The Effects of $I(1)$ Series on Cointegration Inference

YAN-XIA LIN[†]   yanxia@uow.edu.au
School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia

MICHAEL MCCRAE   mccrae@uow.edu.au
Department of Accounting and Finance, University of Wollongong, Wollongong, NSW 2522, Australia

Abstract. Under traditional cointegration tests, some eligible $I(1)$ time series systems $\mathbf{X}_t$, that are not cointegrated over a given time period, say $(0, T_1]$, sometimes test as cointegrated over sub-periods. That is, the system appears to have a stationary linear structure $\xi' \mathbf{X}_t$ for certain vector ξ in the period $0 < t \leq T_1$. Understanding the dynamics between cointegration test power and restricted sample size that causes this inversion of results is a crucial issue when forecasting over extended future time periods. In this paper, we consider non-cointegrated systems that are closely related to collinear systems. We apply a residual based procedure to such systems and establish a criterion for making the decision whether or not $\mathbf{X}_t$ can be continuously accepted as $I(0)$ for $t > T_1$ when $\mathbf{X}_t$ was accepted as $I(0)$ for $t \leq T_1$.

Keywords: cointegration, cointegrating vector, ARIMA model, eigenvalue.

1. Introduction

Cointegration testing of pre-suppositions of long run equilibrium behaviour within financial and economic time series vectors is usually based on sample data drawn from a selected period at a given frequency (eg. daily, monthly, annually). Forecasts are then made on the assumption that the test results are invariant to either data frequency over a given period or to extensions of the sample horizon – subject to constant system dynamics.

However, several studies cast doubt upon this assumption. A system with no long run equilibrium tendency over a given time period may test as cointegrated over sub-periods within the sample period (see Shiller and Perron [11]; Perron [10]; Otero and Smith [9]). The false acceptance of cointegration is attributed to the low discriminatory power of cointegration tests due to small sample sizes over restricted time periods or low frequency

[†] Requests for reprints should be sent to Yan-Xia Lin, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia.

data. Data set expansion may then invert results as increased test power detects the true lack of cointegration.

The effect of sample size on the discriminatory power of cointegration tests is well documented. See, for example, Kwiatkowski, et al. [6]. But, to our knowledge, the dynamics of the relationship are not well understood. The aim of this paper is to explore these dynamics, suggest a theoretical basis for their operation and develop criteria about minimum sample size for detection of the true non-cointegration situation.

This article will focus on non-cointegrated systems $\mathbf{X}_t$ that are closely related to collinear systems (for detailed explanations, see Section 2). For data given for the time period $(0, T_1]$, if $\mathbf{X}_t$ is accepted as a cointegrated system, this means that there is a vector ξ such that $\xi' \mathbf{X}_t$ can be accepted as $I(0)$ for $t \in (0, T_1]$. Since the system $\mathbf{X}_t$ is essentially non cointegrated, it is not necessary that $\xi' \mathbf{X}_t$ can be continuously accepted as $I(0)$ as the time period $(0, T_1]$ is extended to $(0, T]$. To detect the smallest T such that $\xi' \mathbf{X}_t$ can be continuously accepted as $I(0)$ in $(0, T]$ this paper uses a residual based (RB) procedure (see Section 2) to decompose time series $\xi' \mathbf{X}_t$ into two independent time series such that one is $I(0)$ and the other is $I(1)$. Then a criterion for determining the smallest T based on the impact of $I(1)$ to the system can be established.

This paper is organized into five sections. The procedure for identifying stationary and non-stationary components is explained in Section Two. A simulation approach is used in Section Three to show that the significance of the $I(1)$ component determines the power of a DF test (Dickey and Fuller [2]) applied to the system. Test power increases in direct relation to the significance of the $I(1)$ component as time horizons are expanded. Section Four describes a method for measuring this impact of the non-stationary component within the system over expanded sample horizons. The method is used for calculating the ratio of the variances of the non-stationary and stationary components of $\xi' \mathbf{X}_t$. This method enables us to plot differences in the minimum number of data points necessary to reverse a cointegration result (span of observation period) against changes in the variance ratio values for each system. Section 5 contains a discussion on a special case.

2.   Residual Based Procedure

This section briefly describes the RB procedure approach (Lin and McCrae [7], [8]) and its association with traditional cointegration tests. We start with the definition of $I(1)$ time series and cointegration.

DEFINITION 1 *A time series X_t is called an $I(d)$, $d > 0$, time series iff $(1 - B)^d X_t$ is stationary and has an ARIMA form, where B is the back shift operator.*

A certain form of linear combination of $I(d)$ series may become an $I(d')$ series with $d' < d$. If $d' = 0$, then the linear combination of these non-stationary time series is a stationary series. Thus, these non-stationary time series form a cointegrated system. The coefficients of the linear combination form a cointegral vector. Here, we give the cointegration definition based on a $I(1)$ series.

DEFINITION 2 *Let $X_{1,t}, \cdots, X_{p,t}$ be $I(1)$ time series. If there is a non zero vector $\xi' = (\xi_1, \cdots, \xi_p)$ such that $\xi_1 X_{1,t} + \cdots + \xi_p X_{p,t}$ is an $I(0)$ series, then, $X_{1,t}, \cdots, X_{p,t}$ are said to be cointegrated and ξ is called a cointegral vector for $\mathbf{X}_t = (X_{1,t}, \cdots, X_{p,t})'$.*

Remark: Given an $I(1)$ system $\mathbf{X}_t$, if $cov((1-B)\mathbf{X}_t) = cov(\mathbf{X}_t - \mathbf{X}_{t-1})$ is not full rank, then there is a vector ξ such that $\xi' \mathbf{X}_t$ is stationary. Therefore the system $\mathbf{X}_t$ is cointegrated. This system is sometimes defined as a collinear system. Dhrymes [3] claimed that a collinear system does not indicate any "long-run equilibrium" relation. In practice, given the system information within finite time period, a non collinear system might be accepted as collinear; for example, some spot and future price systems show this outcome (see Biondini, Lin and McCrae [1]). Such systems are of interest for the study carried out in this paper and these systems have a close relationship with colliner systems.

For cointegrated systems, several techniques for estimating cointegral vectors are available. These include the univariate two-step Engle-Granger procedure [4], Johansen's [5] multivariate maximum likelihood method and, more recently, the RB procedure, developed by Lin and McCrae [7], [8]. The procedure for determining linearly independent cointegrating vectors introduced by Lin and McCrae [7], [8] is based on the residual error processes from fitting ARIMA models to individual time series. As it will be shown below, this procedure is suitable for systems which are closely related to collinearity.

Consider $\mathbf{X}_t$, a $p \times 1$ $I(1)$ time series vector. To simplify the following discussion, assume that each component of $\mathbf{X}_t$ has the following expression:

$$\Phi_i(B)(1 - B)X_{it} = \Theta_i(B)\varepsilon_{it}, \quad t > 0, \tag{1}$$

where, both $\Phi_i(B)$ and $\Theta_i(B)$ are finite-order polynomial functions of B that have roots outside the unit circle, $i = 1, 2, \cdots, p$; ε_{it} are white noise residuals and $\varepsilon_t = (\varepsilon_{1,t}, \cdots, \varepsilon_{p,t})'$ is stationary. Furthermore, the ε_t are assumed to be multi-normally distributed.

Assume that $var(\varepsilon_t) = \Sigma$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{p-r} \geq \lambda_{p-r+1} \geq \cdots \geq \lambda_p$. If the last r eigenvalues are equal to 0, i.e. $\lambda_{p-r+1} = \cdots = \lambda_p = 0$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{p-r} > 0$ then the system X_t is collinear and cointegrated (see Lin and McCrae, [7] and [8]). This is not what we are interested in this paper. In the following discussion, we assume that $\lambda_{p-r+1}, \cdots, \lambda_p$ are not equal to zero but are close to zero. Thus, the RB procedure can be used to determine cointegral vector of $\mathbf{X}_t$ if $\mathbf{X}_t$ is accepted as cointegrated based on the given sample.

The procedure involves three steps.

Step 1: Since $cov(\varepsilon_t) = \Sigma$ has eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$, a matrix A^* will be determined such that $\varepsilon_t = A^* \mathbf{v}_t^* = A_1^* \mathbf{v}_{(1),t}^* + A_2^* \mathbf{v}_{(2),t}^*$, where $\mathbf{v}_{(1),t}^* = (v_{1,t}^*, \cdots, v_{p-r,t}^*)'$, $\mathbf{v}_{(2),t} = (v_{p-r+1,t}^*, \cdots, v_{p,t}^*)'$ and $var(\mathbf{v}_t^*) = diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$ is a diagonal matrix, i.e. the components of $\mathbf{v}_t^*$ are uncorrelated.

Step 2: Express $(1 - B)\mathbf{X}_t$ in the following way

$$
\begin{aligned}
(1 - B)\mathbf{X}_t &= \Phi(B)^{-1}\Theta(B)\epsilon_t = C(B)\mathbf{v}_{(1),t}^* + C_1(B)\mathbf{v}_{(2),t}^* \\
&= [C(B) - C(1)]\mathbf{v}_{(1),t}^* + C(1)\mathbf{v}_{(1),t}^* + C_1(B)\mathbf{v}_{(2),t}^* \quad (2)
\end{aligned}
$$

where $\Phi(B)$ and $\Theta(B)$ are diagonal matrices with element $\phi_1(B), \cdots, \phi_p(B)$ and $\theta_1(B), \cdots, \theta_p(B)$ respectively; $C(1) = C(B)$ with $B = 1$.

Step 3: As the rank of $C(1)$ is less than or equal to $p - r$, r linearly independent vectors ξ can be determined by solving $\xi' C(1) = 0$. Therefore, from (1), $\xi'(1 - B)\mathbf{X}_t = (1 - B)\xi' \frac{C(B) - C(1)}{1 - B}\mathbf{v}_{(1),t}^* + \xi' C_1(B)\mathbf{v}_{(2),t}^*$, that is,

$$
\xi' \mathbf{X}_t = \xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi' \mathbf{W}_t + \xi' C_1(B) \sum_{i=1}^{t} \mathbf{v}_{(2),i}^* \quad (3)
$$

with $\mathbf{W}_t = \frac{C(B) - C(1)}{1 - B}\mathbf{v}_{(1),t}^*$.

The above three steps, allow $\xi' \mathbf{X}_t$ to be expressed as a sum of two independent time series. In general, the series $\xi' \mathbf{X}_t$ is not an $I(0)$ series, as $\xi' C_1(B) \sum_{i=1}^{t} \mathbf{v}_{(2),i}^*$ is not equal to zero. However, $\xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi' \mathbf{W}_t$ is stationary. The difference between $\xi' \mathbf{X}_t$ and $\xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi' \mathbf{W}_t$ is given by

$$
\xi' C_1(B) \sum_{i=1}^{t} \mathbf{v}_{(2),i}^*. \quad (4)
$$

Thus, whether $\xi' \mathbf{X}_t$ can be accepted as $I(0)$ will depend on whether we can ignore the effect of the $I(1)$ series (4).

If the last r eigenvalues of Σ are zero, (4) will disappear and the system $\mathbf{X}_t$ will be classified as a cointegrating collinear system with cointegral vector ξ. In practice, given a sample from a finite time period $[0, T_1]$, sometimes

$\xi'\mathbf{X}_t$ can be accepted as $I(0)$ even though $\xi'C_1(B)\sum_{i=1}^{t}\mathbf{v}_{(2),i}^{*}$ is not strictly equal to zero. In this situation, ξ can be interpreted as an estimate of the cointegral vector for the system $\{X_t, 0 < t \leq T_1\}$.

In this paper, we always assume that: (i) $\xi'C_1(B)\sum_{i=1}^{t}\mathbf{v}_{(2),i}^{*}$ are not all equal to zero, for all $0 < t \leq T_1$; (ii) for $0 < t \leq T_1$, the system $\{X_t\}$ is accepted as cointegrated .

If the system $\{\mathbf{X}_t\}_{t>0}$ is essential non-cointegrated, then a crucial question for forecasting is to identify the maximum period over which a spurious cointegration result from a restricted sample period $\{\mathbf{X}_t\}_{0<t\leq T}$ will persist as $(0, T_1]$ extends out to $(0, T]$ with $T > T_1$, where all $t > T_1$ represent (unobservable) future times not amenable to direct testing (assuming constant system dynamics).

We can now give a more precise description on the relationship between cointegrating test power and sample size. The question of whether $\xi'\mathbf{X}$ remains spuriously cointegrated as the sample period expands is equivalent to determing whether the impact of the term $\xi'C_1(B)\sum_{i=1}^{t}\mathbf{v}_{(2),i}^{*}$ on the cointegrating structure remains sufficiently insignificant. In the next section, we use simulated data to show how the $I(1)$ time series in (3) impacts on the stationary analysis of $\xi'\mathbf{X}$. The simulation results suggest a practical method for measuring the impact of the $I(1)$ series on $\xi'\mathbf{X}$.

3. Simulations

In this section, we use the following model to show the impact of the $I(1)$ series

$$y_t = \varepsilon_t + \sum_{k=1}^{t} u_k, \qquad 0 < t \leq T, \tag{5}$$

where $\{\varepsilon_i\}$ and $\{u_i\}$ are i.i.d. respectively. In this model, ε_t is stationary, but y_t is not. The difference between these two series is $\sum_{k=1}^{t} u_k$, an $I(1)$ series.

By using simulated data, we show how T and the ratio of the variance ε_t and u_t impact on the stationality conclusion of y_t. For related references see Kwiatkowski and et al. [6].

Let $\lambda(t) = var(\sum_{k=1}^{t} u_k)/var(\varepsilon_t) = t\sigma_{u_1}^2/\sigma_{\varepsilon_1}^2$. In the following study, we set $\sigma_{\varepsilon_1}^2 = 1$ and $\sigma_{u_1}^2 = 0.1, 0.01$ and 0.001 respectively. Thus, for each case, $\max_{0<t\leq T}\lambda(t) = \lambda(T) = T\sigma_{u_1}^2/\sigma_{\varepsilon_1}^2 = 0.1T, 0.01T$ and $0.001T$ respectively.

In the simulation, we let $T = 1000$. Thus, there are 1000 data points for each case. The time series plots for y_t and ε_t are given in Figure 1.

The plot in Figure 1 suggests that (i) for $\sigma_{u_1} = 0.001$, y_t is likely to be accepted as a stationary series for t up to $t = 1000$, where $\lambda(1000) = 1$; (ii)
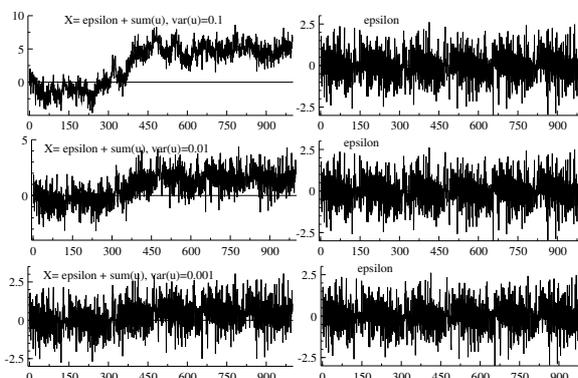
Figure 1. The plots of y_t and ε_t.

for $\sigma_{u_1} = 0.01$, y_t could be accepted to be stationary for $t \leq 200$, where $\lambda(200) = 2$; (iii) for $\sigma_{u_1} = 0.1$, y_t is definitely not acceptable as stationary, even for $t < 150$, where $\lambda(150) = 15$. The above conclusions are confirmed by applying DF test (Dickey and Fuller [2]) to each set of data. The DF test outcomes are given by Table 1.

Table 1: DF test results for simulated data.

σ_u	Size	DF Value	5 %	1 %
0.1	1000	-1.324	-2.865	-3.44
	500	-0.6023	-2.865	-3.446
	200	-2.703	-2.877	-3.466
0.01	1000	-2.236	-2.865	-3.44
	500	-1.227	-2.865	-3.446
	200	-4.281*	-2.877	-3.466
0.001	1000	-5.8211*	-2.865	-3.04
	500	-4.203*	-2.865	-3.446
	200	-6.023*	-2.877	-3.466

These results show that the likelihood of accepting $\{y_t\}_{0 \leq t \leq T}$ as stationary is small when the maximum ratio of $var(\sum_{k=1}^{t} u_k)$ to $var(\varepsilon_t)$ is large for $t \in [0, T]$. This suggests that the value of $\lambda(t)$ maybe useful for identifying the size of the impact of the $I(1)$ series on the system of y_t. Thus, if y_t is accepted as stationary for $0 < t \leq T_1$ and if, for $T > T_1$, the value of $\lambda(T)$ can be estimated through the information of y_t given up to "present time" T_1, then it is possible to predict if y_t can be continuously accepted as stationary for $0 < t < T$. Therefore, to answer the questions addressed in Section 1, we need a quantity which can play the same role as $\lambda(T)$ for

predicting the future behaviour the system. Such a quantity will be studied in the next session.

4.   The Relationship Between $\sigma_u/\sigma_\varepsilon$ and Stationarity

The example in Section 3 shows that whenever a linear structure of a system can be expressed as a sum of two independent parts - stationary and non-stationary, the behaviour of the non-stationary part will have a direct impact on the power of testing the null hypothesis of stationarity of the linear structure of the system against non-stationarity. The lower the significance of the non-stationary part is, the more likely y_t in (5) is to be accepted as stationary, i.e. the power of testing stationarity is lower. The example also indicates that the size of the impact of the non stationary component in the linear structure of y_t could be measured by the ratio of variances given by non stationary and stationary parts.

In this section, we introduce a method of measuring the size of impact of the non-stationary time series in model (5). The measurement relates to the span of the sample and the ratio of the variances.

As described in Section 2, given a system $\mathbf{X}_t$ with components satisfying (1), there is a vector ξ such that

$$\xi'\mathbf{X}_t = \xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi'\mathbf{W}_t + \xi'C_1(B)\sum_{i=1}^{t} \mathbf{v}^*_{(2),i}, \tag{6}$$

where $\xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi'\mathbf{W}_t$ is $I(0)$. If the $\mathbf{v}^*_{(2),i}$ are not zero vectors, then, $\xi'C_1(B)\sum_{i=1}^{t} \mathbf{v}^*_{(2),i}$ is $I(1)$. Whether or not $\mathbf{X}_t$ can be accepted as cointegrated (associated with the vector ξ) will depend upon the significance of the impact of $\xi'C_1(B)\sum_{i=1}^{t} \mathbf{v}^*_{(2),i}$ on the system.

In general, $\xi'C_1(B)\sum_{i=1}^{t} \mathbf{v}^*_{(2),i}$ can be expressed as $\sum_{i=1}^{t} f_1(B)v_{p-r+1,i} + \sum_{i=1}^{t} f_2(B)v_{p-r+2,i} + \cdots + \sum_{i=1}^{t} f_p(B)v_{p,i} = I_{1,t} + I_{2,t} + \cdots + I_{r,t}$, where $I_{1,t}$, $I_{2,t}$, $\cdots$, $I_{r,t}$ are independent $I(1)$ series, $r < p$. Thus, the impact of $\xi'C_1(B)\sum_{i=1}^{t} \mathbf{v}^*_{(2),i}$ on the system $\mathbf{X}_t$ is the aggregate impact of these independent $I(1)$ series.

Model (6) can be written as

$$y_t = z_t + h_t, \tag{7}$$

where z_t is $I(0)$ and h_t is $I(1)$. Whether y_t can be accepted as $I(0)$ at a given level of significance will depend upon whether the impact of h_t is significant at that level. The following lemma gives a general conclusion on this issue.

PROPOSITION 1 *Assume that process y_t, $0 < t \leq T$, satisfies the model $y_t = z_t + h_t$, where $\{z_t\}$ are i.i.d. and independent of $\{h_t\}$; z_t and h_t are normally distributed with mean 0. Let $\sigma_1 = var(z_1)$, $\mathbf{\Sigma}_{h,T}$ be the covariance matrix given by $h_1, h_2, \cdots, h_T$ and $\mathbf{A}$ be an orthogonal matrix such that $\mathbf{A}\Sigma_{h,T}\mathbf{A}' = \sigma_2^2 diag(\lambda_1, \cdots, \lambda_T)$, where $\sigma_2^2\lambda_1 \geq \sigma_2^2\lambda_2 \geq \cdots \geq \sigma_2^2\lambda_T$ are the eigenvalues of $\mathbf{\Sigma}_{h,T}$. Let*

$$F = \frac{\sum_{i=1}^{m} \frac{x_i^2}{m}}{\sum_{i=m+1}^{T} \frac{x_i^2}{(T-m)}} \;\; and \;\; F^* = \frac{\sum_{i=1}^{m} \frac{x_i^2}{\sigma_1^2+\sigma_2^2\lambda_i}/m}{\sum_{i=m+1}^{T} \frac{x_i^2}{\sigma_1^2+\sigma_2^2\lambda_i}/(T-m)},$$

where $\mathbf{X} = (x_1, x_2, \cdots, x_T)' = \mathbf{A}(y_1, y_2, \cdots, y_T)'$. Then, F^* has the F-distribution with degrees of freedom m and $T - m$, and

$$\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2\lambda_1} \leq \frac{F}{F^*} \leq \frac{\sigma_1^2 + \sigma_2^2\lambda_1}{\sigma_1^2}. \tag{8}$$

The proof of Proposition 1 is straightforward and is omitted.

Assuming that the difference between λ_1 and λ_T is large (compared with σ_2^2), then, the necessary condition for accepting $\{Y_t\}_{0<t\leq T}$ as an $I(0)$ time series (that is, accept $\sigma_2^2 = 0$) at the level of significance α is that the value of F should fall within the interval $(F_{m-1,T-m-1,\alpha/2}, F_{m-1,T-m-1,1-\alpha/2})$, denoted (F_1, F_2).

Since (8) holds, the values of F and F^* will be very close if $\sigma_2^2\lambda_1/\sigma_1^2$ is small. Since F^* has the F-distribution with degrees of freedom m and $T - m$, with $(1 - \alpha)100\%$ chance, the value of F^* will be in the interval (F_1, F_2).

Therefore, when both m and $T - m$ are large, say greater than 100, and if $\sigma_2^2\lambda_1/\sigma_1^2 < 0.01$, then, for $\alpha \leq 0.05$, $F_1 \approx \sigma_1^2/(\sigma_1^2 + \sigma_2^2\lambda_1)F_1$ and $F_2 \approx (\sigma_1^2 + \sigma_2^2\lambda_1)/\sigma_1^2 F_2$. That is, with $(1 - \alpha)100\%$ chance F will fall in the interval (F_1, F_2). Then, when testing if H_0: $\{y_t\}_{0<t\leq T}$ is stationary against H_1: $\{y_t\}_{0<t\leq T}$ is $I(1)$, we will accept H_0. That is, the size of the impact of h_t on y_t is not significant.

Based on the above discussion, we suggest that, for the model (7), if $\sigma_2^2\lambda_1/\sigma_1^2 < 0.01$, we should be able to say the impact due to h_t is not significant and y_t can be accepted as $I(0)$.

In Section 5 we give a detailed discussion of a special case of model (6).

5. Special Discussion

In Section 4, we have shown how to test whether the $I(1)$ term in (7) is significant or not. Now we apply the method to (6). As an example, we

show how to test for the significance of $\xi'C_1(B)\sum_{i=1}^t \mathbf{v}^*_{(2),i}$ for $t \leq T_1$ when $r = 1$. In this case, $\xi'C_1(B)\sum_{i=1}^t \mathbf{v}^*_{(2),i} = \sum_{i=1}^t f_p(B)v_{p,i}$, $t \leq T_1$, where the variance of $v_{p,t}$ is estimated by λ_p, the smallest eigenvalue of Σ. (See Section 2. In practice, Σ is estimated through sample covariance $\hat{\Sigma}_{T_1}$.)

In (6), denote $\xi'\mathbf{X}_t$, $\xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi'\mathbf{W}_t$ and $\xi'C_1(B)\mathbf{v}^*_{(2),i}$ by y_t, z_t and u_i respectively, so that

$$y_t = z_t + \sum_{i=1}^t u_i, \ \ t \leq T_1. \tag{9}$$

In (9) z_t is stationary and can be expressed as $f(B)\varepsilon_t$ with i.i.d. $\varepsilon_t \sim N(0,\sigma_1^2)$; $u_t = f_p(B)v^*_{p,t}$ with i.i.d. $v^*_{p,t} \sim N(0,\sigma_2^2)$ (as mentioned before, σ_2^2 is estimated by λ_p); $\{\varepsilon_i\}$ are independent of $\{v^*_{p,t}\}$; and $f(B)$ and $f_p(B)$ are polynomial functions of the backward shift operator B. Thus we have $y_t = f(B)\varepsilon_t + f_p(B)\sum_{i=1}^t v^*_{p,i} = f(B)B^{T_1-t}\varepsilon_{T_1} + f_p(B)(\sum_{i=1}^t B^{T_1-i})v^*_{p,t}$, $t = 1,2,\cdots,T_1$. Let $\mathbf{y} = (y_1,y_2,\cdots,y_{T_1})'$. Then, $\mathbf{y} \sim N(0,\sigma_1^2\mathbf{\Sigma}_{T_1,1} + \sigma_2^2\mathbf{\Sigma}_{T_1,2})$, where $\Sigma_{T_1,1} = (<f(B)B^{T_1-i}, f(B)B^{T_1-j}>)_{T_1 \times T_1}$ and $\Sigma_{T_1,2} = (<f_p(B)\sum_{k=1}^i B^{T_1-k}, f_p(B)\sum_{k=1}^j B^{T_1-k}>)_{T_1 \times T_1}$. The inner product of $f(B) = \sum \varphi_{f,j}B^j$ and $g(B) = \sum \varphi_{g,j}B^j$ is defined by $<f(B),g(B)> = \sum_j \varphi_{f,j}\varphi_{g,j}$.

According to the definitions of $\mathbf{\Sigma}_{T_1,1}$ and $\mathbf{\Sigma}_{T_1,2}$, we can see that, once the functions $f(B)$ and $f_p(B)$ are given, $\mathbf{\Sigma}_{T_1,1}$ and $\mathbf{\Sigma}_{T_1,2}$ will be functions of T_1 only.

For $\mathbf{\Sigma}_{T_1,1}$ there is a matrix A_{T_1} such that $A_{T_1}\mathbf{\Sigma}_{T_1,1}A'_{T_1} = I$. Thus, $A_{T_1}\mathbf{y} \sim N(0,\sigma_1^2 I + \sigma_2^2 A_{T_1}\mathbf{\Sigma}_{T_1,2}A'_{T_1})$. For $\sigma_2^2(A_{T_1}\mathbf{\Sigma}_{T_1,2}A'_{T_1})$, there is an orthogonal matrix D_{T_1} such that $D_{T_1}D'_{T_1} = I$ and

$$D_{T_1}A_{T_1}\mathbf{\Sigma}_{T_1,2}A'_{T_1}D'_{T_1} = diag(\lambda_1^{(T_1)},\cdots,\lambda_{T_1}^{(T_1)}). \tag{10}$$

We now impose a condition:

Condition 1: The eigenvalues $\lambda_1^{(T_1)} \geq \cdots \geq \lambda_{T_1}^{(T_1)} \geq 0$ in (10) are not all equal and the difference between $\lambda_1^{(T_1)}$ and $\lambda_{T_1}^{(T_1)}$ is large (compared with σ_2^2).

The restriction on $\lambda_1^{(T_1)}$, $\cdots$, $\lambda_{T_1}^{(T_1)}$ given by Condition 1 is very weak (see the examples below). We always assume Condition 1 in the following discussion.

For any $0 < m < T_1$, $D_{T_1}A_{T_1}\mathbf{y}$ can be expressed as

$$D_{T_1}A_{T_1}\mathbf{y} = \begin{pmatrix} M_{T_1,1} \\ M_{T_1,2} \end{pmatrix} \mathbf{y},$$

with $M_{T_1,1}\mathbf{y} = (x_1, x_2, \cdots, x_m)' \sim N(0, \sigma_1 I + \sigma_2 diag(\lambda_1^{(T_1)}, \cdots, \lambda_m^{(T_1)}))$,
and $M_{T_1,2}\mathbf{y} = (x_{m+1}, x_{m+2}, \cdots, x_{T_1})' \sim N(0, \sigma_1 I + \sigma_2 diag(\lambda_{m+1}^{(T_1)}, \cdots, \lambda_{T_1}^{(T_1)}))$.
Thus

$$\frac{\sum_{i=1}^m \frac{x_i^2}{\sigma_1^2 + \sigma_2^2 \lambda_i^{(T_1)}}/m}{\sum_{i=m+1}^{T_1} \frac{x_i^2}{\sigma_1^2 + \sigma_2^2 \lambda_i^{(T_1)}}/(T-m)} \sim F_{m-1, T-m-1}.$$

Under Condition 1, the $\lambda_i^{(T_1)}$ are not all equal. Thus, as discussed in the previous section, if $(\sigma_2^2 \lambda_1^{(T_1)})/\sigma_1^2 < 0.01$, we will almost always accept $H_0 : \sigma_2 = 0$ at the level $\alpha \le 0.05$. This means that if $(\sigma_2^2 \lambda_1^{(T_1)})/(\sigma_1^2) < 0.01$ is true, we can accept that $\mathbf{y}_t$ is $I(0)$ at the significance level $\alpha \le 0.05$.

As mentioned in Section 1, for the study carried out in this paper, we always assume that $\xi'\mathbf{X}_t = y_t$ has been accepted as $I(0)$ for $t \le T_1$. In order to understand if $\{y_t\}_{t \le T}$ can be continuously accepted as $I(0)$ over an extended interval $(0, T]$ with $T > T_1$, what we need to check is whether $\frac{\sigma_2^2 \lambda_1^{(T)}}{\sigma_1^2} < 0.01$.

For $T > T_1$, the value of $\lambda_1^{(T)}$ can be determined by matrices $\Sigma_{T,1}$ and $\Sigma_{T,2}$, free from the knowledge of y_t and $\mathbf{X}_t$ for $t > T_1$.

Summary: For the above example, two steps have to be followed if we wish to check whether $\mathbf{X}_t$ can be continuously accepted as cointegrated in the extended time period $[0, T]$ with $T > T_1$.

Step 1: For given observations $\{\mathbf{X}_t\}_{0 < t \le T_1}$, (i) express $\mathbf{X}_t$ in the form (1); (ii) use RB procedure to identify ξ; (iii) express $z_t = \xi'(\mathbf{X}_0 - \mathbf{W}_0) + \xi'\mathbf{W}_t = f(B)\varepsilon_t$ and $u_t = \xi' C_1(B)v_{p,t} = f_p(B)v_{p,t}$; (iv) estimate $\sigma_1^2 = var(\varepsilon_1)$ and $\sigma_2^2 = var(v_{p,1})$ based on the information updated to T_1.

Step 2: (i) identify matrices $\Sigma_{T,1}$ and $\Sigma_{T,2}$ based on $f(B)$ and $f_p(B)$ given in Step 1; (ii) evaluate $\lambda_1^{(T)}$ from $\Sigma_{T,1}$ and $\Sigma_{T,2}$; (iii) compare $\sigma_2^2 \lambda_1^{(T)}/\sigma_1^2$ with 0.01.

In the following, a simulation study is carried out. For simplicity, we ignore Step 1 and assume that, based on the information of $\{\mathbf{X}_t\}_{0 < t \le T_1}$, we have already obtained $f(B)$, $f_p(B)$, σ_1^2 and σ_2^2 in (9). We assume that after Step 1, $\xi'\mathbf{X}_t = y_t$ is expressed as

$$y_t = \varepsilon_t + \sum_{k=1}^t (\phi_0 + \phi_1 B + \phi_2 B^2)\delta_k, \qquad 0 < t \le T_1, \qquad (11)$$

in which $\phi_0 = 0.1$, $\phi_1 = 0.01$ and $\phi_2 = 0.001$, ε_t are i.i.d. normally distributed with zero mean and variance $\sigma_{\varepsilon_1}^2 = \sigma_1^2 = 1$, δ_t are i.i.d. normal distributed with mean 0 and variance $\sigma_\delta^2 = \sigma_2^2$. Given T_1, since $f(B) = 1$

and $f_p(B) = \phi_0 + \phi_1 B + \phi_2 B^2$ are known, it is easy to evaluate $\lambda_1^{(T_1)}$ from $\Sigma_{T_1,1}$ and $\Sigma_{T_1,2}$.

In the following study, we use time series plots of simulation data to show how y_t changes as t increased for different values of $\sigma_2^2 \lambda_1^{(T_1)}/\sigma_1^2$. We put $T_1 = 100$ and consider two cases corresponding to $\sigma_2^2 = 0.00019$ and $\sigma_2^2 = 0.19$.

Based on Model (11), the minimum and maximum eigenvalues of (10) are 0.00207 and 50.33686 respectively when $T = 100$. Therefore, if $\sigma_2^2 = 0.00019$, the ratio $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 = 0.0095$; if $\sigma_2^2 = 0.19$, the ratio $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 = 9.5$. With these ratios we expect that, during $0 < t \le 100$, the impact of $\sum_{k=1}^{t} (\phi_0 + \phi_1 B + \phi_2 B^2)\delta_k$ on y_t given $\sigma_2^2 = 0.00019$ will be more significant than that given $\sigma_2^2 = 0.19$. This fact can be seen in Figures 2. Since given any T we can easily obtain $\lambda_1^{(T)}$ base on $\Sigma_{T,1}$ and $\Sigma_{T,2}$, we can conclude that y_t should be able to be continuously accepted as a stationary time series given $\sigma_2^2 = 0.00019$ if the time period is extended to $[0, 120]$ because $\lambda_1^{(T)} = \lambda_1^{(120)} = 72.37$ and $(\sigma_2^2 \lambda_1^{(120)})/\sigma_1^2 \approx 0.01$.
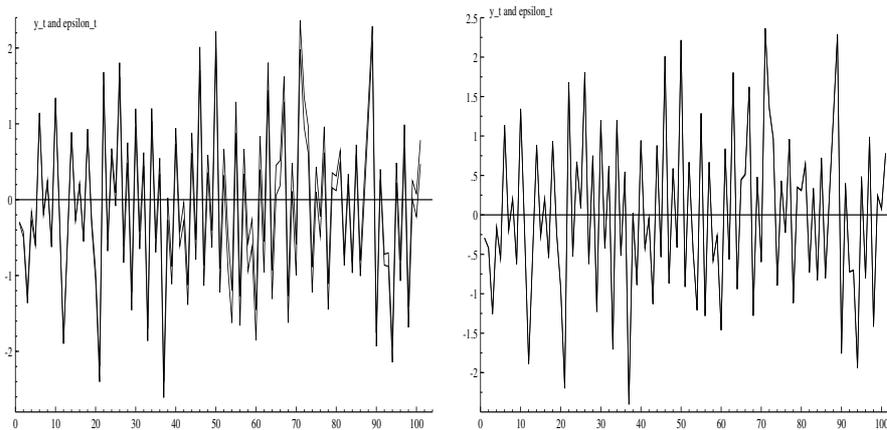

Figure 2. The plots of y_t and ε_t, for $\sigma_\delta^2 = 0.19$ and $\sigma_\delta^2 = 0.00019$ respectively.

Remark: The criterion $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 < 0.01$ is not a necessary condition for ensuring y_t be accepted as $I(0)$. Compared to the power of the DF test, the criterion $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 < 0.01$ is rather conservative. We independently simulated more than 100 sets of data from model (11) based on different values of σ_1 and σ_2 and found that in many cases we can always accept y_t as $I(0)$ by using DF test while the ratio $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 = 10$. Since, the DF

test is a common method for testing $I(0)$ series, the conservative criterion $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 < 0.01$ should be a fairly reasonable criterion for deciding whether a series can be continuously accepted as a stationary series when the time period is extended.

We have shown in Section 3 that $\max_{0<t\leq T}[var(\sum_{k=1}^{t} u_k)/\sigma_{z_1}^2]$ could be used to indicate the level of impact given by the $I(1)$ series in a model like (5). It is obvious that evaluating $\max_{0<t\leq T}[var(\sum_{k=1}^{t} u_k)/\sigma_{z_1}^2]$ is easier than evaluating $(\sigma_2^2 \lambda_1^{(T)}/\sigma_1^2$. It is of interest to ask whether the criterion $(\sigma_2^2 \lambda_1^{(T)})/\sigma_1^2 < 0.01$ can be replaced by $\max_{0<t\leq T}[var(\sum_{k=1}^{t} u_k)]/\sigma_{z_1}^2 < 0.01$. Based on our simulation studies, it seems the criterion $\max_{0<t\leq T}[var(\sum_{k=1}^{t} u_k)]/\sigma_{z_1}^2 < 0.01$ works well in practice. However, as yet we are unable to prove this.

References

1. R. Biondini, Y.-X. Lin and M. McCrae. A case study of the residual-based cointegration procedure. To appear in *J. App. Math. and Decision Science*.
2. D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Ass.* 74:427–431, 1979.
3. P. Dhrymes (1998). *Time Series, Unit Roots and Cointegration*. Sydney: Academic Press, 1998.
4. R. F. Engle and C. W. Granger. Co-integrated and error correction: representation, estimation and testing. *Econometrica* 55:251–276, 1987.
5. S. Johansen (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12:231–254, 1988.
6. D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin. Testing the null hypothesis of stationarity against the alternative of unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* 54:159–178, 1992.
7. Y.-X. Lin and M. McCrae. A new approach of determining cointegrating vectors. *MODSIM99 Proceedings, University of Waikato, Hamilton, New Zealand, 6-9th December 1999, Vol. 2*, 265–270, 1999.
8. Y.-X. Lin and M. McCrae. Estimation of cointegrating vectors based on residuals of ARIMA Models. *System Analysis - Modelling - Simulation, Vol. 41*, 735-753, 2001.
9. J. Otero and J. Smith. Testing for cointegration: power versus frequency of observation - further Monte Carlo results. *Economics Letters* 67:5–9, 2000.
10. P. Perron. Testing for a random walk: a simulating experiment of power when the sampling interval is varied. In *Advances in Econometrics and Modelling*. London: Kluwer, 1989.
11. R. J. Shiller and P. Perron. Testing the random walk hypothesis: power versus frequency of observation. *Economics Letters* 18:381–386, 1995.