating the cointegration rank when $\nabla \mathbf{y}_t - \nabla \bar{\mathbf{y}}$ is used, whereas other procedures using the original series are sensitive to the existence of breaks. The advantage of our method is that it avoids estimating and testing for the trends. Second, the proposed method allows a larger p, because the matrix for recovering the cointegration space is based on the stationary process $\nabla \mathbf{y}_t$, whereas other methods can deal only with smaller values of p.

Example 6. We consider the monthly exchange rates of the U.S. dollar versus the currencies of 19 OECD countries (Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, and the United Kingdom) from January 1973 to December 2000, published by the Federal Reserve Economic Data (FRED), and analyzed in Engel, Mark and West (2015) and Trapani (2021). All exchange rates are transformed by taking the logarithm, and each

of the transformed series are shown to follow a unit-root model based on the PP.test at significance levels 0.05 and 0.1. Applying the proposed method to the transformed data, both (4.4) and the PP.test at significance level 0.1 for the series $\widehat{\mathbf{z}}_t = \widehat{\mathbf{A}}'\mathbf{x}_t$ lead to $\widehat{r} = 2$ with $M = n^{1/4}$ and $\widehat{r} = 4$ with $M = 4(n/100)^{1/4}$ (given as in the KPSS test); $\widehat{r} = 1$ for both $M = n^{1/4}$ and $M = 4(n/100)^{1/4}$ by the PP.test at significance level 0.05. To shed further light on the estimated rank, we plot the four possible cointegrated series and their ACF corresponding to the first four smallest eigenvalues based on (4.4) with $M = 4(n/100)^{1/4}$ and $M = n^{1/4}$ in Figure 5 and Figure 6, respectively. Both figures show that the two estimated series are stationary, whereas the other two series tend to be nonstationary. It would therefore be more reasonable to estimate the cointegration rank as two. On the other hand, all procedures show that cointegration exists in the exchange rates, which is consistent with the findings of Engel, Mark and West (2015), who showed that these exchange rates can be modeled by a nonstationary factor model.

5. Conclusion

We have proposed a simple, direct, and model-free method for identifying cointegration relationships between high-dimensional integrated time series with possible breaks in trend. To remove the effects of the trend and the possible breaks, we propose using the first difference of the observed series. Based on a nonnegative-definite matrix consisting of the Bartlett-weighted sample covariance, the cointegration space can be recovered by the eigenvectors corresponding to the smallest eigenvalues. An information criterion is proposed to estimate the cointegration rank, which estimates the rank consistently. In addition, unit-root tests may be applied to determine the number of stationary components of $\widehat{\mathbf{z}}_t$. The proposed method differs from that of ZRY, which is based on the quadratic form of the sample covariances of the observed series. Under similar conditions, the proposed procedure allows p to be much larger.

This study focuses only on inference for the cointegration rank r and the cointegration space $\mathcal{M}(\mathbf{A}_2)$ when all the components of the cointegrated series are $I(1)$ processes and there are possible breaks in level. With some extra effort, it is feasible to generalize the procedure to the case with breaks in trend. Under this circumstance, the change points can be detected easily and the proposed method can be applied to the subsamples in each regime; see Example 5 in Section 4. Furthermore, it would be interesting to extend the procedure to cases in which the integrated orders of the cointegrated series are different (see ZRY) or have

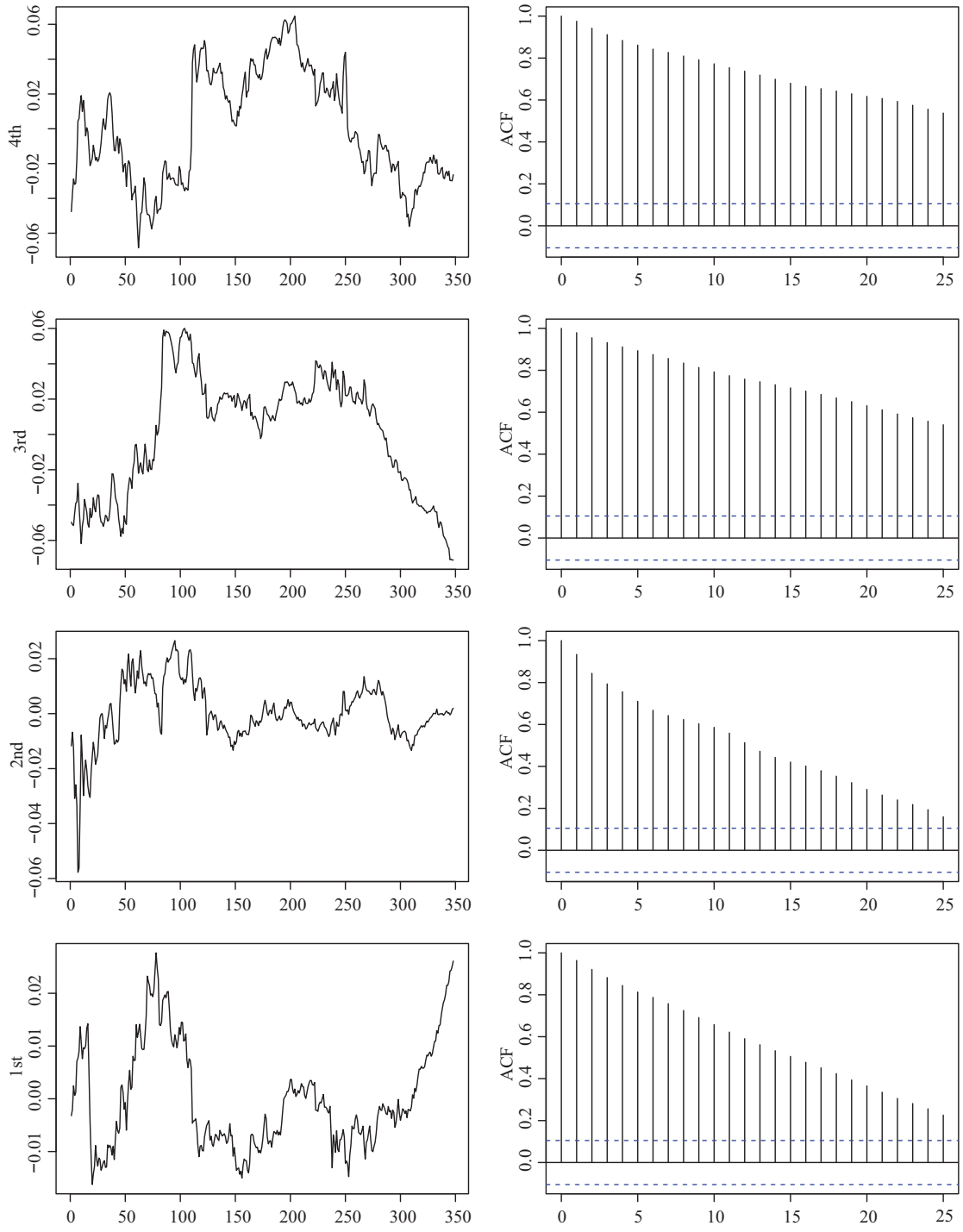

Figure 5. Time series plots of the estimated cointegrated series by the proposed method and their sample ACF for the logarithmic exchange rates based on $M = 4(n/100)^{1/4}$.

fractional values (i.e., fractional cointegration (see, for example, Robinson and Hualde (2003); Robinson (2008); Chen and Hurvich (2006)). These topics are left to future research.

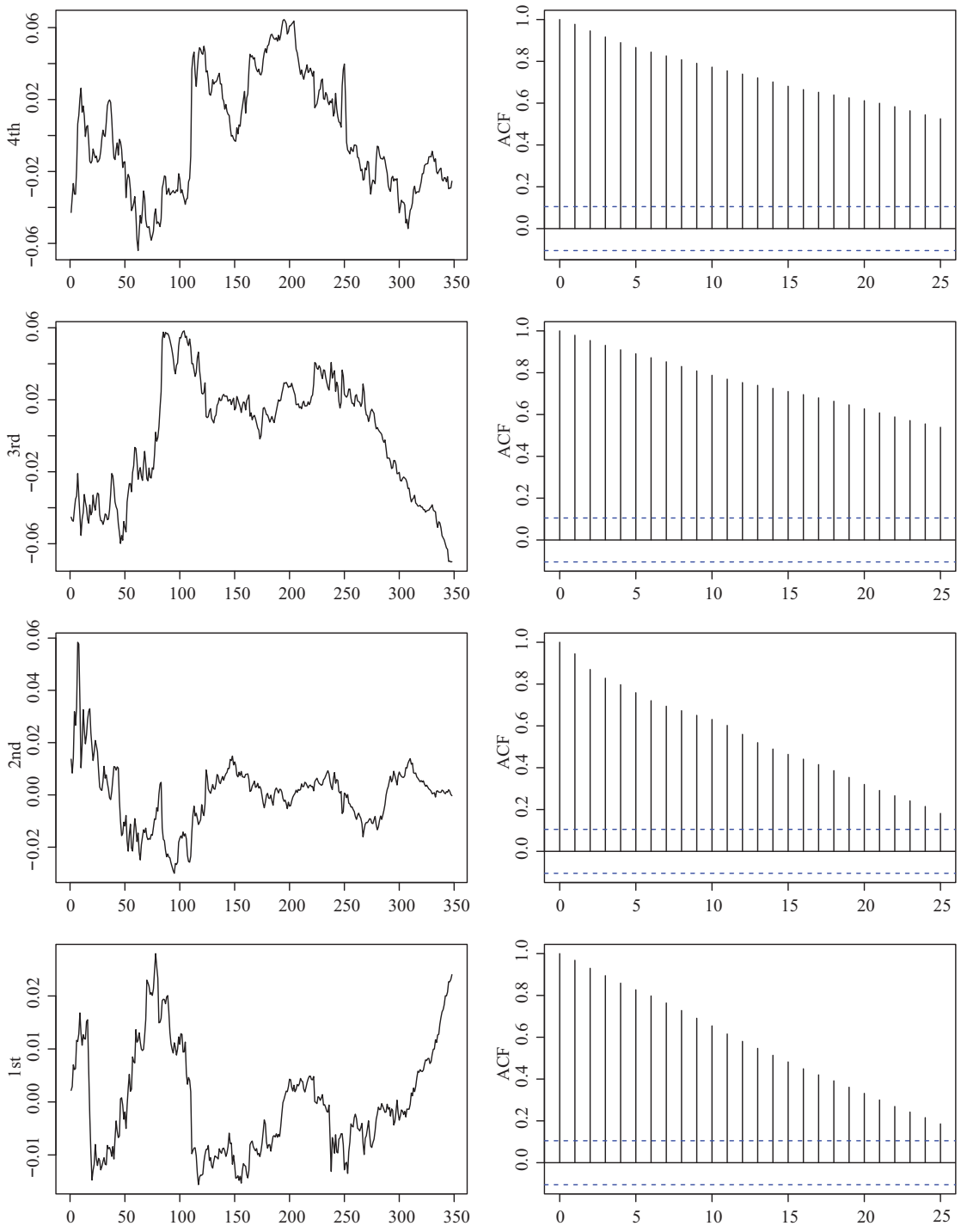

Figure 6. Time series plots of the estimated cointegrated series by the proposed method and their sample ACF for the logarithmic exchange rates based on $M = n^{1/4}$.

Acknowledgements

We would like to thank the co-editor, an associate editor and three anonymous referees for their critical comments and thoughtful suggestions. This re-

search was supported in part by grants from HKSAR-RGC-GRF No. 14325216, 14307921, and the Theme-based Research Scheme of HKSAR-RGC-TBS T32-101/15-R (Chan), and NSFC (Nos. 12171427, 11771390, U21A20426), Zhejiang provincial natural science foundation (No. LZ21A010002) and the Fundamental Research Funds for the Central Universities (Zhang).

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd Edition. Wiley, New York.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* **43**, 1535–1567.

Campos, J., Ericsson, N. R. and Hendry, D. F. (1996). Cointegration tests in the presence of structural breaks. *Journal of Econometrics* **70**, 187–220.

Cavaliere, G. and Georgiev, I. (2007). Testing for unit roots in autoregressions with multiple level shifts. *Econometric Theory* **23**, 1162–1215.

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

Chen, L. K. and Wu, W. B. (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association* **114**, 869–881.

Engel, C., Mark, N. C. and West, K. D. (2015). Factor model forecasts of exchange rates. *Econometric Reviews* **34**, 32–55.

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

Gonzalo, J. and Pitarakis, J. Y. (1999). Dimensionality effect in cointegration analysis. In *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger* (Edited by R. F Engle and H. White), 212–229. Oxford University Press, Oxford.

Harris, D. (1997). Principal components analysis of cointegrated time series. *Econometric Theory* **13** 529–557.

Harris, D., Leybourne, S. J. and Taylor, A. M. R. (2016). Tests of the co-integration rank in VAR models in the presence of a possible break in trend at an unknown point. *Journal of Econometrics* **192**, 451–467.

Ho, M. S. and Sorensen, B. E. (1996). Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based monte carlo simulations. *The Review of Economics and Statistics* **78**, 726–732.

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

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 Models*. Oxford University Press, Oxford.

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on Cointegration–with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* **52**, 169–210.

Johansen, S., Mosconi, R. and Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometric Journal* **3**, 216–249.

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.

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

Müller, U. K. and Watson, M. W. (2013). Low-frequency robust cointegration testing. *Journal of Econometrics* **174**, 66–81.

Newey, W. K. and West, K. (1987). A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.

Onatski, A. and Wang, C. (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica* **86**, 1465–1478.

Perron, P. (1997). Further evidence of breaking trend functions in macroeconomic variables. *Journal of Econometrics* **80**, 355–385.

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.

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 cointegration rank of a VAR process with structural shifts. *Journal of Business & Economic Statistics* **18**, 451–464.

Trapani, L. (2021). Inferential theory for heterogeneity and cointegration in large panels. *Journal of Econometrics* **220**, 474–503.

Trenkler, C., Saikkonen, P. and Lütkepohl, H. (2008). Testing for the cointegrating rank of a VAR process with level shift and trend break. *Journal of Time Series Analysis* **29**, 331–358.

Zhang, R. M. and Chan, N. H. (2018). Portmanteau-type tests for unit-root and cointegration. *Journal of Econometrics* **207**, 307–324.

Zhang, R. M., Robinson, P. and Yao, Q. (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association* **114**, 916–927.

Zivot, E. and Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics* **10**, 251–270.

Ngai Hang Chan

Department of Biostatistics, City University of Hong Kong, Kowloon, Hong Kong.

E-mail: nhchan@cityu.edu.hk

Rongmao Zhang

School of Mathematics, Minnan Normal University, Zhangzhou, Fujian, China.
School of Mathematics, Zhejiang University, Hangzhou, Zhejiang 310058, China.

E-mail: rmzhang@zju.edu.cn

(Received March 2021; accepted August 2021)