Econometric Theory, **19**, 2003, 620–639. Printed in the United States of America.
DOI: 10.1017/S0266466603194078

INFERENCE ON SEGMENTED COINTEGRATION

JAE-YOUNG KIM
State University of New York at Albany
and
Seoul National University

A cointegration relation is often interpreted as a long-run equilibrium relation whereas short-run deviations are allowed to a certain extent. In practice, however, real data often fail to confirm cointegration for a well understood economic relation. This paper proposes that in many cases failure of confirming cointegration is due to nonstationary deviations in a relatively small portion of the data period ("short-run") whereas a cointegration relation prevails in the other periods. We call such a situation *segmented cointegration*. We study in this paper how to detect segmented cointegration and how to identify the period of short-run deviations.

1. INTRODUCTION

The concept of cointegration defined and introduced by Engle and Granger (1987) has become a useful concept for analyzing linear dynamic systems in economics. A cointegration relation is often interpreted as a long-run equilibrium in a system of variables whereas short-run deviations are allowed to a certain extent. Therefore, the cointegration concept applies for statistical inference on many economic theories that are concerned with equilibrium relationships. In practice, however, real data often fail to confirm cointegration for a well understood economic relation.

Failure of rejection of the null of no cointegration could be explained by a variety of reasons. For example, test results are sensitive to various factors such as the span of data set (Hendry, 1995), the choice of lags in the testing equations (Banerjee, Dolado, Galbraith, and Hendry, 1993), the possibility of structural breaks in the underlying relationship (Gregory and Hansen, 1996; Hansen and Johansen, 1993; Stock and Watson, 1994), and threshold effects in a possible cointegration relation (Balke and Fomby, 1996). Siklos and Granger (1997) consider the case when the given variables are cointegrated only during some periods and not at others, which is applied for interest rate parity of the United

I thank Pentti Saikkonen and three anonymous referees for helpful comments. I am deeply indebted to Peter Phillips for encouragement, discussion, and helpful comments on this work. The research of this paper is supported by the Research Grants Council of Hong Kong (grant HKUST6178/98H) and FRAP B of SUNY-Albany (account 320-9709W). Address correspondence to: Jae-Young Kim, School of Economics, Seoul National University, San 56-1 Shilim-dong Kwanak-gu, Seoul, Korea; e-mail: jykim017@snu.ac.kr.

 © 2003 Cambridge University Press 0266-4666/03 $12.00

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

States and Canada. Many of the preceding reasons would be related to the presence of temporary nonstationary deviations from a long-run relation.

In this paper we propose that failure of rejection of the null of no cointegration in many cases is due to nonstationary deviations in a relatively small portion of the data period ("short-run") whereas a cointegration relation prevails in the other periods. We call such a situation *segmented cointegration*. This situation is possible if some important events or changes in an economy interrupt the underlying economic relation for a certain length of time.[1] A trivial reason for the failure of confirming cointegration in the presence of such "short-run" nonstationary deviations is that an estimate of the cointegrating vector may be biased or not consistent as a result of the short-run nonstationarity. As we show in this paper, however, even when the cointegrating vector is known or correctly estimated, the chance of failure of confirming cointegration can be high in the presence of short-run nonstationary deviations. In general, the chance of the failure of confirming cointegration depends on various factors such as duration and the level of the short-run nonstationary deviations and the root of the stationary cointegration error. A Monte Carlo study shows that a relatively short duration of nonstationary deviations from cointegration may cause failure of confirming cointegration.

We study in this paper how to detect segmented cointegration and how to identify the period of short-run deviations. An idea for detecting a segmented cointegration is to use a test for testing the null of no cointegration for each of the two regressions, one regression based on the sample without segmentation and the other based on the sample with segmentation. The regression based on segmentation is a special case of the weighted least square, which is the only way of getting a consistent estimator under the alternative of segmented cointegration. As the period of noncointegration is unknown in practice, we search over a set of all possible segmentations. If there exists a segmentation for which the null is rejected whereas the null is not rejected for the sample without segmentation, we can conclude that the basic cointegration relation prevails in the sample with the segmentation. The actual testing procedure is to determine whether the smallest value of a statistic over a set of possible segmentations falls in a critical region. The reason for using the smallest value of a test statistic is that it constitutes evidence against the null hypothesis. As shown in Section 3, the tests based on segmentation are consistent tests. Monte Carlo simulation is performed for exploring performance of the procedures for detecting segmented cointegration. For identifying the period of noncointegration we propose an extremum estimator. We show that the estimator is a consistent estimator with T-rate of convergence. Alternatively, the noncointegration period can be identified by the period that corresponds to the segmentation yielding the minimum value of the test statistic for testing the existence of segmented cointegration.

The paper is organized as follows. Section 2 introduces definitions and assumptions. Also, Section 2 explains why real data often fail to confirm a co-

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

integration relation in the presence of short-run deviations. Section 3 discusses how to detect segmented cointegration and how to locate the period of noncointegration. A set of Monte Carlo results is provided in Section 4. Section 5 concludes the paper.

2. PRELIMINARIES

2.1. Basic Concepts and Notations

Let x_{1t} be a scalar $I(1)$ process and x_{2t} be an $(n-1)$ vector $I(1)$ process. Suppose that these variables have one of the following relations:

$$x_{1t} = \beta' x_{2t} + \varepsilon_t, \tag{2.1a}$$

$$x_{1t} = \alpha + \beta' x_{2t} + \varepsilon_t, \tag{2.1b}$$

$$x_{1t} = \alpha + \gamma t + \beta' x_{2t} + \varepsilon_t. \tag{2.1c}$$

Define the innovation vector $u_t = \Delta x_t$ or $u_t = \Delta x_t - \gamma$ in the presence of some trending variable(s). Also, define its cumulative process $S_t = \sum_{i=1}^{t} u_i$ and its long-run variance $\Omega = \lim_T T^{-1} E S_T S_T'$. We use the following regularity conditions.

Assumption 1.

(a) $\{u_t\}$ is mean-zero and strong mixing with mixing coefficient of size $-r\kappa/(r-\kappa)$, and $E|u_t|^r < \infty$ for some $r > \kappa > \frac{5}{2}$.
(b) The matrix Ω exists with finite elements and is positive definite.
(c) x_0 is a random vector with $E|x_0| < \infty$.

The conditions in Assumption 1 allow a broad class of weakly dependent time series and have been used by Phillips (1987), Phillips and Perron (1988), Phillips and Ouliaris (1990), and Gregory and Hansen (1996), among others, to study limiting behavior of some stochastic processes.

Now, consider the possibility that the error term ε_t is a unit root process in a subset of the data period, whereas it may follow a stationary process for the rest of the data period. We reserve notation for these two different types of periods, $\mathcal{N}_T$ and $\mathcal{C}_T$, respectively. We assume that $\mathcal{N}_T$ consists of a single interval of periods. More specifically,

$$\mathcal{N}_T = \{k_0 + 1, \ldots, k_1\}$$

for two time periods k_0 and k_1, $k_0 < k_1$.

Assumption 2. Let $\#(\mathcal{C}_T)$ be the number of elements in $\mathcal{C}_T$. We assume that $\#(\mathcal{C}_T) \geq n$.

Assumption 2 rules out the case of a "trivial" segmented cointegration with $\#(\mathcal{C}_T) \leq (n-1)$: for example, for $n = 2$ if $\#(\mathcal{C}_T) = (n-1) = 1$ then any

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

process forms a segmented cointegration by defining $\beta = x_{1h}/x_{2h}$ and $\varepsilon_h = 0$ for each $h = 1,\ldots,T$.

When $\mathcal{N}_T$ consists of more than one interval of periods, we need to determine the number of segmentations in practice. Formal analysis of the problem of multiple segmentations is beyond the scope of this paper. However, we briefly discuss this problem in Section 3.

Testing for cointegration for (2.1) is often based on testing for the null of no cointegration. Thus, writing the disturbance process as an AR(1):

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t, \tag{2.2}$$

the hypotheses are for the value of ρ: the null of $\rho = 1$ vs. $|\rho| < 1$. There are several different tests for testing such hypotheses, that is, Dickey–Fuller (1979) tests, Phillips (1987) tests, Phillips and Perron (1988) tests, Phillips and Ouliaris (1990) tests, and augmented Dickey–Fuller tests.

In the case of possible segmented cointegration we can consider the following hypotheses:

$$H_0 : \rho = 1 \quad \text{for } t \in \{t = 1,\ldots,T\}, \tag{2.3}$$

$$H_1 : \rho = \begin{cases} 1 & \text{for } t \in \mathcal{N}_T, \\ r_1, \quad \text{with } |r_1| < 1 & \text{for } t \in \mathcal{C}_T, \end{cases}$$

where $\mathcal{N}_T \subset \{1,\ldots,T\}$ and $\mathcal{C}_T \equiv \{1,\ldots,T\}\backslash\mathcal{N}_T$.

As an example of the situation implied in H_1 we can consider an economy that runs in two different regimes. For example, when the monetary authority changes the policy regime from targeting the interest rate to targeting money volume, the interest rate may be stationary only in the first regime.

A situation similar to that in H_1 is found in the following model of threshold cointegration studied in Balke and Fomby (1996), among others:

$$\varepsilon_t = \rho^{(t)}\varepsilon_{t-1} + \nu_t$$

$$\rho^{(t)} = \begin{cases} 1 & \text{if } |\varepsilon_{t-1}| \le \theta, \\ \rho, \quad \text{with } |\rho| < 1 & \text{if } |\varepsilon_{t-1}| > \theta. \end{cases} \tag{2.4}$$

When the "threshold" θ is a sufficiently small constant, then the preceding process ε_t is effectively a stationary process as is the case in Balke and Fomby (1996). However, if the threshold θ is not a small number or is a random variable, which sometimes has high values, the process ε_t may effectively be nonstationary. On the other hand, there is an important difference between segmented cointegration and threshold cointegration: segmented cointegration assumes a deterministic interval of cointegration whereas threshold cointegration depends on the conditional variables.

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

2.2. Failure to Confirm a Cointegration Relation

In this subsection we briefly examine why usual tests fail to confirm cointegration in the presence of short-run nonstationary deviations. To see this consider an example of the system (2.1a) under H_1 where ε_t is such that

$$\varepsilon_t = \rho^{(t)}\varepsilon_{t-1} + \nu_t, \tag{2.5}$$

$$\rho^{(t)} = \begin{cases} \rho, & \text{with } |\rho| < 1 \quad \text{for } t \in \mathcal{C}_T, \\ 1 & \text{for } t \in \mathcal{N}_T, \end{cases} \tag{2.6}$$

where $\mathcal{N}_T = \{k_0 + 1, \ldots, k_1\}$ for two time periods k_0 and k_1, $k_0 < k_1$.

A trivial reason for the failure of confirming cointegration in the presence of short-run deviations is that the cointegrating vector may not be consistently estimated. This is because the short-run nonstationary deviations dominate in the estimate of the cointegrating vector, which very likely makes the regression spurious. If the cointegrating vector is not correctly estimated, the residual contains nonstationary regressors: in (2.1a) if $\hat{\beta} \neq \beta$, then the residual e_t becomes

$$e_t = x_{1t} - \hat{\beta}' x_{2t} = \varepsilon_t + (\beta - \hat{\beta})' x_{2t}.$$

In the presence of short-run nonstationary deviations, however, even when the cointegrating vector is known or is correctly estimated, the chance of failure of confirming cointegration of usual tests can be very high. Thus, suppose that β is known in (2.1a). Then, the simple Dickey–Fuller t-statistic for testing the null H_0 (2.3) is

$$\begin{aligned} t &= \frac{T^{-1}\sum_{t=1}^{T}\varepsilon_{t-1}\nu_t}{\left\{T^{-2}\sum_{t=1}^{T}\varepsilon_{t-1}^2\right\}^{1/2}\{s_T^2\}^{1/2}} \\ &= \frac{T^{-1}\sum_{t\in\mathcal{C}_T}\varepsilon_{t-1}\nu_t + T^{-1}\sum_{t\in\mathcal{N}_T}\varepsilon_{t-1}\nu_t}{\left\{T^{-2}\sum_{t\in\mathcal{C}_T}\varepsilon_{t-1}^2 + T^{-2}\sum_{t\in\mathcal{N}_T}\varepsilon_{t-1}^2\right\}^{1/2}\{s_T^2\}^{1/2}}, \end{aligned} \tag{2.7}$$

where $s_T^2 = \sum_{t=1}^{T}(\varepsilon_t - \hat{\rho}\varepsilon_{t-1})^2/(T-1)$. But, since $\sum_{t\in\mathcal{C}_T}\varepsilon_{t-1}^2 = O_p(T)$ and $\sum_{t\in\mathcal{C}_T}\varepsilon_{t-1}\nu_t = O_p(T^{1/2})$, the preceding t-statistic becomes

$$t = \frac{o_p(1) + T_{\mathcal{N}}^{-1}\sum_{t\in\mathcal{N}_T}\varepsilon_{t-1}\nu_t}{\left\{o_p(1) + T_{\mathcal{N}}^{-2}\sum_{t\in\mathcal{N}_T}\varepsilon_{t-1}^2\right\}^{1/2}\{s_T^2\}^{1/2}}, \tag{2.8}$$

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

where $T_{\mathcal{N}}$, the number of elements in $\mathcal{N}_T$, replaces T in the t-statistic. Therefore, the behavior of the t-statistic in (2.8) under H_1 for the model (2.5) and (2.6) is arbitrarily close to that under H_0 (2.3). In fact, under the condition that $E|\varepsilon_{k_0}| < \infty$ asymptotic behavior of the t-statistic (2.8) is the *same* as that under H_0. But, for $t = k_0$ in a stationary regime the condition $E|\varepsilon_{k_0}| < \infty$ is not restrictive. As a consequence, the statistic (2.7) would very often fail to reject H_0 under H_1.

3. INFERENCE ON SEGMENTED COINTEGRATION

First, we study how to detect segmented cointegration defined in the previous section. Then, we discuss how to locate the period of noncointegration.

3.1. Detection of Segmented Cointegration

When the error term in a linear regression model such as (2.1) is heterogeneous across different periods, econometric theory suggests estimating the model by a weighted least square rather than by the ordinary least square:

$$\operatorname{argmin}_{\beta} Q(\beta, w) \equiv \sum_{t=1}^{T} w_t (x_{1t} - \beta' x_{2t})^2, \tag{3.1}$$

where w_t is a weight given to the t-period residual and $w = (w_1, \ldots, w_T)$. For the model (2.1) with ε_t having the property in H_1 the weight series $\{w_t\}$ should be such that

$$w_t = \begin{cases} 1 & \text{for } t \in \mathcal{C}_T, \\ 0 & \text{for } t \in \mathcal{N}_T, \end{cases} \tag{3.2}$$

where $w_t = 1$ for $t \in \mathcal{C}_T$ can be replaced by $w_t = \text{constant}$.[2] We use the notation $\mathcal{C}_T$ for a *trial* period for segmented cointegration and also for the *true* period of segmented cointegration, regarding which the reader should not be confused. Under H_0 the regression (2.1) is spurious for any choice of $\{w_t\}$ or $\mathcal{C}_T$. Also, under H_1 the regression yields a consistent estimator of the parameter for a proper choice of $\{w_t\}$ or $\mathcal{C}_T$. For this reason a test for segmented cointegration based on the estimation (3.1) with $\{w_t\}$ (3.2) would have desirable properties.

Notice that the weighting process $\{w_t\}$ depends on the selection of $\mathcal{C}_T$. Thus, we denote it by $w_t = w_t(\mathcal{C}_T)$. Also, denote by $\hat{\beta}(\mathcal{C}_T)$ and $e_t(\mathcal{C}_T)$, respectively, the estimator of β and the residual from (3.1) based on the weight $\{w_t(\mathcal{C}_T)\}$ for a segmentation $\mathcal{C}_T$. Let $\hat{\rho}(\mathcal{C}_T)$ be the least square estimator of the first-order serial correlation coefficient ρ of e_t in $e_t = \rho e_{t-1} + \nu_t$ for $e_t = e_t(\mathcal{C}_T)$ and $t(\mathcal{C}_T)$ be the t-test of $\rho = 1$. That is, $\hat{\rho}(\mathcal{C}_T)$ is obtained from regressing $w_t(\mathcal{C}_T)e_t$ on $w_{t-1}(\mathcal{C}_T)e_{t-1}$. Also, let $\hat{\sigma}^2_{\hat{\rho}}(\mathcal{C}_T)$ be the least square variance of $\hat{\rho}$ and $s^2(\mathcal{C}_T)$ be the least square estimate of the variance of the second stage residual:

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

$$s^2(\mathcal{C}_T) = (T_{\mathcal{C}} - 1)^{-1} \sum w_t(\mathcal{C}_T)(e_t - \hat{\rho} e_{t-1})^2,$$

$$\hat{\sigma}^2_{\hat{\rho}}(\mathcal{C}_T) = s^2(\mathcal{C}_T)\Big(\sum w_t(\mathcal{C}_T) e^2_{t-1}\Big)^{-1},$$

where $T_{\mathcal{C}}$ is the number of elements of $\mathcal{C}_T$. Define

$$Z_\rho(\mathcal{C}_T) = T_{\mathcal{C}}(\hat{\rho} - 1) - \frac{1}{2}(T^2_{\mathcal{C}} \hat{\sigma}^2_{\hat{\rho}}(\mathcal{C}_T)/s^2(\mathcal{C}_T))(\hat{\lambda}^2 - \hat{\gamma}_0), \tag{3.3}$$

$$Z_t(\mathcal{C}_T) = (\hat{\gamma}_0/\hat{\lambda}^2)^{1/2} \cdot t(\mathcal{C}_T) - \left\{\frac{1}{2}(\hat{\lambda}^2 - \hat{\gamma}_0)/\hat{\lambda}\right\} \times \{T_{\mathcal{C}} \hat{\sigma}_{\hat{\rho}}(\mathcal{C}_T)/s(\mathcal{C}_T)\}, \tag{3.4}$$

where $\hat{\lambda}^2$ and $\hat{\gamma}_0$, respectively, are estimators of $\lambda^2 = \lim_{T_{\mathcal{C}} \nearrow \infty} \times T_{\mathcal{C}}^{-1} E[(\sum_{t \in \mathcal{C}_T} \nu_t)^2]$ and $\gamma_0 = \lim_{T_{\mathcal{C}} \nearrow \infty} T_{\mathcal{C}}^{-1} \sum_{t \in \mathcal{C}_T} E(\nu_t^2)$.

Consistent nonparametric estimators of λ and γ_0 under H_1 are obtained as in the following expressions:

$$\hat{\gamma}_j = \frac{1}{T_{\mathcal{C}} - j} \sum_t w_t(\mathcal{C}_T) \hat{\nu}_t \hat{\nu}_{t-j} \quad \text{for } j = 0,1,\ldots,q,$$

$$\hat{\lambda} = \hat{\gamma}_0 + \sum_{j=1}^{q} \left(1 - \frac{j}{1+q}\right) \hat{\gamma}_j \quad \text{for an integer } q.$$

The approach given above is based on an AR(1) for the residual process e_t. Alternatively, we can consider the following variant of an augmented Dickey–Fuller expression for e_t:

$$\Delta e_t = \zeta_1 \Delta e_{t-1} + \zeta_2 \Delta e_{t-2} + \cdots + \zeta_{p-1} \Delta e_{t-p+1} + w_t(\mathcal{C}_T)(\rho - 1) e_{t-1} + \epsilon_t, \tag{3.5}$$

where $\{\epsilon_t\}$ is an independent and identically distributed (i.i.d.) sequence with mean zero, finite variance σ^2_ϵ, and finite fourth moment. Notice that in (3.5) the error correction term $(\rho - 1)e_{t-1}$ is multiplied by the "weight" w_t. Let $\tilde{\rho}$ and $\tilde{t}_T$, respectively, be the least square estimator of ρ in (3.5) and the t-test of $\rho - 1 = 0$.[3] Now, define

$$ADF_\rho(\mathcal{C}_T) \equiv T_{\mathcal{C}}(\tilde{\lambda}/\tilde{\sigma}_\epsilon)(\tilde{\rho} - 1), \tag{3.6}$$

$$ADF_t(\mathcal{C}_T) \equiv \tilde{t}_T, \tag{3.7}$$

where $\tilde{\lambda}/\tilde{\sigma}_\epsilon$ is consistently estimated by

$$\tilde{\lambda}/\tilde{\sigma}_\epsilon = (1 - \tilde{\zeta}_1 - \cdots - \tilde{\zeta}_{p-1})^{-1},$$

where $\tilde{\zeta}_j$ is the least square estimator of ζ_j in (3.5).

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

Let $\mathcal{T}_C$ be a fixed subset of $[0,1]$ such that $[\mathcal{T}_C T] = \mathcal{C}_T$ where $[\mathcal{T}_C T] = \{[\tau T] : \tau \in \mathcal{T}_C\}$ where $[\cdot]$ is the integer part of $\cdot$. Likewise, $[\mathcal{T}_N T] = \mathcal{N}_T$. Since $\mathcal{N}_T$ consists of a single interval of periods, we may write it as

$$\mathcal{N}_T = ([\tau_0 T] + 1, \ldots, [\tau_1 T])$$

for some $\tau_i \in \mathcal{T}_i$, $i = 0,1$ where $\mathcal{T}_0 \subset [0,1)$ and $\mathcal{T}_1 = (\sup \mathcal{T}_0, 1]$.

Now, for a given $\mathcal{T}_C$, the statistics $Z_\rho(\mathcal{C}_T)$, $Z_t(\mathcal{C}_T)$, $ADF_\rho(\mathcal{C}_T)$, and $ADF_t(\mathcal{C}_T)$ converge to some limits given in the following theorem.

THEOREM 1. *Assume the conditions in Assumption 1. Under H_0 it is true that*

$$Z_\rho(\mathcal{C}_T), ADF_\rho(\mathcal{C}_T) \Rightarrow Z_\rho(\mathcal{T}_C), \tag{3.8}$$

$$Z_t(\mathcal{C}_T), ADF_t(\mathcal{C}_T) \Rightarrow Z_t(\mathcal{T}_C), \tag{3.9}$$

where $Z_\rho(\mathcal{T}_C)$ and $Z_t(\mathcal{T}_C)$ are such that

$$Z_\rho(\mathcal{T}_C) = \frac{\int_{\mathcal{T}_C} W(r)\, dW(r)}{\int_{\mathcal{T}_C} [W(r)]^2\, dr}, \tag{3.10}$$

$$Z_t(\mathcal{T}_C) = \frac{\int_{\mathcal{T}_C} W(r)\, dW(r)}{\left\{\int_{\mathcal{T}_C} [W(r)]^2\, dr\right\}^{1/2}}, \tag{3.11}$$

where

$$W(r) = W_1(r) - \int_{\mathcal{T}_C} W_1 W_{2,m}' \left[\int_{\mathcal{T}_C} W_{2,m} W_{2,m}'\right]^{-1} W_{2,m}(r), \tag{3.12}$$

where $W_1(\cdot)$ is a standard Wiener process and $W_{2,m}(\cdot)$ depends on the model. If the residuals are from ordinary least squares (OLS) estimation of the model (2.1a), then

$$W_{2,m}(r) = W_2(r),$$

where $W_2(\cdot)$ is the $(n-1)$-dimensional standard Wiener process. If the residuals are from OLS estimation of the model (2.1b), then

$$W_{2,m}(r) = [1, W_2(r)']'.$$

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

If the residuals are from OLS estimation of the model (2.1c), then

$$W_{2,m}(r) = [1, r, W_2(r)']'.$$

In practice, the period $\mathcal{T}_C$ or $\mathcal{T}_N$ is not known. In this case we adopt the following approach: we get values of the statistic $Z(\mathcal{C}_T)$ for all the possible candidates of $\mathcal{C}_T$ and take the smallest value of the statistic. The smallest value constitutes evidence against the null hypothesis.

THEOREM 2. *Denote by* $Z^*(\mathcal{C}_T) = \inf_{\mathcal{C}_T} Z_T(\mathcal{C}_T)$, *etc. Assume the conditions in Assumption* 1. *Then, under* H_0 *it is true that*

$$Z_\rho^*(\mathcal{C}_T), ADF_\rho^*(\mathcal{C}_T) \Rightarrow \inf Z_\rho(\mathcal{T}_C), \tag{3.13}$$

$$Z_t^*(\mathcal{C}_T), ADF_t^*(\mathcal{C}_T) \Rightarrow \inf Z_t(\mathcal{T}_C). \tag{3.14}$$

Asymptotic critical values of these statistics can be obtained by simulating test statistics for a large sample size for a large number of replications. Because computation in the simulation for the present example for large samples is very time consuming we adopt a procedure due to Mackinnon (1991): using 10,000 replications for each of the sample sizes $T = 50$, 70, 100, 120, 150, 170, 200, 220, 250, 270, and 300 we obtain critical values $Crt(T, p, n)$, where p is the percentile. We then estimate by OLS for each p and n the response surface

$$Crt(T, p, n) = \psi_0 + \psi_1 T^{-1} + \psi_2 T^{-2} + error.$$

Various polynomials were tried (involving T^{-3}, T^{-4}, etc.), but the one given above appeared to be the best fit overall (R^2's were generally over 0.98). The asymptotic critical value is taken to be the OLS estimate $\hat{\psi}_0$. Although the above specification seemed to be the best overall, estimates of ψ_0 changed on occasion by a factor of 2 under alternative specifications. Also, we get the critical values of $\inf Z(\mathcal{T}_C)$ for some given upper limit of the length of $\mathcal{T}_N$ denoted by $\bar{\ell}(\mathcal{T}_N)$.[4] That is,

$$\inf Z_\rho(\mathcal{T}_C)_{\bar{\ell}(\mathcal{T}_N)} = \inf_{\mathcal{T}_C \in \{\mathcal{T}_C : \ell(\mathcal{T}_N) \leq \bar{\ell}(\mathcal{T}_N)\}} Z_\rho(\mathcal{T}_C), \tag{3.15}$$

where $\ell(\mathcal{T}_N)$ is the length of the interval $\mathcal{T}_N$. In Tables 1 and 2 asymptotic critical values of the minimum statistics in (3.13) and (3.14) are provided for $\bar{\ell}(\mathcal{T}_N) = 0.3$. In Tables 1 and 2, Case I is for the model (2.1a), Case II is for the model (2.1b), and Case III is for the model (2.1c).

In practice, critical values depend on the value of $\bar{\ell}(\mathcal{T}_N)$ chosen for simulation. We can show that the critical value for a given size of the test is a decreasing function of $\bar{\ell}(\mathcal{T}_N)$.

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

TABLE 1. Asymptotic critical values of $Z_\rho^*(\mathcal{C})$ and $ADF_\rho^*(\mathcal{C})$

n	0.01	0.025	0.05	0.10	0.15	0.95	0.975	0.99
				Case I				
1	−13	−10.16	−8.18	−5.68	−4.57	1.29	1.64	2.12
2	−37.2	−32.3	−27.9	−23.7	−20.82	−2.4	−1.29	−0.56
3	−46.63	−41.05	−36.41	−31.61	−28.74	−7.54	−5.81	−4.11
4	−55.89	−48.97	−44.59	−39.43	−36.36	−12.73	−10.85	−9.22
5	−63.94	−58.4	−52.79	−47.78	−44.59	−17.92	−15.86	−13.75
6	−70.39	−64.3	−59.83	−54.22	−51.02	−22.47	−20.6	−18.42
				Case II				
1	−20.15	−16.73	−13.96	−11.37	−9.59	−0.13	0.5	1.21
2	−87.37	−64.3	−50.75	−39.58	−34.02	−7.77	−6.19	−4.79
3	−105.48	−84.12	−65.53	−51.53	−44.82	−13.57	−11.25	−9.21
4	−122.91	−96.68	−78.66	−61.95	−54.24	−18.65	−16.37	−13.77
5	−130.86	−106.23	−87.07	−69.28	−60.56	−23.79	−21.33	−18.94
6	−134.72	−109.84	−89.89	−73.65	−66.98	−28.72	−25.86	−22.5
				Case III				
1	−29.17	−24.9	−21.55	−18.22	−16.03	−2.6	−1.67	−0.75
2	−107.65	−80.24	−62.54	−46.59	−38.44	−8.26	−6.42	−4.39
3	−131.66	−101.33	−79.34	−58.96	−49.43	−14.13	−12.05	−9.81
4	−135.59	−109.94	−87.87	−68.31	−59.07	−19.65	−17.64	−15.2
5	−140.29	−115.59	−96.57	−74.32	−65.16	−24.78	−22.25	−19.61
6	−144.09	−118.83	−98.31	−78.46	−70.02	−28.98	−26.27	−23.71

Note: n = the number of variables in a cointegration regression. The case of $n = 1$ is for a univariate process.

LEMMA 1. *Let* $\bar{\ell}_1(\mathcal{T}_\mathcal{N}) < \bar{\ell}_2(\mathcal{T}_\mathcal{N})$. *Denote by* $Z^*(\mathcal{C}_T)_{\alpha,\bar{\ell}}$ *the lower* α*-size critical value of* $Z^*(\mathcal{C}_T)$ *obtained based on* $\bar{\ell}(\mathcal{T}_\mathcal{N})$. *Then, it is true that* $Z^*(\mathcal{C}_T)_{\alpha,\bar{\ell}_2} < Z^*(\mathcal{C}_T)_{\alpha,\bar{\ell}_1}$.

Therefore, the test (3.13) or (3.14) based on critical values $\inf Z(\mathcal{T}_C)_{\alpha,\bar{\ell}_2}$ for a given $\bar{\ell}_2$ is a conservative test when the actual test statistic $Z^*(\mathcal{C}_T)$ is computed for $\bar{\ell}_1$ such that $\bar{\ell}_1 < \bar{\ell}_2$.[5]

As was explained in Section 2, we consider in this paper the case when $\mathcal{N}_T$ consists of a single interval of periods. In practice a researcher would have situations where the number of such intervals is possibly more than one. In this case the researcher would have to determine the number of such intervals. Formal analysis of the problem of multiple segmentations including determination of the number of such intervals is beyond the scope of this paper. However, in this case we could suggest adopting the idea of detecting one by one as in Bai (1997). To describe it briefly, the approach is to divide the sample into several successive subsamples and detect the existence of segmented cointegration in

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

TABLE 2. Asymptotic critical values of $Z_t^*(\mathcal{C})$ and $ADF_t^*(\mathcal{C})$

n	0.01	0.025	0.05	0.10	0.15	0.95	0.975	0.99
				Case I				
1	−2.51	−2.2	−1.96	−1.61	−1.43	1.29	1.7	2.17
2	−4.23	−3.95	−3.65	−3.34	−3.15	−0.88	−0.54	−0.23
3	−4.88	−4.49	−4.23	−3.92	−3.73	−1.74	−1.46	−1.15
4	−5.26	−4.89	−4.68	−4.38	−4.19	−2.32	−2.1	−1.84
5	−5.59	−5.33	−5.09	−4.86	−4.7	−2.84	−2.62	−2.33
6	−5.89	−5.68	−5.45	−5.16	−5	−3.24	−3.05	−2.81
				Case II				
1	−3.49	−3.12	−2.88	−2.58	−2.37	−0.06	0.31	0.66
2	−8.84	−7.43	−6.36	−5.31	−4.71	−1.91	−1.62	−1.28
3	−10.05	−8.68	−7.38	−6.22	−5.47	−2.53	−2.28	−2.01
4	−11.04	−9.51	−8.25	−6.86	−6.05	−3.03	−2.78	−2.51
5	−11.33	−10.03	−8.7	−7.21	−6.35	−3.43	−3.2	−2.93
6	−11.78	−10.14	−8.76	−7.33	−6.56	−3.78	−3.58	−3.37
				Case III				
1	−3.92	−3.66	−3.4	−3.12	−2.94	−0.9	−0.62	−0.3
2	−10.42	−8.78	−7.74	−6.32	−5.52	−1.96	−1.67	−1.27
3	−11.68	−9.98	−8.62	−7.23	−6.2	−2.61	−2.27	−1.91
4	−11.9	−10.37	−9.13	−7.57	−6.62	−3.1	−2.87	−2.61
5	−12.32	−10.8	−9.52	−7.83	−6.86	−3.51	−3.3	−3.05
6	−12.45	−10.98	−9.53	−7.79	−6.91	−3.84	−3.62	−3.38

Note: n = the number of variables in a cointegration regression. The case of $n = 1$ is for a univariate process.

each subsample based on the approach discussed in this section. This approach would save us a lot of computation time for detecting multiple intervals of noncointegration periods when the number of such intervals is unknown.

The tests for segmented cointegration $Z^*(\mathcal{C}_T)$ and $ADF^*(\mathcal{C}_T)$ are consistent tests, which is shown by the following lemma.

LEMMA 2. *Let* $\{x_t\}$, $x_t = (x_{1t}, x_{2t}')'$, *be an n-vector* $I(1)$ *process where* x_{2t}, *an* $(n-1)$ *vector* $I(1)$ *process, has a positive definite long-run variance-covariance matrix* $\Omega_{22} \equiv \lim_{T\nearrow\infty} T^{-1}E[(\sum_{t=1}^{T}\Delta x_{2t})(\sum_{t=1}^{T}\Delta x_{2t}')]$. *Suppose that* $\varepsilon_t = x_{1t} - \beta' x_{2t}$ *is* $I(0)$ *in a subset of data* $\mathcal{C}_T$ *where* ε_t *has a continuous spectral density* $f_{\varepsilon\varepsilon}$ *with* $f_{\varepsilon\varepsilon}(0) > 0$ *in* $\mathcal{C}_T$. *Then,*

$$Z_\rho^*(\mathcal{C}_T) = O_p(T),$$

$$Z_t^*(\mathcal{C}_T) = O_p(T^{1/2}),$$

$$ADF^*(\mathcal{C}_T) = O_p(T^{1/2}).$$

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

A Monte Carlo result in Section 4 shows that empirical power values of the tests (3.13) and (3.14) are quite high for a 5% test and for $T = 200$.

Remark. A model related to the one in our alternative hypothesis (2.4) has been considered by Leybourne, McCabe, and Tremayne (1996) and McCabe and Tremayne (1995):

$$\varepsilon_t = \rho_t \varepsilon_{t-1} + \nu_t,$$

where ρ_t is assumed to be an independent process. McCabe and Tremayne (1995) study how to determine whether or not the null of nonstationarity, $\rho_t = 1$, is rejected against the alternative of the root being independently changing, and Leybourne, McCabe, and Tremayne (1996) provide some applications of McCabe and Tremayne (1995). In our approach the model for alternative hypothesis is a special case of that in McCabe and Tremayne. However, the model for our alternative hypothesis allows richer statistical inference than that of McCabe and Tremayne: First, our alternative allows us to analyze a cointegration relation as well as observable individual series whereas the alternative of McCabe and Tremayne is adequate only for analyzing the latter as ρ_t in McCabe and Tremayne has to be an independent process. Second, our approach allows us to identify the two distinct periods, the period of noncointegration (nonstationarity) and the period of cointegration (stationarity), whereas the approach in McCabe and Tremayne, by its nature, is not appropriate for identifying or estimating the varying root.

3.2. Identifying the Period of Noncointegration

When a segmented cointegration is detected by the procedure discussed previously, we are interested in the location of the period of noncointegration. In this section we study how to identify the location of the period of noncointegration.

As in Section 3.1 (before Theorem 1), let $\mathcal{N}_T = ([\tau_0 T] + 1, \ldots, [\tau_1 T])$ for some $\tau_i \in \mathcal{T}_i$, $i = 0,1$ where $\mathcal{T}_0 \subset [0,1)$, $\mathcal{T}_1 = (\sup \mathcal{T}_0, 1]$, and $[\cdot]$ is the integer part. Notice that we allow the possibility that the nonintegration period lies at the beginning or at the end of the sample. Now, let $\hat{\beta}(\mathcal{C}_T)$ be the estimator of β from (3.1) based on $\{w_t(\mathcal{C}_T)\}$ in (3.2) for $\mathcal{C}_T \equiv \{1, \ldots, T\} \backslash \mathcal{N}_T$. Construct a residual process $\{e_t(\mathcal{C}_T)\}$ from the estimate $\hat{\beta}(\mathcal{C}_T)$ for *all* $t = 1, \ldots, T$. For a given $\tau = (\tau_0, \tau_1)$, let $\Lambda_T(\tau)$ be such that

$$\Lambda_T(\tau) = \frac{[(\tau_1 - \tau_0)T]^{-2} \sum_{t \in \mathcal{N}_T} e_t(\mathcal{C}_T)^2}{T_{\mathcal{C}}^{-1} \sum_{t \in \mathcal{C}_T} e_t(\mathcal{C}_T)^2}. \tag{3.16}$$

Also, let $\hat{\tau} = (\hat{\tau}_0, \hat{\tau}_1)$ be such that

$$\hat{\tau} = \operatorname{argmax}_{\tau \in \mathcal{T}} \Lambda_T(\tau), \tag{3.17}$$

where $\mathcal{T} = \mathcal{T}_0 \times \mathcal{T}_1$.

We assume the following condition.

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

Assumption 3. Under H_1 let $e_{s+1}, e_{s+2}, \ldots, e_{s+q}$ for $s \in \{0, \ldots, T-1\}$ and $q \leq T - s$ be a sequence of residuals of (2.1) in the stationary regime $\mathcal{C}_T$. Assume that $q^{-1} \sum_{t=s+1}^{s+q} e_t^2 \to E[e^2]$ for $E[e^2] < \infty$.

Assumption 3 describes a law of large numbers for a sequence of stationary e's, which usually holds for least square residuals such as $\{e_t\}$ in the previous section.

THEOREM 3. *Suppose that there exists $\mathcal{C}_T$ such that $\varepsilon_t \sim I(0)$ for $t \in \mathcal{C}_T$ whereas $\varepsilon_t \sim I(1)$ for $t \in \{1,2,\ldots,T\} \backslash \mathcal{C}_T$. Assume that the condition of Assumption 3 is satisfied. Then, it is true that for $i = 0,1$*

(a) $(\hat{\tau}_i - \tau_i) = o_p(1)$,
(b) $T(\hat{\tau}_i - \tau_i) = O_p(1)$.

Another approach to identifying $\mathcal{N}_T$ is to find the set $\tilde{\mathcal{N}}_T$ for which the p-value of the test statistic $Z(\mathcal{C}_T)$ or $ADF(\mathcal{C}_T)$ is the smallest: denoting $\boldsymbol{T} = \{1,2,\ldots,T\}$,

$$\tilde{\mathcal{N}}_T = \text{argmin}_{\mathcal{N}_T} \text{Prob}[Z < Z(\boldsymbol{T} \backslash \mathcal{N}_T)].$$

This approach is well approximated by

$$\tilde{\mathcal{N}}_T = \text{argmin}_{\mathcal{N}_T} Z(\boldsymbol{T} \backslash \mathcal{N}_T) \tag{3.18}$$

when the value of the Z-statistic does not heavily depend on the size of $\boldsymbol{T} \backslash \mathcal{N}_T (= \mathcal{C}_T)$. The term Z_t or ADF_t is relatively less dependent on the sample size than Z_ρ or ADF_ρ. In an empirical example in Kim (2001), we find that the two approaches (3.17) and (3.18) yield similar results but with the tendency of $\tilde{\tau}_0 > \hat{\tau}_0$ and $\tilde{\tau}_1 < \hat{\tau}_1$ where $\tilde{\tau}$ is the estimate of τ from (3.18).

Kim (2000) considers change in persistence of linear time series, considers change from stationarity to nonstationarity and vice versa, and studies how to detect the existence of such change and how to identify the period of such change. Chong (2001) considers structural change in AR(1) models including change from stationarity to a unit root and vice versa and studies how to locate the period of such change. The subjects of these papers are related to the work in this paper. However, there is difference in some important aspects: first, whereas this paper studies change in persistence of the (unobservable) error of a cointegration relation, the two previous works study similar change in an observable series. Statistical treatment in practice for the former case is fundamentally different from that for the latter case. Second, whereas the main interest of this paper is the existence of a stable relation in the majority of the data period (segmented cointegration), the two previous works are interested in a one-time break in the associated property of a series.

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

Another related work found in the literature is Bai, Lumsdaine, and Stock (1998). Bai et al. study how to test for the existence of and date a structural break in a cointegration relation. Their analysis includes a vector error correction model as well as a single-equation cointegration relation. The subject of Bai et al. is related to that of this paper in that both are discussing inference on a cointegration relation with possible turbulence. However, the subjects of the two papers are different in an important aspect: whereas Bai et al. study the problem of a structural break (parameter change) assuming that cointegration relation(s) hold in both periods before and after the break, this paper considers the case when there exists an *interval* of periods in which a (stable) cointegration relation does not exist. Statistical inference on the two systems, testing for the existence of such systems and dating the period(s) of break(s), as well as economic implications of the two systems are quite different. The main reason for the difference in inference is that the disturbance of the system considered in this paper undergoes change from stationarity to nonstationarity (and vice versa later) whereas there does not exist such change in the system considered in Bai et al.

4. MONTE CARLO STUDY

In this section we perform a Monte Carlo study to examine properties of the procedures discussed in the previous section. First, we investigate the behavior of the test statistics under the null H_0. More specifically, we get the actual sizes of the tests $Z^*(\mathcal{C}_T)$ and $ADF^*(\mathcal{C}_T)$ for testing H_0. For this purpose, we consider three cases of $n = 1,2,3$ and also three different cases for the underlying model.

Case I. The estimated model is (2.1a), no drift in the variables.

Case II. The estimated model is (2.1b), no drift in the variables.

Case III. The estimated model is (2.1c), drift in some variable(s).

Each variable follows a random walk with an i.i.d. normal error with variance 0.01. The drift in Case III is set to be 0.1. The simulation is based on samples of size 200 with 2,000 replications. The values of test statistics are obtained under H_0 from the segmentation $\mathcal{N}_T = (0.4T, 0.6T)$. Table 3 shows empirical sizes of the tests $Z^*(\mathcal{C}_T)$ and $ADF^*(\mathcal{C}_T)^*$ at the 5% theoretical size. Overall, the Phillips–Perron–Ouliaris Z-tests show better size property than ADF tests. Since the values of test statistics are obtained from the segmentation of $\ell(\mathcal{C}_T) = 0.2$, $\mathcal{N}_T = (0.4T, 0.6T)$, whereas we use corresponding critical values for $\bar{\ell}(\mathcal{C}_T) = 0.3$, we expect that the tests would be conservative tests, having smaller actual sizes than the true size of 5%. This is evident for most cases of $n = 2,3$ for Z^*.

Next, we investigate the behavior of the tests under the alternative hypothesis H_1, power properties of the tests. This is to evaluate the results of Lemma 2 by simulation. The simulation is based on samples of size 200 with 2,000 replications. We consider the following cointegration regression with an AR(1) error:

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

TABLE 3. Empirical sizes of four tests of segmented cointegration (true size = 0.05)

		$Z_\rho^*(\mathcal{C})$	$Z_t^*(\mathcal{C})$	$ADF_\rho^*(\mathcal{C})$	$ADF_t^*(\mathcal{C})$
$n = 1$	Case I	0.048	0.0652	0.148	0.3882
	Case II	0.0542	0.0642	0.1652	0.3728
	Case III	0.0556	0.0892	0.0094	0.1168
$n = 2$	Case I	0.0298	0.0528	0.1292	0.292
	Case II	0.0132	0.0204	0.0064	0.0146
	Case III	0.0056	0.0068	0.0024	0.0064
$n = 3$	Case I	0.0286	0.054	0.0968	0.2426
	Case II	0.008	0.0118	0.0012	0.0088
	Case III	0.0044	0.0068	0.001	0.0042

$$x_{1t} = \alpha + \beta x_{2t} + \varepsilon_t,$$

$$x_{2t} = x_{2t-1} + v_{2t},$$

$$\varepsilon_t = \begin{cases} \rho\varepsilon_{t-1} + \nu_t, & t \in \mathcal{C}_T, \\ \varepsilon_{t-1} + \nu_t, & t \in \mathcal{N}_T, \end{cases} \tag{4.1}$$

where $\mathcal{N}_T = (0.4T, 0.6T)$, $\beta = 1$, $\alpha = 0$, and $\nu_t \sim i.i.d.N(0, \sigma^2)$ for $\sigma^2 = 0.1$. We consider three different values of ρ, $\rho = 0.85, 0.9, 0.95$.

Tables 4 and 5 show power values of the tests $Z^*(\mathcal{C}_T)$ and $ADF^*(\mathcal{C}_T)$ for testing H_0 against H_1 under three different values of ρ in H_1 for the 5% test. Also, in Tables 4 and 5 we have power values of ordinary $Z(T)$ and $ADF(T)$

TABLE 4. Power of P-P-O type tests of segmented cointegration (size = 0.05)

	$\rho = 0.85$		$\rho = 0.9$		$\rho = 0.95$	
	$Z_\rho^*(\mathcal{C})$	$Z_\rho(T)$	$Z_\rho^*(\mathcal{C})$	$Z_\rho(T)$	$Z_\rho^*(\mathcal{C})$	$Z_\rho(T)$
$n = 2$	0.523	0.005	0.265	0	0.071	0.002
$n = 3$	0.144	0	0.052	0	0.011	0
$n = 4$	0.631	0	0.288	0	0.133	0
	$Z_t^*(\mathcal{C})$	$Z_t(T)$	$Z_t^*(\mathcal{C})$	$Z_t(T)$	$Z_t^*(\mathcal{C})$	$Z_t(T)$
$n = 2$	0.266	0	0.086	0	0.039	0
$n = 3$	0.95	0	0.744	0	0.45	0
$n = 4$	0.768	0	0.476	0	0.242	0

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

TABLE 5. Power of ADF tests of segmented cointegration (size = 0.05)

	$\rho = 0.85$		$\rho = 0.9$		$\rho = 0.95$	
	$ADF^*_\rho(\mathcal{C})$	$ADF_\rho(T)$	$ADF^*_\rho(\mathcal{C})$	$ADF_\rho(T)$	$ADF^*_\rho(\mathcal{C})$	$ADF_\rho(T)$
$n = 2$	0.475	0	0.341	0	0.195	0
$n = 3$	0.045	0	0.028	0	0.02	0
$n = 4$	0.136	0	0.087	0	0.04	0
	$ADF^*_t(\mathcal{C})$	$ADF_t(T)$	$ADF^*_t(\mathcal{C})$	$ADF_t(T)$	$ADF^*_t(\mathcal{C})$	$ADF_t(T)$
$n = 2$	0.007	0	0.008	0	0	0
$n = 3$	0.027	0.008	0.007	0.01	0.012	0.002
$n = 4$	0.001	0	0	0	0	0

statistics where $Z(T)$ and $ADF(T)$, respectively, are the Z- and ADF-statistics obtained from the nonsegmented OLS based on the sample generated from the model (4.1). Power values of $Z(T)$ and $ADF(T)$ are very low for all the cases under study. This implies that the test obtained from the nonsegmented OLS tends to lead to the conclusion of noncointegration even if a cointegration relation prevails in 80% of the sample period, $[0,1]\backslash[0.4,0.6]$, an implication of Section 2.2. On the other hand, power values of $Z^*(\mathcal{C}_T)$ and $ADF^*(\mathcal{C}_T)$ are quite high for all the cases under study, a result supporting the conclusion of Lemma 2.

5. CONCLUSION

Many empirical works find that real data often fail to confirm cointegration for some well understood economic relations. Also, if a cointegration is confirmed by real data, the evidence is quite weak in many cases. We hope that this paper provides a probe to examine whether there indeed exists a cointegration relation except for some exceptional periods in the data. In Kim (2001) the Fisher relation is studied based on the procedure proposed in this paper. For U.S. postwar quarterly data, Kim (2001) finds that short-run nonstationary deviations in the data cause failure of the Fisher relation and that there exists a stable Fisher relation in the postwar era except for a period of several years. The Fisher relation is currently under study by the author of this paper much more in depth in the context of segmented cointegration and related inference methods.

NOTES

1. In this sense, the concept of segmented cointegration has much to do with structural change that occurs over several periods of time. Therefore, the approach of this paper contrasts with most

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

of the existing works in the literature of structural change, which are about structural *break(s)* that all occur at single period(s) of time. (See Hansen and Johansen, 1993, and references therein.) A segmented cointegration, however, can result from situations other than structural change.

2. The specification (3.2) for w_t is justified by the following reasons: first, the data in $\mathcal{N}_T$ are not informative at all for estimating the coefficient β because in $\mathcal{N}_T$ there does not exist a β that yields stationary ε_t. Second, in the case of variance heterogeneity the weight w_t is the reciprocal of the "variance" of ε_t. The reciprocal of the "variance" is arbitrarily small for $t \in \mathcal{N}_T$ because the sample "variance" of ε_t in $\mathcal{N}_T$ is of $O_p(T)$, $T_{\mathcal{N}}^{-1} \sum_{t \in \mathcal{N}_T} (x_{1t} - \hat{\beta}' x_{2t})^2 = O_p(T)$ as the relation is a spurious regression in $\mathcal{N}_T$, which is by virtue of Phillips (1986).

3. The initial condition e_{t_0-j} for $j = 1, \ldots, p$ for the starting period t_0 in $\mathcal{C}_T$ can be taken either from the last p observations of the previous noncointegration regime or from the first p observations of $\mathcal{C}_T$ depending on which of the two hypotheses, H_0 or H_1, is favored. To be in favor of H_0 it can be taken from the last p observations of the previous noncointegration regime: $e_{t_0-j} = x_{1t_0-j} - \hat{\beta}(\mathcal{C}_T)' x_{2t_0-j}$ for $j = 1, \ldots, p$. To be in favor of H_1 we can take the first p observations of $\mathcal{C}_T$ for the initial condition.

4. This is again to reduce the computational requirements. In practice the length of the noncointegration period $\ell(\mathcal{T}_N)$ is variable, and we need critical values for each $\ell(\mathcal{T}_N)$, which are almost impossible to obtain. Fortunately, however, for a conservative test we can get an upper limit $\bar{\ell}(\mathcal{T}_N)$ to get the critical values (Lemma 1 in Sect. 3).

5. In general, as the assumed length of the noncointegration period is smaller the distribution of the test statistic is more concentrated, which makes the test more conservative. This is true with respect to the absolute length of the noncointegration period as well as the relative length of the noncointegration period $\ell(\mathcal{T}_N)$. Thus, for a small sample analysis the test based on asymptotic critical values would lead to conservative inference.

REFERENCES

Bai, J. (1997) Estimating multiple breaks one at a time. *Econometric Theory* 13, 315–352.

Bai, J., R. Lumsdaine, & J. Stock (1998) Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65, 395–432.

Balke, N.S. & T.B. Fomby (1996) Threshold cointegration. *International Economic Review* 38, 627–645.

Banerjee, A., J.K. Dolado, J.W. Galbraith, & D.F. Hendry (1993) *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Chan, N.H. & C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 15, 1050–1063.

Chong, T.L. (2001) Structural change in AR(1) models. *Econometric Theory* 17, 87–155.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

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

Gregory, A.W. & B.E. Hansen (1996) Residual based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70, 99–126.

Hansen, H. & S. Johansen (1993) Recursive Estimation in Cointegrated VAR Models. Manuscript, University of Copenhagen.

Hendry, D.G. (1995) *Dynamic Econometrics*. Oxford: Oxford University Press.

Kim, J.Y. (2000) Detection of change in persistence of a linear time series. *Journal of Econometrics* 95, 97–116.

Kim, J.Y. (2001) Inference on Segmented Cointegration and Reconsideration of the Fisher Relation. Discussion paper, Department of Economics, State University of New York at Albany.

Leybourne, S.J., B.P.M. McCabe, & A.R. Tremayne (1996) Can economic time series be differenced to stationarity? *Journal of Business and Economic Statistics* 14, 435–446.

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

MacKinnon, J.G. (1991) Critical values for cointegration tests. In R.F. Engle and C.W.J. Granger (eds.), *Long-Run Economic Relationships: Readings in Cointegration*, pp. 267–276. Oxford: Oxford University Press.
McCabe, B.P.M. & A.R. Tremayne (1995) Testing a time series for difference stationarity. *Annals of Statistics* 23, 1015–1028.
Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.
Phillips, P.C.B. (1987) Time series regression with unit roots. *Econometrica* 55, 277–302.
Phillips, P.C.B. (1998) Econometric Analysis of Fisher's Equation. Cowles Foundation Discussion Paper 1180.
Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.
Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
Siklos, P.L. & C.W.J. Granger (1997) Temporary cointegration with an application to interest rate parity. *Macroeconomic Dynamics* 1, 640–657.
Stock, J.H. & M.W. Watson (1994) Evidence on Structural Instability in Macroeconomic Time Series Relations, NBER Technical working paper 164.

APPENDIX: MATHEMATICAL PROOFS

Proof of Theorem 1. Except for the fact that we have segmentation with $\mathcal{N}_T$ and $\mathcal{T}_C$ the asymptotics in (3.8) and (3.9) are similar to the corresponding results of Phillips and Ouliaris (1990).

Proof of Theorem 2. The result is an immediate consequence of the continuous mapping theorem of Chan and Wei (1989) applied to the statistics $Z(\mathcal{C}_T)$ and $ADF(\mathcal{C}_T)$. ■

Proof of Lemma 1. Let $\ell(\mathcal{T}_\mathcal{N})$ be the volume of the set (i.e., the Lebesque measure) $\mathcal{T}_\mathcal{N}$. Then, it is true that $\{\mathcal{T}_\mathcal{N}: \ell(\mathcal{T}_\mathcal{N}) \leq \bar{\ell}_1\} \subset \{\mathcal{T}_\mathcal{N}: \ell(\mathcal{T}_\mathcal{N}) \leq \bar{\ell}_2\}$ for $\bar{\ell}_1 < \bar{\ell}_2$. Therefore,

$$\min(\{\mathcal{T}_\mathcal{N}: \ell(\mathcal{T}_\mathcal{N}) \leq \bar{\ell}_2\}) < \min(\{\mathcal{T}_\mathcal{N}: \ell(\mathcal{T}_\mathcal{N}) \leq \bar{\ell}_1\})$$

for every sample realization. Then, the conclusion of Lemma 1 follows.

Proof of Lemma 2. Except that they are obtained based on a regression for samples with segmentation the statistics $Z_\rho(\mathcal{C}_T)$, $Z_t(\mathcal{C}_T)$, and $ADF(\mathcal{C}_T)$ under H_1 have the same properties as the corresponding statistics in Phillips (1997), Phillips and Perron (1988), and Phillips and Ouliaris (1990) under H_1 of ε_t, $t = 1, \ldots, T$, being $I(0)$. This is particularly so for the magnitude of the statistics. Then, under the given conditions the conclusion of Lemma 3 immediately follows from Theorem 5.1 of Phillips and Ouliaris (1990). ■

Proof of Theorem 3. (a) Since the same analysis is applicable for all (2.1a)–(2.1c), we consider the regression (2.1a). Let $U(\tau^0)$ be an open neighborhood of τ^0, the true value of τ, and $\mathcal{U}$ be a local base of τ^0 containing all such neighborhoods. To prove that $(\hat{\tau}_i - \tau_i^0) = o_p(1)$ it is sufficient to show that for each $U(\tau^0) \in \mathcal{U}$

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

$$\liminf_{T\nearrow\infty} P\left[\sup_{\tau\in\mathcal{T}\setminus U(\tau^0)} \Lambda_T(\tau) - \Lambda_T(\tau^0) < \eta\right] = 1 \tag{A.1}$$

for any $\eta > 0$. Notice that $P[\Lambda_T(\tau^0) > 0] = 1$ and $\Lambda_T(\tau^0) = O_p(1)$ since the numerator and the denominator of $\Lambda(\tau^0)$ are of $O_p(1)$ with the probability limit of each being nonzero.

Let $\delta > 0$ be an arbitrary number. To prove (A.1) it is sufficient to show that for $\tau_i = \tau_i^0 + \epsilon$ or $\tau_i = \tau_i^0 - \epsilon$, $i = 0,1$, for $0 < \epsilon < \delta$

$$\liminf_{T\nearrow\infty} P[\Lambda_T(\tau) - \Lambda_T(\tau^0) < \eta] = 1 \tag{A.2}$$

uniformly in ϵ. Since the analysis is the same for $i = 0$ and for $i = 1$, we consider only the case of $i = 0$. For simplicity, fix $\tau_1 = \tau_1^0$. Consider first the case of $\tau_0 = \tau_0^0 + \epsilon$. In this case (A.2) follows because the numerator of $\Lambda_T(\tau)$ is of $O_p(1)$ whereas the denominator is of $O_p(\kappa T)$ for a real number κ, $0 < \kappa < \lim(T_C/T)$, which is the result of the following: data in a portion of the noncointegration period, (τ_0^0, τ_0), are used for estimating $\hat{\beta}$. Therefore, the estimator $\hat{\beta}$ is not consistent, and thus the residual process e_t is nonstationary for the whole data period in this case.

Now, consider the case $\tau_0 = \tau_0^0 - \epsilon$ for $\epsilon \in (0,\delta)$. Notice that in this case $\hat{\beta}$ is consistently estimated. Hence, it follows that

$$\Lambda_T(\tau) = \frac{((\tau_1^0 - \tau_0)T)^{-2}\left\{\sum_{\tau_0 T}^{\tau_0^0 T} e_t(\tau)^2 + \sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2\right\}}{T_C^{-1}\sum_{t\in C_T} e_t(\tau)^2}$$

$$= \frac{((\tau_1^0 - \tau_0)T)^{-2}\{O_p((\tau_0^0 - \tau_0)T) + O_p(((\tau_1^0 - \tau_0^0)T)^2)\}}{O_p(1)} \tag{A.3}$$

uniformly in ϵ where $\sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 = O_p(((\tau_1^0 - \tau_0^0)T)^2)$ under the assumption that the noncointegration period is $\{\tau_0^0 T + 1, \ldots, \tau_1^0 T\}$. Also notice that the dominant term in (A.3) is

$$((\tau_1^0 - \tau_0)T)^{-2}\sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 = ((\tau_1^0 - \tau_0)T)^{-2} O_p(((\tau_1^0 - \tau_0^0)T)^2).$$

But this dominant term is less than the numerator of $\Lambda_T(\tau^0)$:

$$((\tau_1^0 - \tau_0)T)^{-2}\sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 < ((\tau_1^0 - \tau_0^0)T)^{-2}\sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 \tag{A.4}$$

since $\tau_0 = \tau_0^0 - \epsilon$ for $\epsilon > 0$. Then, it follows that $\Lambda_T(\tau) < \Lambda_T(\tau^0)$ for $\tau_0 = \tau_0^0 - \epsilon$ uniformly in $\epsilon \in (0,\delta)$. Notice that under Assumption 2 the denominator of (A.3) and that of $\Lambda_T(\tau_0)$ are asymptotically equivalent. This completes the proof. ■

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press

(b) Define $U_T(\tau^0, C) = \{\tau : |T(\tau - \tau^0)| < C\}$ for $C > 0$. It is sufficient to show that for all $T > T_0$ for a sufficiently large T_0 there exists a C such that $C > 0$ and

$$P\left[\sup_{\mathcal{T}\setminus U_T(\tau^0, C)} \Lambda_T(\tau) - \Lambda_T(\tau^0) \geq 0\right] < \eta \quad \text{for each } \eta > 0. \tag{A.5}$$

To prove (A.5) it is sufficient to show that there exists a finite real number $C > 0$ such that, for $\tau = (\tau_0, \tau_1)$ with $\tau_i T = \tau_i^0 T + M_i$ or $\tau_i T = \tau_i^0 T - M_i$ for $M_i > C$ for $i = 0,1$,

$$P[\Lambda_T(\tau) - \Lambda_T(\tau_0) \geq 0] < \eta \tag{A.6}$$

uniformly in $M = (M_0, M_1)$ for all $T > T_0$ and for each $\eta > 0$. Since the analysis is the same for $i = 0$ and for $i = 1$, we consider only the case of $i = 0$. For simplicity, fix $\tau_1 = \tau_1^0$. Consider first the case of $\tau_0 T = \tau_0^0 T + M_0$. In this case (A.6) follows because for a suitable choice of C the numerator of $\Lambda_T(\tau)$ is of $O_p(1)$ whereas the denominator is of $O_p(\kappa T)$ for a real number κ, $0 < \kappa < \lim(T_C/T)$, which results from the following: data in a portion of the noncointegration period, $(\tau_0^0 T, \tau_0 T)$, are used for estimating $\hat{\beta}$. Therefore, for a suitable choice of C the estimator $\hat{\beta}$ cannot be consistent, and thus the residual process e_t is nonstationary for the whole data period in this case.

Now, consider the case $\tau_0 T = \tau_0^0 T - M_0$ for $M_0 > C$. Notice that in this case $\hat{\beta}$ is consistently estimated. Then, it follows that

$$\begin{aligned}\Lambda_T(\tau) &= \frac{((\tau_1^0 - \tau_0)T)^{-2}\left\{\sum_{\tau_0 T}^{\tau_0^0 T} e_t(\tau)^2 + \sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2\right\}}{T_C^{-1}\sum_{t\in C_T} e_t(\tau)^2} \\ &= \frac{((\tau_1^0 - \tau_0)T)^{-2}\{O_p((\tau_0^0 - \tau_0)T) + O_p(((\tau_1^0 - \tau_0^0)T)^2)\}}{O_p(1)}\end{aligned} \tag{A.7}$$

uniformly in M_0 where $\sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 = O_p(((\tau_1^0 - \tau_0^0)T)^2)$ under the assumption that the noncointegration period is $\{\tau_0^0 T + 1, \ldots, \tau_1^0 T\}$. Also notice that the dominant term in (A.7) is

$$((\tau_1^0 - \tau_0)T)^{-2} \sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 = ((\tau_1^0 - \tau_0)T)^{-2} O_p(((\tau_1^0 - \tau_0^0)T)^2).$$

But this dominant term is less than the numerator of $\Lambda_T(\tau^0)$:

$$((\tau_1^0 - \tau_0)T)^{-2} \sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 < ((\tau_1^0 - \tau_0^0)T)^{-2} \sum_{\tau_0^0 T+1}^{\tau_1^0 T} e_t(\tau)^2 \tag{A.8}$$

since $\tau_0 T = \tau_0^0 T - M_0$ for $M_0 > C$. Then, it follows that $\Lambda_T(\tau) < \Lambda_T(\tau^0)$ for $\tau_0 T = \tau_0^0 T - M_0$ uniformly in $M_0 > C$. This completes the proof. ■

https://doi.org/10.1017/S0266466603194078 Published online by Cambridge University Press