A CUSUM Test for Cointegration Using Regression Residuals

By

Zhijie Xiao and Peter C.B. Phillips

September 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1329

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

A CUSUM TEST FOR COINTEGRATION USING REGRESSION RESIDUALS *

Zhijie Xiao
Department of Economics
University of Illinois at Urbana-Champaign

and

Peter C. B. Phillips
Cowles Foundation, Yale University,
University of Auckland & University of York

March 16, 2001

Abstract

We show that the conventional CUSUM test for structural change can be applied to cointegrating regression residuals leading to a consistent residual based test for the null hypothesis of cointegration. The proposed tests are semiparametric and utilize fully modified residuals to correct for endogeneity and serial correlation and to scale out nuisance parameters. The limit distribution of the test is derived under both the null and the alternative hypothesis. The tests are easy to use and are found to perform quite well in a Monte Carlo experiment.

JEL Classification: C22

Key Words and Phrases: Bandwidth, CUSUM test, Fully modified regression, Null of cointegration, Residual based test, Semiparametric method.

*We thank Peter Robinson, John Chao, Roger Koenker, Anil Bera, Ted Juhl, an Associate Editor, two referees, and participants of seminars in University of Maryland and University of Illinois for helpful comments. Phillips thanks the NSF for support under Grant No. SBR 97-30295. Xiao thanks the Hong Kong Research Grants Council for support under Grant No. CUHK 4078/98.

1 Introduction

In the last ten years, a substantial body of econometric theory has developed for testing the presence of cointegration in time series models (see Phillips, 1991a; Stock and Watson,1988; Phillips and Ouliaris, 1990; Hansen, 1992a; Johansen, 1991, 1995; and Park, Ouliaris and Choi, 1988, among others). In their original study, Engle and Granger (1987) suggest testing cointegration by examining whether or not the residuals from the cointegrating regression contain a unit root, and Phillips and Ouliaris (1990) study the asymptotic properties of these so-called residual-based tests. Stock and Watson (1988) propose the 'common trends' approach based on the fact that a vector time series cointegrated with order r can be written as the sum of $n - r$ common trends and an $I(0)$ component. Using the reduced rank regression technique, Johansen (1991, 1995) studied likelihood inference based on a Gaussian error correction model and showed that the asymptotic distribution of the likelihood ratio tests for cointegration is determined by a generalized eigenvalue problem and has the form of a multivariate unit root distribution. More recently, Phillips (1996) has proposed treating cointegrating rank as a matter of order selection and gives methods for the joint determination of cointegrating rank and lag order selection in vector autoregressions that have been shown to be consistent in Chao and Phillips (1999).

Among these various tests for cointegration, the residual-based procedure has been one of the more frequently used approaches in empirical research. The residual-based procedure was analyzed and critical values were reported in Phillips and Ouliaris (1990) and later by MacKinnon (1991). These tests take the residuals calculated from the cointegrating regression (conventionally, a simple OLS regression among the levels of economic time series) and apply unit root tests to the residuals. If there is no cointegration among the individual time series, the residual process should contain a unit root. Otherwise, if there is cointegration, the residuals will be stationary. Thus, unit root tests can be applied to the residual process and the null hypothesis that there is a unit root in the residual process corresponds to the null hypothesis of no cointegration in the vector time series. These procedures are used in the same way as unit root tests, but the data are the residuals from the cointegrating regression, and the alternative hypothesis of cointegration is now the main hypothesis of interest.

Being unit root tests, these procedures are designed to test the null hypothesis of no cointegration. Since cointegration is the primary interest, it is natural to consider residual-based procedures that seek to test a null hypothesis of cointegration. Shin (1994) used a component representation of the time series and proposed a residual-based test for the null hypothesis of cointegration based on the KPSS (Kwiatkowski et al. 1991) test for stationarity. Related, but less popular, methods have been considered by Park (1990), and Park, Ouliaris, and Choi (1988), among others.

As this paper shows, the null hypothesis of cointegration can be tested by directly looking at the fluctuation in the residual process of a cointegrating regression. In particular, we show that the conventional CUSUM (or MOSUM) test for structural stability can also be applied to the residuals in a cointegrating regression and provide another way of testing the null hypothesis of cointegration. The CUSUM test was introduced by Brown, Durbin and Evans (1975) for the study of structural change and the original test statistic was constructed based on cumulated sums of recursive residuals. Ploberger and Kramer (1992) extended the CUSUM test to OLS residuals. Nowadays, these tests are widely used in econometrics and statistics, and have become especially popular because they draw attention to structural change and breakpoints in the data. This paper follows up previous work on CUSUM tests and considers its extension to the residuals in cointegrating regressions. To obtain a valid cointegrating CUSUM test, we modify OLS regression by means of semiparametric corrections for serial correlation and endogeneity and construct fully modified (Phillips and Hansen 1990) residuals so that the limit process of the corresponding partial sums can be expressed as a variance parameter multiplied by a functional of Brownian motion that is free of nuisance parameters. Under the null hypothesis of cointegration, fluctuations in the residual process u_t are simply "equilibrium errors" and thus the cumulated sums of the cointegrating residuals are of order $n^{1/2}$. If the residuals display too much fluctuation, we should reject the null hypothesis of cointegration.

Closely related work to the work of the present paper is Hao and Inder (1996)[1] who considered testing structural change in cointegrated regressions (i.e. $\beta =$ constant in regression (4) in Section 2) based on the CUSUM statistic. While these authors focus on the constancy of the regression parameter β, we consider a residual-based test for the null hypothesis of cointegration against the alternative of no cointegration. The two models have the same behavior under the null hypotheses but are different under the alternatives. Readers are referred to Hao and Inder (1996) or an earlier version of the current paper for a detailed asymptotic analysis on the null model.

The paper is organized as follows. Section 2 develops the fully modified CUSUM test for cointegration. Consistency is studied in Section 3. Section 4 reports finite sample size and power based on a Monte Carlo experiment and Section 5 concludes. Proofs are given in the Appendix.

In matters of notation, we use "$\Rightarrow$" to signify weak convergence of the associated probability measures, $[nr]$ to signify the integer part of nr, $:=$ to signify definitional equality, and $I(k)$ to denote integration of order k. Continuous stochastic processes

[1]We thank the referees for bringing to our attention this related work.

such as the Brownian motion $B(r)$ on $[0,1]$ are usually written simply as B and integrals $\int$ are understood to be taken over the interval $[0,1]$, unless otherwise specified.

2 A CUSUM Test for the Null of Cointegration

Consider an m-vector time series $\{z_t\}$ generated by

$$z_t = \Phi d_t + z_t^s, \tag{1}$$

where d_t is the deterministic trend component with unknown coefficient matrix Φ and z_t^s is the stochastic component. The leading cases of interest for d_t are: (1) no trend $d_t = 0$; (2) a constant term $d_t = 1$; and (3) a linear time trend $d_t = (1,t)'$. The stochastic component is an integrated process

$$z_t^s = z_{t-1}^s + \xi_t, \ t = 1, ..., n, \tag{2}$$

where the initial observation of z_t^s is taken to be $z_0^s = 0$, and the process $\{\xi_t\}$ is strictly stationary and ergodic with zero mean, finite variance, and spectral density matrix $f_{\xi\xi}(\lambda)$. For convenience in deriving asymptotic theory, we assume that the random sequence $\{\xi_t\}$ follows a general linear process (Phillips and Solo 1992) whose coefficients satisfy the summability conditions given in the following Assumption.

Assumption L (Linear Process): *$\xi_t = C(L)\varepsilon_t$, where ε_t is a white noise process with zero mean and variance matrix $\Sigma_\varepsilon > 0$, and $C(L) = \sum_{j=0}^{\infty} C_j L^j$, $|C(1)| \neq 0$, and $\sum_{j=1}^{\infty} j^2 |C_j| < \infty$.*

This assumption ensures that the partial sum process constructed from ξ_t satisfies a multivariate invariance principle: $n^{-1/2}\sum_{t=1}^{[nr]} \xi_t \Rightarrow B_z(r), 0 \leq r \leq 1$, where $B_z(r)$ is a Brownian motion with long-run variance $\Omega = C(1)\Sigma_\varepsilon C(1)'$. We partition $z_t^s = (y_t^s, x_t^{s\prime})'$ into the scalar variate y_t^s and the p-vector x_t^s $(m = p+1)$ with the following conformable partition of z_t, ξ_t, $B_z(r)$, and Ω :

$$z_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} \begin{matrix} 1 \\ p \end{matrix}, \ \xi_t = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} \begin{matrix} 1 \\ p \end{matrix}, \ B_z(r) = \begin{bmatrix} B_y(r) \\ B_x(r) \end{bmatrix} \begin{matrix} 1 \\ p \end{matrix}, \ \Omega = \begin{bmatrix} \omega_{yy} & \omega_{xy}' \\ \omega_{xy} & \Omega_{xx} \end{bmatrix}.$$

Following Phillips and Ouliaris (1990) and Hansen (1992a), we assume $\Omega_{xx} > 0$ so that cointegration among the elements of x_t^s is excluded. If a p-vector β exists such that

$$u_t = y_t^s - \beta' x_t^s \tag{3}$$

is stationary with continuous spectral density $f_{uu}(\lambda)$, then y_t^s and x_t^s are cointegrated in the sense of Engle and Granger (1987). If the process u_t satisfies Assumption L, then a functional limit theorem holds for the partial sums constructed from u_t, viz. $n^{-1/2}\sum_{t=1}^{[nr]} u_t \Rightarrow B_u(r) = BM(\omega_u^2), 0 \leq r \leq 1$, where $\omega_u^2 = 2\pi f_{uu}(0)$ is the long-run variance of the process u_t.

As discussed in Phillips (1986) and Phillips and Ouliaris (1990), the hypothesis of interest can be formulated in terms of the variance parameters. Notice that $|\Omega| = \left(\omega_{yy} - \omega_{xy}'\Omega_{xx}^{-1}\omega_{xy}\right)|\Omega_{xx}|$, so that if $\omega_{y.x} = \omega_{yy} - \omega_{xy}'\Omega_{xx}^{-1}\omega_{xy} = 0$, Ω is singular, and y_t^s and x_t^s are cointegrated (Phillips, 1986). Thus, in terms of the conditional variance parameter $\omega_{y.x}$, the null hypothesis of cointegration corresponds to $H_0 : \omega_{y.x} = 0$, and the alternative hypothesis of no cointegration is $H_1 : \omega_{y.x} > 0$. Tests based on this formulation in terms of the variance parameter were originally developed in Phillips and Ouliaris (1988) and involve a null in which the parameter $\omega_{y.x}$ is on the boundary. Their approach uses principal components and leads to a confidence limit rather than a formal statistical test of the null. For that reason, residual based cointegration tests (where H_1 is the null) are considered to be more practical.

Combining models (1) and (3), we have the following representation

$$y_t = \gamma' d_t + \beta' x_t + u_t = \Pi w_t + u_t, \tag{4}$$

where

$$\Pi = \begin{bmatrix} \gamma', & \beta' \end{bmatrix}, \text{ and } w_t = \begin{bmatrix} d_t \\ x_t \end{bmatrix}.$$

The residual-based test (Engle and Granger, 1987; Phillips and Ouliaris, 1990; Hansen 1992a) considers a null hypothesis of no cointegration and applies scalar unit root tests to the residuals of the cointegrating regression (4):

$$\widehat{u}_t = y_t - \widehat{\gamma}' d_t - \widehat{\beta}' x_t = y_t - \widehat{\Pi} w_t. \tag{5}$$

In this paper, we propose testing the null hypothesis of cointegration (H_0) directly by looking at the fluctuation in the residual process $\widehat{u}_t$. The intuition of the suggested tests is as follows. If a long term equilibrium relationship exists between y_t and x_t, i.e., y_t and x_t are cointegrated, then the residual process in the cointegrating regression (5) should be stable and fluctuations in $\widehat{u}_t$ reflect only equilibrium errors. Otherwise, the fluctuations in $\widehat{u}_t$ can be expected to be of a larger order of magnitude. Thus, the null hypothesis of cointegration should be rejected whenever there is excessive fluctuation in the residuals of (5).

To measure the fluctuation in the residual process, we consider the following cumulated sum statistic (also see Remarks 1 and 2 below for related discussion)

$$\max_{k=1,\ldots,n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} \widehat{u}_t \right|. \tag{6}$$

We assume that there is a standardizing matrix D_n such that $D_n d_{[nr]} \to d(r)$ as $n \to \infty$. For the case of a linear trend, $D_n = diag[1, n^{-1}]$ and $d(r) = (1, r)'$. Under Assumption L and the null of cointegration, the least squares estimator of the cointegrating vector β, $\widehat{\beta}$, is n-consistent and the asymptotic properties of $\widehat{\beta}$ are well known (Park and Phillips 1988; Phillips and Hansen 1990). Indeed,

$$n(\widehat{\beta} - \beta) \Rightarrow \left[\int_0^1 \underline{B}_{xd} \underline{B}'_{xd}\right]^{-1} \left[\int_0^1 \underline{B}_{xd} dB_u + \Lambda_{xu}\right], \tag{7}$$

where $\Lambda_{xu} = \sum_{t=0}^{\infty} E(\xi_{2t} u_0)$ is a one-sided long-run covariance and $\underline{B}_{xd}(r) = B_x(r) - \left(\int B_x d'\right)\left(\int dd'\right)^{-1} d(r)$. As a result, the partial sum of the residuals $\widehat{u}_t$ satisfies the following invariance principle

$$n^{-1/2} \sum_{t=1}^{[nr]} \widehat{u}_t \Rightarrow f(r, B_u, \underline{B}_x, \Lambda_{ux}), \tag{8}$$

where $f(r, B_u, \underline{B}_x, \Lambda_{ux})$ is a function of $B_u(r)$, $\underline{B}_x(r)$, and Λ_{ux}. However, under the alternative hypothesis of no cointegration (H_1), the residual process of regression (5) is nonstationary and the cumulated sum process constructed from $\widehat{u}_t$ has a larger order of magnitude. It is easy to verify that in this case $\sum_{t=1}^{[nr]} \widehat{u}_t = O_p(n^{3/2})$. As a result, the statistic (6) diverges at rate n under H_1.

A valid test statistic should have a limit distribution that is free of nuisance parameters under the null hypothesis. Notice that the limiting processes $B_x(r)$ and $B_u(r)$ in (8) will be correlated Brownian motions whenever contemporaneous correlation between ξ_{2t} and u_t exists. Despite super-consistency, $\widehat{\beta}$ is second-order biased and the miscentering effect in the limit distribution (7) is reflected in Λ_{xu}. As a result, the statistic (6) can not be used directly for testing cointegration.

To eliminate the nuisance parameters from the limiting null distribution, we construct fully modified (FM) residuals. Fully modified least squares regression was originally proposed by Phillips and Hansen (1990) and further studied in Phillips (1995) with the intent of providing a regression based procedure for the optimal estimation of cointegrating regressions. To construct fully modified residuals, we consider the following kernel estimates of ω_u^2, Λ_{xu}, Ω_{xx}, $\Lambda_{xx} = \sum_{t=0}^{\infty} E(\xi_{2t}\xi'_{20})$, and $\Omega_{xu} = \sum_{t=-\infty}^{\infty} E(\xi_{2t} u_0)$ (see, e.g., Phillips, 1995):

$$\begin{aligned}
\widehat{\omega}_u^2 &= \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) C_{uu}(h), \ \widehat{\Lambda}_{xu} = \sum_{h=0}^{M} k\left(\frac{h}{M}\right) C_{xu}(h), \\
\widehat{\Lambda}_{xx} &= \sum_{h=0}^{M} k\left(\frac{h}{M}\right) C_{xx}(h), \ \widehat{\Omega}_{xu} = \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) C_{xu}(h), \\
\widehat{\Omega}_{xx} &= \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) C_{xx}(h),
\end{aligned}$$

where $k(\cdot)$ is the lag window defined on $[-1,1]$ with $k(0)=1$, and M is the bandwidth parameter satisfying the property that $M \to \infty$ and $M/n \to 0$ as the sample size $n \to \infty$. In later analysis, for notational convenience, we will also use the spectral window corresponding to $k(\cdot)$, viz., $K(\lambda) = \lim(2\pi M)^{-1} \sum_h k(h/M) e^{ih\lambda}$. The quantities $C_{uu}(h)$, $C_{xu}(h)$, and $C_{xx}(h)$ are sample covariances defined by $C_{uu}(h) = n^{-1} \sum' \widehat{u}_t \widehat{u}_{t+h}$, $C_{xu}(h) = n^{-1} \sum' \Delta x_t \widehat{u}_{t+h}$, $C_{xx}(h) = n^{-1} \sum' \Delta x_t \Delta x'_{t+h}$, where $\sum'$ signifies summation over $1 \le t, t+h \le n$. Thus, $C_{xx}(h) = n^{-1} \sum_{t=1}^{n-h} \Delta x_t \Delta x'_{t+h}$ for $h \ge 0$. Under the null hypothesis of cointegration, $\widehat{\omega}_u^2$, $\widehat{\Lambda}_{xu}$, $\widehat{\Lambda}_{xx}$, $\widehat{\Omega}_{xu}$, and $\widehat{\Omega}_{xx}$ are consistent estimates of ω_u^2, Λ_{xu}, Λ_{xx}, Ω_{xu}, and Ω_{xx}.

Following Phillips and Hansen (1990) and Phillips (1995), we define

$$y_t^+ = y_t - \Delta x_t' \widehat{\Omega}_{xx}^{-1} \widehat{\Omega}_{xu}, \text{ and } \widehat{\Lambda}_{xu}^+ = \widehat{\Lambda}_{xu} - \widehat{\Lambda}_{xx} \widehat{\Omega}_{xx}^{-1} \widehat{\Omega}_{xu}.$$

The fully modified estimator of Π is then defined by the following formula

$$\widehat{\Pi}^+ = \left(\sum_t y_t^+ w_t' - \left[\begin{array}{cc} 0, & n\widehat{\Lambda}_{xu}^+ \end{array} \right] \right) \left(\sum_t w_t w_t' \right)^{-1}$$

and has an asymptotically mixed normal distribution.

The fully modified residual process can be constructed based on y_t^+ and $\widehat{\Pi}^+$:

$$\widehat{u}_t^+ = y_t^+ - \widehat{\Pi}^+ w_t,$$

and

$$n^{-1/2} \sum_{t=1}^{[nr]} \widehat{u}_t^+ \Rightarrow \omega_{u.x} \left\{ W_1(r) - \left[\int_0^1 dW_1 S' \right] \left[\int_0^1 SS' \right]^{-1} \int_0^r S \right\} := \omega_{u.x} \underline{W}_d(r),$$

where $\omega_{u.x}^2 = \omega_u^2 - \Omega_{ux} \Omega_{xx}^{-1} \Omega_{xu}$, $S(r)' = (d(r)', W_2(r)')$, $W_1(r)$ and $W_2(r)$ are 1 and p-dimensional standard Brownian motions that are independent of each other.

Since $\widehat{\omega}_{u.x}^2 = \widehat{\omega}_u^2 - \widehat{\Omega}_{ux} \widehat{\Omega}_{xx}^{-1} \widehat{\Omega}_{xu}$ is a consistent nonparametric estimator of $\omega_{u.x}^2$, we obtain the following CUSUM test that is asymptotically free of nuisance parameters and can be used in testing the null hypothesis of cointegration:

$$CS_n = \max_{k=1,\ldots,n} \frac{1}{\widehat{\omega}_{u.x} \sqrt{n}} \left| \sum_{t=1}^k \widehat{u}_t^+ \right|. \tag{9}$$

We summarize the asymptotic distribution of CS_n in Theorem 1.

THEOREM 1: *Under the null hypothesis of cointegration, if u_t satisfies Assumption L, then, as $n \to \infty$,*

$$CS_n \Rightarrow \sup_{0 \le r \le 1} \left| \underline{W}_d(r) \right|, \tag{10}$$

where $\underline{W}_d(r) = W_1(r) - \left[\int_0^1 dW_1 S'\right]\left[\int_0^1 SS'\right]^{-1} \int_0^r S.$

Just like the limit distribution of the residual-based test for the null of no cointegration, the limiting variate (10) is dependent on the limiting function of the deterministic trend, as well as the known dimensional constant p. For the case with no trend, critical values have been computed based on a direct simulation with a sample size of $n = 2000$ and $20,000$ replications. These critical values are reported in Table 1. For cases where d_t equals a constant and a linear trend, Tables of critical values can be found in Hao and Inder (1996) and an early version of the current paper.

Table 1: Upper Tail Critical Values for CS_n: Case with No Deterministic Trend

Critical level	15%	10%	7.5%	5%	2.5%	1%
$p = 1$	1.480	1.616	1.714	1.842	2.063	2.326
$p = 2$	1.285	1.411	1.486	1.601	1.782	2.043
$p = 3$	1.148	1.242	1.325	1.414	1.547	1.761
$p = 4$	1.034	1.128	1.190	1.277	1.445	1.632

Remark 1. In principle, any other fluctuation test statistics can be applied in the same way. Generally speaking, if we consider a continuous functional $h(\cdot)$ that measures the fluctuation of in the residual process and denote $Y_n(r) = \frac{1}{\widehat{\omega}_{u.x}\sqrt{n}}\left|\sum_{t=1}^{[nr]} \widehat{u}_t^+\right|$, $h(Y_n(r))$ may be used as a test statistic. For example, the MOSUM test with cointegrating regression residuals can be constructed as follows

$$MS_n = \max_{k=1,\ldots,n-[nh]} \frac{1}{\widehat{\omega}_{u.x}\sqrt{n}} \left| \sum_{t=k}^{k+[nh]} \widehat{u}_t^+ \right|, \tag{11}$$

where $0 < h < 1$ is a bandwidth parameter for the moving window, indicating the proportion of $\widehat{u}_t^+$ used to construct the moving sum. Differing from the CUSUM test, where more and more residual terms enter the cumulated sum, each moving sum $(\sum_{t=k+1}^{k+[nh]} \widehat{u}_t^+)$ in the MOSUM test contains a fixed number of $\widehat{u}_t^+$. From the construction of the MOSUM test, it is apparent that the critical values of the detrended MOSUM test will depend on the width of the moving window. It can be shown that under the assumptions given in Theorem 1, the MOSUM test with cointegrating regression residuals converges weakly to $\sup_{0\le r\le 1-h} \left|\underline{W}_d(r+h) - \underline{W}_d(r)\right|$. Since the use of other types of fluctuation test statistics does not affect our analysis in any substantive way, we will focus our discussion on the CUSUM test to keep the discussion simple.

Remark 2. The limiting distribution of $\sup_{0\le r\le 1} \left|\underline{W}_d(r)\right|$ may be treated as an generalized version of the classical Kolmogoroff-Smirnoff type distribution. Notice that by using the Cramer-von Mises type measure for the fluctuation in residuals, we can derive the Shin (1994) test. In this sense, both the CUSUM test and the Shin test can be obtained by testing the fluctuations in the residuals. Our procedure corresponds to the Kolmogoroff-Smirnoff test and the Shin test is of Cramer-von Mises type.

3 Test Consistency

This section considers the asymptotic behavior of the cointegrating CUSUM test under the alternative hypothesis of no cointegration. It is important that a statistical test be able to fully discriminate between the null and the alternative in large samples. This is a nontrivial matter for the proposed test because both the numerator and the denominator of the test statistic diverge under the alternative.

Under the alternative hypothesis, there is no cointegration between y_t and x_t and (4) is commonly regarded as a spurious regression, whose asymptotic behavior was studied by Phillips (1986). As a result, the partial sum of $\widehat{u}_t^+$ has a different order of magnitude. In particular, it can be shown that $\max_{k=1,\ldots,n} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} \widehat{u}_t^+ \right|$ diverges as $n \to \infty$. However, notice that under H_1, the nonparametric kernel estimators $\widehat{\omega}_u^2$, $\widehat{\Delta}_{xu}$, and $\widehat{\Omega}_{xu}$, which are constructed from the residuals of (5), diverge as well. In order to prove consistency, we need to show that the denominator diverges at a slower rate.

Under H_1 and Assumption L, the least squares estimator of β has a non-degenerate limiting distribution as $n \to \infty$, viz.

$$\widehat{\beta} \Rightarrow \left(\int \underline{B}_{xd} \underline{B}_{xd}' \right)^{-1} \left(\int \underline{B}_{xd} \underline{B}_{yd} \right) = \zeta_d,$$

where $\underline{B}_{yd}$, defined similarly to $\underline{B}_{xd}$, is the Hilbert projection of $B_y(r)$ onto the space orthogonal to $d(r)$. The least-squares residual $\widehat{u}_t$ and its partial sum process have the following asymptotic behavior:

$$\begin{aligned} n^{-1/2} \widehat{u}_{[nr]} &\Rightarrow \underline{B}_{yd}(r) - \left(\int \underline{B}_{yd} \underline{B}_{xd}' \right) \left(\int \underline{B}_{xd} \underline{B}_{xd}' \right)^{-1} \underline{B}_{xd}(r) := \underline{Q}(r), \\ n^{-3/2} \sum_{t=1}^{[nr]} \widehat{u}_t &\Rightarrow \int_0^r \underline{Q}(s) := \overline{Q}(r). \end{aligned}$$

The nonparametric kernel estimates play an important role in the proposed procedure and affect the order of magnitude of the test statistic. The following Lemma, which is based on Phillips (1991b), gives some useful limit results for the nonparametric estimates under the alternative hypothesis.

Lemma 1: *Under H_1 and Assumption L, as $n \to \infty$, $M \to \infty$,* and $M/n \to 0$,

$$\begin{aligned}
\frac{1}{M}\widehat{\Omega}_{xu} &\Rightarrow 2\pi K(0)\left(\int_0^1 dB_x \underline{Q}\right) + \Omega_*, \\
\frac{1}{M}\widehat{\Lambda}_{xu} &\Rightarrow 2\pi K_1(0)\left(\int_0^1 dB_x \underline{Q}\right) + \Lambda_*, \\
\frac{1}{nM}\widehat{\omega}_{u.x}^2 &\Rightarrow 2\pi K(0)\int_0^1 \underline{Q}^2,
\end{aligned}$$

where $\Omega_* = \Omega_{xz}\eta$, $\Lambda_* = \Lambda_{xz}\eta$, $\eta' = (1, -\zeta_d')$, $\Omega_{xz} = [\omega_{xy} : \Omega_{xx}]$, $\Lambda_{xz} = \sum_{t=0}^{\infty} E(\Delta x_0 \Delta z_t') = [\Lambda_{xy} : \Lambda_{xx}]$, $K(0) = \frac{1}{2\pi}\int k(x)dx$ *is the spectral window evaluated at 0, and* $K_1(0) = \lim_{M\to\infty} \frac{1}{2\pi M}\sum_{h=0}^{M} k(h/M)$ *is the one-sided counterpart of* $K(0)$.

Using the result of Lemma 1, we obtain the following Theorem on the consistency of the test CS_n.

Theorem 2: *Under H_1 and Assumption L, as $n \to \infty$,* $\Pr[CS_n > B_n] \to 1$ *for any nonstochastic sequence* $B_n = o(n^{1/2}M^{-1/2})$.

From theorem 2, it is apparent that the behavior of CS_n under the alternative hypothesis is similar to that of the KPSS (Kwiatkowski et al., 1992) test in that the divergence rate of CS_n under H_1 is dependent on the bandwidth expansion rate.

Remark 3. If we compare the proposed test with a CUSUM test for a structural break (say, Hao and Inder (1996), Ploberger and Kramer (1992)), the statistical properties of the residual process u_t in these models do not change between the null and the alternative hypothesis, and thus the order of magnitude of the scalar (the variance estimator $\widehat{\omega}_{u.x}$) also does not change under the alternative hypothesis. However, in the present paper, the behavior of u_t changes fundamentally between the null and the alternative, and both the cumulated sum (in the numerator) and the variance estimator (in the denominator) diverge under the alternative hypothesis, complicating the consistency issue.

Remark 4. In stationary time series regression (i.e. with I(0) regressors, see Ploberger and Kramer, 1992), the limiting distribution of the CUSUM statistic is asymptotically invariant to the limit behavior of the regressors (only the constant term plays a role in the limit). However, in the cointegrating CUSUM test (9), because of the larger signal contained in the nonstationary regressors, the regressors do influence the limiting distribution, and critical values of the test are dependent on the dimension of x_t, reproducing the well known phenomenon that arises in conventional residual based tests (Phillips and Ouliaris, 1990).

REMARK 5. The nonparametric variance estimates that are used in the fully modified regression entail a choice of bandwidth M. Just like other statistical procedures that use nonparametric estimators, the finite sample performance of the proposed test depends on the choice of bandwidth. The empirical size and power of the test can vary considerably with bandwidth selection although, as long as the bandwidth M satisfies a certain expansion rate, the semiparametric tests are asymptotically equivalent.

In the literature of covariance (or, more generally, spectral density) estimation, several automatic bandwidth methods have been developed by Sheather (1986), Robinson (1988), and Andrews (1991) among others. Since the focus of those papers was density estimation itself, not hypothesis tests that may rely on such estimators, the criterion functions are different and it can be anticipated that optimal bandwidth choices will change with the criterion functions, depending on the focus of interest in the study. More importantly, traditional bandwidth selection methods in spectral density estimation were derived for stationary time series with weak dependence. However, the stationarity properties of the time series in the current paper change between H_0 and H_1. Thus, taking into account the trade-off between correct size and reasonable power, the optimal bandwidth choices obtained in past studies may not be appropriate in our test. Indeed, if we use the traditional data dependent bandwidth choices in cointegration tests, the sampling performance of the tests may be good under the null hypothesis, but will be poor under the alternative. For example, Andrews (1991) studied bandwidth selection in nonparametric variance estimation based on minimizing the mean squared error of the kernel variance estimator and found that the optimal bandwidth is given by a rule of the form $c(f,k)n^{1/3}$ for the Bartlett estimator, where $c(f,k)$ is a function of the unknown spectral density. A data dependent bandwidth choice can then be proposed by estimating $c(f,k)$ using a plug-in method. If we use this estimator in the cointegration tests, the order of magnitude of $\widehat{c(f,k)}$ may change between H_0 and H_1. Actually, $\widehat{c(f,k)}$ may diverge under the alternative, reducing the power of the test.

4 Finite Sample Performance

A Monte Carlo experiment was conducted to examine the finite sample performance of the CUSUM test for cointegration. From the construction of the test, it is apparent that its finite sample performance depends on the sample size n and the bandwidth parameter M used to calculate the long-run variance and covariance parameters. Thus, special attention is paid here to the effects of the bandwidth and sample size on the performance of these tests.

The data were generated from the following bivariate regression model

$$y_t = \beta x_t + u_t,\ \beta = 1, \tag{12}$$

where

$$u_t = \alpha u_{t-1} + \varepsilon_t$$

and

$$\Delta x_t = v_t, t = 1, ..., n.$$

The random vector (ε_t, v_t) is independently distributed as bivariate normal $N(0, \Sigma)$, and

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The initial values are all set to be zero. We examined the test without trend (i.e. $d_t = 0$) and the linear trend case (i.e. $d_t = (1, t)$). Thus, the corresponding 5% level critical values are 1.842 (no trend case) and 0.834 (linear trend case), respectively.

In this model, the AR coefficient α is a convenient nuisance parameter to investigate. For $|\alpha| < 1$, y_t and x_t are cointegrated, and when $\alpha = 1$, there is no cointegration and regression (12) is a spurious regression. As α approaches unity, there is more and more persistence in the residual process. In consequence, it is anticipated that the empirical rejection rate of these tests will increase as α increases, and the test will overreject the null hypothesis of cointegration for large α, depending on how close α is to unity. A wide range of α values has been considered in the simulation, including $\alpha = 0, 0.2.0.4, 0.6, 0.8, 0.9, 0.95$. We are especially interested in the case where α is close to unity. In this case, the system is nearly cointegrated and overrejection will happen more frequently. For this reason, three values ($\alpha = 0.8, 0.9, 0.95$) that are close to unity are studied. These values are also commonly used in the Monte Carlo study for unit root tests.

Notice that the bandwidth parameter M corresponds to the number of lags used to calculate the long-run variance parameters. Intuitively, for $\alpha > 0$, the larger the α value, the longer the lags we should need. Thus, for small α, we expect that a small bandwidth value should provide reasonably good finite sample performance. As α increases, we need a larger M to estimate the long-run variance parameters. This intuition is confirmed in the simulation results. In the experiment, we have also examined the performance of the test for different choices of the correlation coefficient ρ. Since the results are qualitatively similar, we report the results for the case where $\rho = 0.7$. The experiment considered the following sample sizes: $n = 100, 200, 300, 500$. These sample sizes were chosen because they represent the most relevant range of sample sizes in empirical work. All experiments use 5,000 replications. For the kernel function, the Bartlett window is used, following Kwiatkowski et al. (1992), so that nonnegativity of the variance estimator is guaranteed.

Since the bandwidth is an important parameter in the construction of the test statistic, we examined the test based on quite a few different choices of bandwidth: $M1 = 1$, $M2 = 3$, $M3 = [4(n/100)^{1/4}]$, $M4 = [0.5n^{1/3}]$, $M5 = [n^{1/3}]$, $M6 = [2n^{1/3}]$, $M7 =$ Andrews (1991) AR(1) plug-in estimator. The first two bandwidth values are fixed and relatively small. Thus, only when α is close to zero will the test using these bandwidth choices have reasonably good performance. The other five bandwidth choices are all increasing with the sample size. $M3$ is of order $n^{1/4}$ and has been used in Schwert (1989), Kwiatkowski et al. (1992), and other simulations. In the spectral density estimation of stationary time series, the optimal bandwidth is $O(n^{1/3})$ for the Bartlett kernel. For this reason, we considered several bandwidth choice that are of order $n^{1/3}$. In particular, $M4, M5$, and $M6$ are all of order $n^{1/3}$ but have different fixed scalars. $M7 = \widehat{\delta}(f)n^{1/3}$ is the data dependent bandwidth using the AR(1) plug-in.

We examined the size and power of the residual-based procedures. In particular, Table 2 reports the empirical size of tests corresponding to different serial correlation (α) and bandwidth (M) choices at the 5% level and for the sample size $n = 100$. Tables 3, 4, and 5 report the empirical size for cases of $n = 200, 300$, and 500. When the value of α is taken to be unity, the rejection rates provide the empirical power of the residual-based test and these are reported in Table 6 for different choices of the bandwidth parameter. Both the case without a trend and the linear trend case are reported. In particular, the top panels in these tables reports the results with no trend and the bottom panels gives results with a linear trend.

From Tables 2, 3, 4, and 5, we can see that the tests have reasonable size if the bandwidth is appropriately chosen.. The poor cases are those with large α and small M, and those with a small sample and large M. For cases with large α (≥ 0.6), the problem of overrejection is severe when $M = 1$ or 3 because, according to the asymptotic theory, the validity of the tests requires M to increase with n in this case. Size distortion is especially large when $\alpha = 0.9$, or 0.95 because in these cases the residuals are nearly integrated. A very large bandwidth is needed to reduce the size distortion for cases with large α. We also see that as the sample size increases, for large enough bandwidth choices, the size property is reasonably good, corroborating the asymptotic theory.

Table 4: Empirical Size, $n = 100$

No Trend Case							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.025	0.020	0.017	0.022	0.017	0.009	0.023
$\alpha = 0.2$	0.041	0.022	0.019	0.031	0.019	0.009	0.025
$\alpha = 0.4$	0.080	0.031	0.022	0.049	0.022	0.013	0.022
$\alpha = 0.6$	0.180	0.046	0.028	0.086	0.028	0.014	0.017
$\alpha = 0.8$	0.304	0.112	0.075	0.215	0.075	0.014	0.016
$\alpha = 0.9$	0.511	0.285	0.158	0.435	0.158	0.025	0.021
$\alpha = 0.95$	0.606	0.396	0.308	0.525	0.308	0.052	0.025
Linear Time Trend							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.023	0.021	0.019	0.023	0.019	0.034	0.023
$\alpha = 0.2$	0.061	0.031	0.027	0.043	0.027	0.030	0.031
$\alpha = 0.4$	0.151	0.053	0.042	0.078	0.042	0.035	0.032
$\alpha = 0.6$	0.325	0.105	0.069	0.182	0.069	0.039	0.040
$\alpha = 0.8$	0.717	0.289	0.184	0.463	0.184	0.047	0.159
$\alpha = 0.9$	0.862	0.509	0.352	0.698	0.352	0.051	0.215
$\alpha = 0.95$	0.905	0.606	0.445	0.752	0.445	0.048	0.365

Table 5: Empirical Size, $n = 200$

No Trend Case							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.027	0.021	0.019	0.021	0.014	0.010	0.026
$\alpha = 0.2$	0.044	0.023	0.022	0.035	0.018	0.012	0.032
$\alpha = 0.4$	0.090	0.031	0.029	0.046	0.025	0.014	0.031
$\alpha = 0.6$	0.210	0.049	0.034	0.095	0.028	0.016	0.024
$\alpha = 0.8$	0.412	0.163	0.092	0.297	0.075	0.029	0.019
$\alpha = 0.9$	0.646	0.381	0.248	0.556	0.236	0.049	0.015
$\alpha = 0.95$	0.855	0.656	0.483	0.802	0.459	0.109	0.027
Linear Time Trend							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.038	0.033	0.032	0.036	0.031	0.032	0.036
$\alpha = 0.2$	0.080	0.047	0.041	0.060	0.036	0.027	0.047
$\alpha = 0.4$	0.182	0.075	0.060	0.110	0.050	0.036	0.044
$\alpha = 0.6$	0.445	0.152	0.113	0.235	0.088	0.052	0.056
$\alpha = 0.8$	0.807	0.452	0.316	0.650	0.231	0.086	0.069
$\alpha = 0.9$	0.906	0.772	0.645	0.885	0.526	0.156	0.094
$\alpha = 0.95$	0.949	0.855	0.788	0.915	0.718	0.265	0.096

Table 6: Empirical Size, $n = 300$

No Trend Case							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.029	0.027	0.024	0.027	0.019	0.011	0.031
$\alpha = 0.2$	0.045	0.034	0.028	0.034	0.021	0.018	0.046
$\alpha = 0.4$	0.094	0.036	0.032	0.036	0.028	0.022	0.043
$\alpha = 0.6$	0.225	0.061	0.036	0.061	0.035	0.028	0.041
$\alpha = 0.8$	0.433	0.135	0.078	0.135	0.065	0.036	0.039
$\alpha = 0.9$	0.698	0.361	0.202	0.361	0.175	0.049	0.031
$\alpha = 0.95$	0.801	0.603	0.446	0.603	0.429	0.122	0.034
Linear Time Trend							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.039	0.038	0.039	0.038	0.036	0.034	0.039
$\alpha = 0.2$	0.091	0.053	0.046	0.053	0.044	0.042	0.049
$\alpha = 0.4$	0.205	0.091	0.062	0.091	0.055	0.046	0.052
$\alpha = 0.6$	0.489	0.169	0.095	0.169	0.085	0.056	0.060
$\alpha = 0.8$	0.839	0.514	0.279	0.514	0.212	0.085	0.077
$\alpha = 0.9$	0.941	0.826	0.617	0.826	0.522	0.165	0.094
$\alpha = 0.95$	0.965	0.935	0.833	0.935	0.762	0.345	0.083

Table 7: Empirical Size, $n = 500$

No Trend Case							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$\alpha = 0$	0.035	0.034	0.034	0.034	0.032	0.027	0.055
$\alpha = 0.2$	0.059	0.045	0.043	0.045	0.038	0.028	0.056
$\alpha = 0.4$	0.099	0.051	0.046	0.051	0.040	0.030	0.051
$\alpha = 0.6$	0.230	0.069	0.048	0.069	0.041	0.030	0.050
$\alpha = 0.8$	0.489	0.201	0.109	0.201	0.060	0.039	0.047
$\alpha = 0.9$	0.722	0.509	0.262	0.509	0.191	0.051	0.039
$\alpha = 0.95$	0.848	0.775	0.655	0.775	0.502	0.163	0.036
Linear Time Trend							
	$M1$	M2	M3	M4	M5	M6	M7
$\alpha = 0$	0.041	0.038	0.039	0.038	0.037	0.032	0.041
$\alpha = 0.2$	0.097	0.057	0.046	0.057	0.044	0.039	0.047
$\alpha = 0.4$	0.213	0.089	0.065	0.089	0.060	0.055	0.057
$\alpha = 0.6$	0.524	0.172	0.103	0.172	0.087	0.078	0.078
$\alpha = 0.8$	0.883	0.517	0.265	0.517	0.179	0.092	0.090
$\alpha = 0.9$	0.952	0.863	0.705	0.863	0.506	0.171	0.102
$\alpha = 0.95$	0.989	0.962	0.895	0.962	0.810	0.385	0.135

Table 8: Empirical Power

No Trend Case							
	$M1$	$M2$	$M3$	$M4$	$M5$	$M6$	$M7$
$n = 100$	0.784	0.501	0.400	0.631	0.400	0.145	0.031
$n = 200$	0.947	0.798	0.727	0.875	0.665	0.358	0.025
$n = 300$	0.973	0.896	0.794	0.896	0.762	0.452	0.045
$n = 500$	0.992	0.962	0.914	0.962	0.806	0.635	0.047

Linear Time Trend							
	M1	M2	M3	M4	M5	M6	M7
$n = 100$	0.925	0.666	0.529	0.815	0.529	0.092	0.322
$n = 200$	0.992	0.938	0.886	0.972	0.825	0.416	0.112
$n = 300$	0.999	0.982	0.935	0.982	0.902	0.606	0.091
$n = 500$	0.999	0.998	0.992	0.998	0.979	0.804	0.253

Table 6 reports simulation results on the power of these tests. Again, the effects of the bandwidth and the sample size on the power of the tests are considered. The tests have good power in many cases (except for those with small samples and large bandwidths, and those using $M7$). As anticipated from test consistency, for each bandwidth choice, the power increases as n increases. It is also apparent from Table 8 that for each sample size, the power decreases as the bandwidth parameter increases. According to the asymptotic analysis, the distribution of the tests under the alternative hypothesis depends on n/M, with large M (given n) generally reducing power.

A final word on the test using bandwidth $M7$. In this case, the test has good size property but almost no power. Notice that, corresponding to the AR(1) plug-in method, $M7 = 1.1447(\widehat{\delta}n)^{1/3}$ where $\widehat{\delta} = 4\,\widehat{\rho}^2/(1-\widehat{\rho}^2)^2$ and $\widehat{\rho}$ is the AR(1) coefficient estimator. Under the alternative hypothesis, $\widehat{\delta}$ actually diverges to ∞. In fact, $M7$ is of order n under the alternative. As a result, the test is no longer consistent and has no asymptotic power against H_1.

5 Conclusion

This paper shows how the CUSUM test for structural change can be applied to cointegrating regression residuals. In particular, residual-based tests for the null hypothesis of cointegration can be constructed by looking at fluctuations in the residuals from the cointegrating regression. Asymptotic distributions of these tests were derived under both the null hypothesis and the alternative of no cointegration. The limit distributions are nonstandard and are functions of several Brownian motions, but depend only on the dimension of the system and on the form of any deterministic detrending. The new tests are shown to be consistent and their asymptotic behavior under the alternative is similar to that of the KPSS tests in the sense that their divergence rate depends on the bandwidth parameter. These tests complement conventional residual-based procedures and are a companion to the work of Hao and Inder (1996) on the use of CUSUM tests for structural breaks in cointegrating regression.

6 Appendix: Proofs

In this Appendix, we give proofs for the results under H_1. For asymptotic analysis of the test under H_0, readers are referred to Hao and Inder (1996) or an earlier version of the current paper which may be obtained from the authors.

6.1 Proof of Lemma 1

By definition,

$$\widehat{\Omega}_{xu} = \sum_{h=-M}^{M} k(\frac{h}{M}) C_{xu}(h),$$

where

$$C_{xu}(h) = \frac{1}{n} \sum_{t} \Delta x_t \widehat{u}_{t+h}, \ 1 \leq t, t+h \leq n.$$

Let and $\underline{z}_t' = (\underline{y}_t, \underline{x}_t')$ be the OLS detrended z_t, then

$$\widehat{u}_t = y_t - w_t' \widehat{\Pi} = (y_t, w_t') \begin{pmatrix} 1 \\ -\widehat{\Pi} \end{pmatrix} = \underline{z}_t' \begin{pmatrix} 1 \\ -\widehat{\beta} \end{pmatrix}.$$

Thus

$$\widehat{\Omega}_{xu} = \sum_{h=-M}^{M} k(\frac{h}{M}) \left(\frac{1}{n} \sum_{t} \Delta x_t \underline{z}_{t+h}' \right) \begin{pmatrix} 1 \\ -\widehat{\beta} \end{pmatrix}, 1 \leq t, t+h \leq n.$$

Following similar arguments as in Phillips (1991) and with $M = O(n^{1/3})$ as $n \to \infty$, we have

$$\begin{aligned} & \frac{1}{M} \sum_{h=-M}^{M} k(\frac{h}{M}) \left[\frac{1}{n} \sum_{t} \Delta x_t \underline{z}_{t+h}' \right] \\ \Rightarrow \quad & 2\pi \left(\frac{1}{2\pi} \int_{-1}^{1} k(s) ds \right) \int_0^1 dB_x \underline{B}_{zd}' + \Omega_{xz} \\ = \quad & 2\pi K(0) \int_0^1 dB_x \underline{B}_{zd}' + \Omega_{xz}. \end{aligned}$$

where $\underline{B}_{zd}(r) = B_z(r) - \left(\int B_z d'\right)\left(\int dd'\right)^{-1} d(r)$. Under the alternative hypothesis,

$$\widehat{\beta} \Rightarrow \left(\int \underline{B}_{xd}\underline{B}_{xd}'\right)^{-1}\left(\int \underline{B}_{xd}\underline{B}_{yd}\right) = \zeta_d,$$

thus

$$\begin{aligned}
\frac{1}{M}\widehat{\Omega}_{xu} &= \left[\frac{1}{M}\sum_{h=-M}^{M} k(\frac{h}{M})\left(\frac{1}{n}\sum_t \Delta x_t \underline{z}_{t+h}'\right)\right]\begin{pmatrix} 1 \\ -\widehat{\beta} \end{pmatrix} \\
&\Rightarrow \left[2\pi K(0)\int_0^1 dB_x \underline{B}_{zd}' + \Omega_{xz}\right]\begin{pmatrix} 1 \\ -\zeta_d \end{pmatrix} \\
&= 2\pi K(0)\int_0^1 dB_x \underline{B}_{zd}'\eta + \Omega_{xz}\eta \\
&= 2\pi K(0)\int_0^1 dB_x \left[\underline{B}_{yd} - \zeta_d'\underline{B}_{xd}\right] + \Omega_{xz}\eta \\
&= 2\pi K(0)\int_0^1 dB_x \underline{Q} + \Omega_*.
\end{aligned}$$

Similarly, we can show that

$$\frac{1}{M}\widehat{\Delta}_{xu} \Rightarrow 2\pi K_1(0)\int_0^1 dB_x \underline{B}_{zd}'\eta + \Lambda_{xz}\eta = 2\pi K_1(0)\int_0^1 dB_x \underline{Q} + \Lambda_*.$$

For $\widehat{\omega}_{u.x}^2 = \widehat{\omega}_u^2 - \widehat{\Omega}_{ux}\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu}$, we first look at $\widehat{\omega}_u^2$. Notice that

$$\widehat{\omega}_u^2 = \sum_{h=-M}^{M} k(\frac{h}{M}) C_{uu}(h),$$

where

$$C_{uu}(h) = \frac{1}{n}\sum_t \widehat{u}_t\widehat{u}_{t+h} = \frac{1}{n}\sum_t \begin{pmatrix} 1, & -\widehat{\beta} \end{pmatrix} \underline{z}_t\underline{z}_{t+h}' \begin{pmatrix} 1 \\ -\widehat{\beta} \end{pmatrix}, \ 1 \le t, t+h \le n.$$

Hence, by the same argument as before, for $M = O(n^{1/3})$ as $n \to \infty$, we have

$$\begin{aligned}
\frac{1}{nM}\widehat{\omega}_u^2 &= \frac{1}{M}\sum_{h=-M}^{M} k(\frac{h}{M})\left[\frac{1}{n}C_{uu}\right] \\
&\Rightarrow 2\pi K(0)\int_0^1 \underline{Q}^2.
\end{aligned}$$

Thus, under the alternative,

$$\widehat{\omega}_u^2 = O_p(nM),\ \widehat{\Omega}_{ux} = O_p(M),\ \text{and}\ \widehat{\Omega}_{ux}\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} = O_p(M^2),$$

$$\begin{aligned}
\frac{1}{nM}\widehat{\omega}_{u.x}^2 &= \frac{1}{nM}\widehat{\omega}_u^2 - \frac{1}{nM}\widehat{\Omega}_{ux}\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} \\
&= \frac{1}{nM}\widehat{\omega}_u^2 + o_p(1) \\
&\Rightarrow 2\pi K(0)\int_0^1 \underline{Q}^2.
\end{aligned}$$

6.2 Proof of Theorem 2

Under the alternative hypothesis,

$$\begin{aligned}
\widehat{u}_t^+ &= y_t^+ - w_t'\widehat{\Pi}^+ \\
&= y_t - w_t'\widehat{\Pi} - \Delta x_t'\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} + w_t'\left(\sum_s w_s w_s'\right)^{-1}\left(\sum_s w_s \Delta x_s'\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} + \left[\begin{array}{cc} n\widehat{\Lambda}_{xu}^+ & 0 \end{array}\right]\right) \\
&= \widehat{u}_t - \Delta x_t'\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} + w_t'\left(\sum_s w_s w_s'\right)^{-1}\left(\sum_s w_s \Delta x_s'\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} + \left[\begin{array}{cc} n\widehat{\Lambda}_{xu}^+ & 0 \end{array}\right]\right).
\end{aligned}$$

By the result of Lemma 1 and notice that

$$\widehat{\Lambda}_{xu}^+ = \widehat{\Lambda}_{xu} - \widehat{\Lambda}_{xx}\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} = O_p(M),$$

we have

$$\begin{aligned}
& \frac{1}{n^{3/2}}\sum_{t=1}^{[nr]}\widehat{u}_t^+ \\
=\; & \frac{1}{n^{3/2}}\sum_{t=1}^{[nr]}\widehat{u}_t - \frac{1}{n^{3/2}}x_{[nr]}'\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} \\
& + \frac{1}{n^{3/2}}\left(\sum_{t=1}^{[nr]}w_t'\right)\left(\sum_s w_s w_s'\right)^{-1}\left(\sum_s w_s \Delta x_s'\widehat{\Omega}_{xx}^{-1}\widehat{\Omega}_{xu} + \left[\begin{array}{cc} n\widehat{\Lambda}_{xu}^+ & 0 \end{array}\right]\right) \\
=\; & \frac{1}{n^{3/2}}\sum_{t=1}^{[nr]}\widehat{u}_t + o_p(1) \\
\Rightarrow\; & \overline{Q}(r).
\end{aligned}$$

Thus

$$\begin{aligned}
\sqrt{\frac{M}{n}}CS_n &= \max_{k=1,\ldots,n}\sqrt{\frac{M}{n}}\frac{1}{\widehat{\omega}_{u.x}\sqrt{n}}\left|\sum_{t=1}^{k}\widehat{u}_t^+\right| \\
&= \max_{k=1,\ldots,n}\frac{1}{\sqrt{n^{-1}M^{-1}\widehat{\omega}_{u.x}^2}}\left|\frac{1}{n}\sum_{t=1}^{k}\frac{\widehat{u}_t^+}{\sqrt{n}}\right| \\
&= \sup_{0\le r\le 1}\frac{1}{\sqrt{n^{-1}M^{-1}\widehat{\omega}_{u.x}^2}}\left|\frac{1}{n}\sum_{t=1}^{[nr]}\frac{\widehat{u}_t^+}{\sqrt{n}}\right| \\
&\Rightarrow \left[2\pi K(0)\int_0^1 \underline{Q}^2\right]^{-1/2}\sup_{0\le r\le 1}\left|\overline{Q}(r)\right|.
\end{aligned}$$

The result of Theorem 2 then follows immediately.

7 References

Andrews, D.W.K., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, **59**, 817–858.

Brown, R.L., J. Durbin, and J. Evans, 1975, "Techniques for testing the constancy of regression relationship over time," *J.R.S.S. (*B), 37, 149-63.

Chao, J. and P. C. B. Phillips (1999). "Bayesian model selection in partially nonstationary vector autoregressive processes with reduced rank structure," *Journal of Econometrics*, Vol 91, 227-271.

Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica* 55, 251 - 276

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

Johansen, S. (1995), *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford: Oxford University Press.

Hannan, E. J., 1970, Multiple Time Series, Wiley, New York.

Hansen, B., 1992a, "Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends", *Journal of Econometrics, 53, 87-121.*

Hansen, B., 1992b, "Tests for parameter instability in regressions with I(1) processes", *Journal of Business and Economic Statistics*, Vol. 10,321-335.

Hao, K. and B. Inder, 1996, "Diagnostic test for structural change in cointegrated regression models," *Economic Letters* 50, 179-187.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?", *Journal of Econometrics*, **54**, 159–178.

Park, J., S. Ouliaris, and B. Choi, 1988, "Spurious regressions and tests for cointegration", mimeo, Cornell University.

Park, J., P.C.B. Phillips, 1988, "Statistical inference in regressions with integrated process: Part I", *Econometric Theory*, **4**, 468-498.

Phillips, P.C.B., 1986, "Understanding spurious regression," *Journal of Econometrics*, **57**, 1361–1401.

Phillips, P.C.B., 1991a, "Optimal inference in cointegrated systems", *Econometrica*, Vol 59, 283-306

Phillips, P.C.B., 1991b, "Spectral regression for cointegrated time series", in *Nonparametric and semiparametric methods in econometrics and statistics*, edited by W.Barnett, J. Powell and G. Tauchen. Cambridge: Cambridge University Press.

Phillips, P.C.B., 1995, "Fully modified least squares and vector autoregression," *Econometrica*, 63, 1023-1078.

Phillips, P. C. B., 1996, "Econometric model determination". *Econometrica*, 64, pp. 763-812.

Phillips, P. C. B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies* 57, 99–125.

Phillips, P. C. B. and S. Ouliaris (1988), "Testing for Cointegration Using Principal Components Methods", *Journal of Economic Dynamics and Control*, 12, 205-230.

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

Phillips, P.C.B. and V. Solo, 1992, "Asymptotics for linear processes," *Annals of Statistics*, **20**, 971–1001.

Ploberger, W., and W. Kramer, 1992, "The CUSUM test with OLS residuals," *Econometrica*, **60**, 271–85.

Robinson, P. M. (1991), Automatic frequency domain inference on semiparametric and nonparametric models, *Econometrica* 59, 755-786.

Shin, Y., 1994, "A residual based test of the null of cointegration against the alternative of no cointegration," *Econometric Theory*, 10, 91—115.

Stock, J. H. and M. W. Watson (1988), "Testing for common trends", *Journal of American Statistical Associatio*n, 83, 1097 - 1107.

Xiao, Z., 2001, "Testing the null hypothesis of stationarity against an autoregressive unit root alternative," *Journal of Time Series Analysis, 22, 87-105*.

Xiao, Z. and P.C.B. Phillips, 1998, "A CUSUM test for Cointegration Using Regression Residuals," Working Paper, University of Illinois.