Testing the null of cointegration with structural breaks

Josep Lluís Carrion-i-Silvestre*
Anàlisi Quantitativa Regional
Parc Científic de Barcelona
Universitat de Barcelona

Andreu Sansó
Departament of Economics
Universitat de les Illes Balears

June 2005

Abstract

In this paper we propose an LM-Type statistic to test the null hypothesis of cointegration allowing for the possibility of a structural break, in both the deterministic and the cointegration vector. Our proposal focuses on the presence of endogenous regressors and analyses which estimation method provides better results. The test has been designed to be used as a complement to the usual non-cointegration tests in order to obtain stronger evidence of cointegration. We consider the cases of known and unknown break date. In the latter case, we show that minimizing the SSR results in a super-consistent estimator of the break fraction. Finally, the behaviour of the tests is studied through Monte Carlo experiments.

JEL classification: C12, C22

Keywords: cointegration, structural break, KPSS test

1 Introduction

Cointegration has made a great contribution in the field of economics and has given rise to a huge amount of theoretical and applied research. Since Granger (1981) and Engle and Granger (1987)

*We would like to thank Pierre Perron, the editor, Anindya Banerjee, and two anonymous referees for helpful comments. Corresponding author: Av. Diagonal, 690. 08034 Barcelona. Tel: +34 934021826, Fax: +34 934021821, E-mail: carrion@ub.edu The authors gratefully acknowledge the financial support of the Ministerio de Ciencia y Tecnología SEJ2004-05052/ECON and SEC2001-3672.

defined the concept, several methods have been proposed in the econometric literature to test the stationarity of a linear combination of non-stationary time series. Most of these procedures specify the null hypothesis of no cointegration against the alternative hypothesis of cointegration. However, this specification has been criticised in Engle and Granger (1987), Phillips and Ouliaris (1990) and Engle and Yoo (1991), where it is argued that the natural specification to test should be the null hypothesis of cointegration rather than the null hypothesis of lack of cointegration. Papers such as Hansen (1992b), Leybourne and McCabe (1993), Harris and Inder (1994), Shin (1994) and McCabe, Leybourne and Shin (1997) have addressed this question, reversing these hypotheses and designing tests for the null hypothesis of cointegration.

In fact, the definition of cointegration cannot be disassociated from the long-run. The property of cointegration implies that a set of variables moves jointly defining an equilibrium relationship, which means that it has to be tested using time series covering long periods. Practitioners have to bear in mind that the larger the sample period covering the time series, more likely it is that there will be a structural change. As shown by Perron (1989), this possibility changes the distribution of the standard unit root tests since the specification for the deterministic component of the time series needs to be modified. In these terms, the definition of cointegration under consideration falls into the classification of deterministic and stochastic cointegration of Park (1990). Perron's proposal was generalized to the cointegration tests of Engle and Granger (1987) and Phillips and Ouliaris (1990) by Gregory and Hansen (1996a, 1996b). They modify the usual procedures specifying a regime and trend shift model as the most general formulation. Departing from this general specification it is possible to obtain other simplified models that account for a level shift, with and without a time trend. One important thing to bear in mind is that the Gregory-Hansen tests detect regime-shift as well as stable cointegration relationships. Thus, the rejection of the null hypothesis of these tests does not entangle the instability of the cointegration relationship. In order to discern between these situations they suggest applying instability tests such as the ones in Hansen (1992b). However, this concern can be also addressed through the application of stationarity tests including structural breaks.

There have been some approximations in the literature to analyse cointegration with structural breaks. To this end, Hao (1996) generalizes the test of Kwiatkowski, Phillips, Schmidt and Shin (1992) to allow for a structural break that shifts the independent term of the cointegrating vector. In a recent paper, Bartley, Lee and Strazicich (2001) also deal with the KPSS cointegration test but include a shift in the mean and in the trend of the deterministic elements of the cointegrating relationship. Finally, Busetti (2002) adopts a multivariate framework when testing the stationarity of a set of variables allowing for one structural break. In brief, he extends the multivariate Nyblom and Harvey (2000) and Kwiatkowski et al. (1992) test statistic for those situations in which the time trend is affected by a structural break that shifts the level and/or the slope of the variables. Hence, assuming the date of the break to be known, his proposal is designed to test the breaking-stationarity hypothesis for a set of variables at once, although avoiding the presence of endogenous regressors. Moreover, the test can also be used to determine the dimension of the cointegration space, *i.e.* the number of cointegrating relationships among the variables.

In the present paper we extend the test of the null hypothesis of cointegration to allow for a structural break in both the parameters of the deterministic component and the parameters of the stochastic component. Our proposal generalizes the previous contributions and develops a test for cointegration around a break-cointegrating relationship. For instance, this approach is different from the one in Busetti (2002) since, firstly, we are primarily interested on the estimation of the cointegration relationship under structural breaks as in Gregory and Hansen (1996a) and, secondly, we account for the presence of endogenous regressors. Throughout the paper we highlight the differences that can be established between these above contributions and the ones proposed here. The break point is estimated through the minimisation of the sum of squared residuals, for which the rates of convergence of the estimated break fraction parameter are derived.

In section 2 we present the models and test statistics and derive their asymptotic distribution under two assumptions: (i) that the stochastic regressors are strictly exogenous and (ii) that the break point is assumed to be known. In section 3 we relax the first assumption and consider efficient estimators of the cointegrating vector. Section 4 deals with the estimation of the break point. Section 5 shows the consistency of the test statistics proposed in section 2. Section 6 analyses the performance of the tests in finite samples. Finally, section 7 concludes. All the proofs of the Theorems are presented together in the Appendix.

2 The models and tests

The model we deal with is a multivariate extension of the one specified by Kwiatkowski et al. (1992) where deterministic and/or stochastic components are allowed to change at a point of time (T_b). The data generating process (DGP) is of the form:

$$y_t = \alpha_t + \xi t + x_t' \beta_1 + \varepsilon_t, \tag{1}$$

$$x_t = x_{t-1} + \varsigma_t, \tag{2}$$

$$\alpha_t = f(t) + \alpha_{t-1} + \eta_t, \tag{3}$$

where $\eta_t \sim iid\left(0, \sigma_\eta^2\right)$, x_t is a k-vector of I(1) processes and $\alpha_0 = \alpha$, a constant. We define $f(t)$ as a function collecting the set of deterministic and/or stochastic components. The different models under consideration are specified through the definition of the function $f(t)$. Thus, following Perron (1989, 1990), the models that affect the deterministic component are:

- Model An, where $\xi = 0$ and $f(t) = \theta D(T_b)_t$;
- Model A, where $\xi \neq 0$ and $f(t) = \theta D(T_b)_t$;
- Model B, where $\xi \neq 0$ and $f(t) = \gamma DU_t$;
- Model C, where $\xi \neq 0$ and $f(t) = \theta D(T_b)_t + \gamma DU_t$.

where $D(T_b)_t = 1$ for $t = T_b + 1$ and 0 otherwise, $DU_t = 1$ for $t > T_b$ and 0 otherwise, with $T_b = \lambda T$, $0 < \lambda < 1$, indicating the date of the break. Notice that the null hypothesis of cointegration is equivalent to $\sigma_\eta^2 = 0$, under which the model given by (1), (2) and (3) turns out to be:

$$y_t = g_i(t) + x_t' \beta + \varepsilon_t, \tag{4}$$

where $g_i(t)$, $i = \{An, A, B, C\}$, denotes the deterministic function under the null hypothesis. To be exact, $g_{An}(t) = \alpha + \theta DU_t$ for model An, $g_A(t) = \alpha + \theta DU_t + \xi t$, $g_B(t) = \alpha + \xi t + \gamma DT_t^*$ and $g_C(t) = \alpha + \theta DU_t + \xi t + \gamma DT_t^*$, where $DT_t^* = (t - T_b)$ for $t > T_b$ and 0 otherwise. The specification given by the model An was proposed first in Hao (1996) whereas the one given by model C can be found in Bartley et al. (2001). These models account for the structural break effect shifting the deterministic part of the long-run relationship, although they assume that the cointegrating vector remains unchanged –see Gabriel, Da Silva and Nunes (2003) for an application to the long-run money demand for Portugal. The specification given for Model An accounts for a level shift without a time trend, while Model A contains a trend and allows for a break in level. Model B captures a change in the slope of the time trend but not in the level, whereas Model C accounts for both level and slope shifts.

It is possible to introduce the dummy variables in a way that affect the cointegrating vector. In this case, we have allowed for two different effects:

- Model D, where $\xi = 0$ and $f(t) = \theta D(T_b)_t + x_t' \beta_2 D(T_b)_t$;
- Model E, where $\xi \neq 0$ and $f(t) = \theta D(T_b)_t + \gamma DU_t + x_t' \beta_2 D(T_b)_t$.

Hence, under the null hypothesis, the model described by (1), (2) and (3) reduces to:

$$y_t = g_i(t) + x_t' \beta_1 + x_t' \beta_2 DU_t + \varepsilon_t, \tag{5}$$

$i = \{D, E\}$ with $g_D(t) = \alpha + \theta DU_t$ for model D and $g_E(t) = \alpha + \xi t + \theta DU_t + \gamma DT_t^*$ for model E. Now, the specification allows for a structural break that not only shifts the deterministic component but also changes the cointegrating vector. Thus, in some situations, practitioners would be interested in modelling a cointegration relationship that at a point in time might has shifted from one long-run path to another one –see, for instance, the modellisation of the aggregate consumption function in Hansen (1992b) and the estimation of the long-run money demand for the US in Gregory and Hansen (1996a). The difference between Models D and E is the time trend –Model D does not consider a time trend while Model E does. Notice that under the alternative hypothesis that $\sigma_\eta^2 > 0$, $(y_t, x_t')'$ will not be cointegrated.

As for the disturbance terms of the model, let us assume that the long-run variance matrix of $\vartheta_t = (\varepsilon_t, \varsigma_t', \eta_t)'$ is given by:

$$\Omega_a = \begin{bmatrix} \omega_1^2 & & 0 \\ & \Omega_{22} & \\ 0 & & \sigma_\eta^2 \end{bmatrix},$$

a block diagonal matrix which ensures, on the one hand, that ε_t and η_t, and, on the other hand, that ε_t and ς_t', are mutually uncorrelated. The assumption of no correlation between the disturbance terms of (1) and (2) introduces the restriction that x_t is strictly exogenous –see Gonzalo (1994). Section 3 shows how to proceed when this is not the case.

The (LM-type) statistic to test the null hypothesis of cointegration against the alternative of no cointegration is given by:

$$SC_i(\lambda) = T^{-2}\hat{\omega}_1^{-2} \sum_{t=1}^{T} S_{i,t}^2, \tag{6}$$

where $\lambda = T_b/T$, $S_{i,t} = \sum_{j=1}^{t} \hat{e}_{i,j}$, $\{\hat{e}_{i,t}\}_{t=1}^{T}$, are the OLS estimated residuals driven from (4) or (5), depending on the model, and $\hat{\omega}_1^2$ denotes a consistent estimator of the long-run variance of $\{\varepsilon_{i,t}\}_{t=1}^{T}, i = \{An, A, B, C, D, E\}$. The long-run variance can be estimated in a non-parametric way using the Bartlett Kernel. However, it should be noticed that the use of the standard automatic bandwidth selection methods leads to an inconsistent test. In order to avoid the inconsistency Kurozumi (2002) suggest to apply a modified version of the Andrews' data dependent procedure, which can be also applied in this framework. The asymptotic distribution of (6) is stated in the following Theorem.

Theorem 1 *Let $\{y_t\}_{t=1}^{T}$ be generated by (1), (2) and (3), and let $\zeta_t = \sum_{i=1}^{t} (\varepsilon_i, \varsigma_i')'$ satisfy the multivariate invariance principle of Phillips and Durlauf (1986). If $T_b = \lambda T$ $(0 < \lambda < 1)$ and as $T \to \infty$, $T_b \to \infty$, so λ remains constant, under the null hypothesis of cointegration $\left(\sigma_\eta^2 = 0\right)$:*

$$SC_i(\lambda) \Rightarrow \lambda^2 \int_0^1 V_{k,i}^2(b_1)\, db_1 + (1-\lambda)^2 \int_0^1 V_{k,i}^2(b_2)\, db_2,$$

for $i = \{An, A, B, C, D, E\}$, where $\Rightarrow$ denotes weak convergence and $V_{k,i}(\cdot)$ are functions of Wiener process shown in the Appendix.

The proof of Theorem 1 and the expressions for $V_{k,E}(b_1)$ and $V_{k,E}(b_2)$ can be found in the Appendix. The expressions for $V_{k,i}(b_1)$ and $V_{k,i}(b_1)$, $\{i = An,\ A,\ B,\ C,\ D\}$ are omitted although they are particular cases of the latter. Notice that the test is performed using the upper tail of the distribution so that the null hypothesis of cointegration is rejected when $SC_i(\lambda) >$ *critical value*, $i = \{An, A, B, C, D, E\}$. Two remarks are in order. First, the asymptotic distribution of Theorem 1 can be expressed as a weighted sum of two independent functionals of Wiener processes. As pointed out in Lee and Strazicich (2001) and Kurozumi (2002) for the univariate KPSS test with one structural break, the symmetry of the distribution around $\lambda = 0.5$ is given by the fact that we can interchange λ and $(1 - \lambda)$ in the asymptotic distribution and obtain the same result. Second, notice that the asymptotic distributions of Theorem 1 depend both on k, the number of elements of x_t that involves the specification, and on $\lambda = T_b/T$, the break fraction parameter. Therefore, they depend on the nuisance parameter λ. The econometric literature has addressed this concern in two different ways. First, in some cases, the date of the break can be assumed to be known and, hence, exogenous with respect to the model, so the break fraction does not need to be estimated –see Busetti (2002). This situation might be suitable, for instance, for many German macroeconomic time series for which the reunification process of 1990 had caused a shift in the deterministic part of the time series –see Lütkepohl, Muller and Saikkonen (1999). Hence, it seems desirable to have a set of critical values that assumes the date of the break to be exogenous. Critical values computed using 20,000 replications through direct simulation of the Wiener processes for up to five values of the break fraction, $\lambda = \{0.1, 0.2, \ldots, 0.5\}$, and for up to four stochastic regressors, $k = \{1, 2, 3, 4\}$, are shown in Tables 1 and 2.[1]

Up to now we have focused on those situations for which the date of the break is presumed to be known. Notwithstanding, sometimes the information about an event that might have caused a break is not sufficiently clear. In this situation, a second approach to deal with the dependency of the asymptotic distributions on the nuisance parameter λ is required and this parameter requires to be estimated. The procedures that can be applied to do so are considered in Section 4.

3 Non strictly exogenous regressors

The previous analysis has been based on the assumption that x_t is strictly exogenous, although this assumption may be very restrictive in practice. When this is the case, the asymptotic results

[1] As mentioned above, Bartley et al. (2001) analyse the specifications that correspond to models A and C, for which they compute the corresponding critical value sets. However, there is no mention anywhere of the fact that the asymptotic distribution of the test depends on k, as it does –see Theorem 1. Actually, if we compare their critical values sets with ours we see that theirs are almost equivalent to the ones computed here for $k = 1$. Practitioners should be aware that a size distortion problem (underrejection of the null hypothesis) is expected when using their critical values for those models with $k > 1$.

Table 1: Asymptotic critical values for the models An, A, B and C

	Model An					Model A				
	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$
			$k = 1$					$k = 1$		
90%	0.1932	0.1583	0.1395	0.1281	0.1256	0.0827	0.0736	0.0747	0.0821	0.0840
95%	0.2582	0.2087	0.1855	0.1632	0.1553	0.1028	0.0885	0.0907	0.1021	0.1060
97.5%	0.3367	0.2676	0.2341	0.1991	0.1855	0.1228	0.1054	0.1062	0.1229	0.1315
99%	0.4546	0.3543	0.2948	0.2503	0.2287	0.1537	0.1305	0.1251	0.1508	0.1642
			$k = 2$					$k = 2$		
90%	0.1336	0.1157	0.1079	0.1020	0.1029	0.0700	0.0630	0.0650	0.0690	0.0693
95%	0.1796	0.1557	0.1400	0.1306	0.1292	0.0865	0.0759	0.0774	0.0852	0.0858
97.5%	0.2325	0.2007	0.1759	0.1622	0.1557	0.1033	0.0891	0.0909	0.1023	0.1037
99%	0.3116	0.2631	0.2259	0.2035	0.1903	0.1273	0.1095	0.1083	0.1254	0.1348
			$k = 3$					$k = 3$		
90%	0.1007	0.0907	0.0856	0.0847	0.0840	0.0594	0.0554	0.0571	0.0581	0.0584
95%	0.1319	0.1179	0.1094	0.1063	0.1051	0.0728	0.0670	0.0692	0.0712	0.0725
97.5%	0.1670	0.1490	0.1338	0.1276	0.1271	0.0871	0.0784	0.0803	0.0843	0.0877
99%	0.2238	0.1989	0.1773	0.1602	0.1594	0.1064	0.0941	0.0971	0.1035	0.1103
			$k = 4$					$k = 4$		
90%	0.0799	0.0738	0.0712	0.0704	0.0706	0.0507	0.0490	0.0501	0.0509	0.0510
95%	0.1037	0.0924	0.0873	0.0878	0.0874	0.0616	0.0588	0.0606	0.0617	0.0621
97.5%	0.1304	0.1151	0.1091	0.1073	0.1056	0.0729	0.0691	0.0724	0.0728	0.0741
99%	0.1754	0.1502	0.1385	0.1365	0.1350	0.0898	0.0846	0.0864	0.0886	0.0938

	Model B					Model C				
	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$
			$k = 1$					$k = 1$		
90%	0.0842	0.0747	0.0663	0.0614	0.0604	0.0802	0.0661	0.0559	0.0493	0.0484
95%	0.1059	0.0919	0.0809	0.0730	0.0729	0.1000	0.0813	0.0673	0.0579	0.0562
97.5%	0.1280	0.1102	0.0973	0.0863	0.0844	0.1200	0.0973	0.0802	0.0669	0.0645
99%	0.1578	0.1313	0.1197	0.1039	0.1013	0.1470	0.1208	0.0967	0.0799	0.0746
			$k = 2$					$k = 2$		
90%	0.0723	0.0632	0.0579	0.0542	0.0533	0.0667	0.0567	0.0487	0.0450	0.0430
95%	0.0892	0.0775	0.0694	0.0651	0.0639	0.0828	0.0696	0.0583	0.0530	0.0498
97.5%	0.1069	0.0925	0.0825	0.0761	0.0752	0.0998	0.0824	0.0687	0.0614	0.0574
99%	0.1314	0.1161	0.1019	0.0940	0.0903	0.1235	0.1010	0.0843	0.0731	0.0680
			$k = 3$					$k = 3$		
90%	0.0602	0.0536	0.0507	0.0475	0.0470	0.0563	0.0493	0.0430	0.0403	0.0393
95%	0.0740	0.0657	0.0613	0.0575	0.0561	0.0688	0.0602	0.0518	0.0476	0.0463
97.5%	0.0884	0.0784	0.0724	0.0675	0.0663	0.0819	0.0716	0.0616	0.0551	0.0527
99%	0.1106	0.0969	0.0888	0.0805	0.0788	0.1023	0.0887	0.0754	0.0657	0.0618
			$k = 4$					$k = 4$		
90%	0.0523	0.0472	0.0443	0.0429	0.0421	0.0491	0.0428	0.0389	0.0368	0.0359
95%	0.0638	0.0574	0.0529	0.0511	0.0498	0.0603	0.0510	0.0464	0.0431	0.0415
97.5%	0.0757	0.0684	0.0626	0.0596	0.0578	0.0714	0.0614	0.0544	0.0495	0.0474
99%	0.0921	0.0834	0.0756	0.0711	0.0699	0.0871	0.0748	0.0659	0.0592	0.0559

Note: Percentage points of the asymptotic distribution are based on $n = 20,000$ replications using partial sums of $\varepsilon \sim iidN(0,1)$ random variables of $T = 2,000$ observations to approximate the Wiener process. λ denotes the break fraction and k is the number of stochastic regressors in the model.

Table 2: Asymptotic critical values for the models D and E

	Model D					Model E				
	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$
			$k = 1$					$k = 1$		
90%	0.1908	0.1547	0.1265	0.1098	0.1044	0.0808	0.0654	0.0538	0.0463	0.0436
95%	0.2560	0.2067	0.1670	0.1395	0.1309	0.1004	0.0804	0.0659	0.0552	0.0512
97.5%	0.3295	0.2631	0.2098	0.1729	0.1603	0.1205	0.0974	0.0784	0.0645	0.0587
99%	0.4463	0.3449	0.2699	0.2224	0.1941	0.1480	0.1223	0.0960	0.0763	0.0681
			$k = 2$					$k = 2$		
90%	0.1319	0.1087	0.0885	0.0760	0.0735	0.0671	0.0540	0.0448	0.0387	0.0363
95%	0.1759	0.1459	0.1163	0.0969	0.0922	0.0832	0.0661	0.0544	0.0462	0.0423
97.5%	0.2288	0.1873	0.1485	0.1198	0.1123	0.0994	0.0790	0.0639	0.0534	0.0488
99%	0.3068	0.2510	0.1942	0.1578	0.1419	0.1218	0.0980	0.0795	0.0641	0.0574
			$k = 3$					$k = 3$		
90%	0.0983	0.0803	0.0664	0.0572	0.0542	0.0561	0.0457	0.0375	0.0323	0.0309
95%	0.1286	0.1049	0.0851	0.0721	0.0672	0.0696	0.0559	0.0454	0.0379	0.0360
97.5%	0.1638	0.1363	0.1079	0.0883	0.0819	0.0828	0.0658	0.0542	0.0444	0.0406
99%	0.2307	0.1816	0.1425	0.1145	0.1039	0.1040	0.0821	0.0660	0.0529	0.0474
			$k = 4$					$k = 4$		
90%	0.0772	0.0616	0.0512	0.0451	0.0423	0.0484	0.0391	0.0326	0.0282	0.0266
95%	0.0981	0.0791	0.0648	0.0548	0.0514	0.0597	0.0476	0.0393	0.0329	0.0308
97.5%	0.1225	0.1002	0.0806	0.0658	0.0613	0.0719	0.0572	0.0463	0.0379	0.0353
99%	0.1579	0.1312	0.1048	0.0852	0.0766	0.0899	0.0703	0.0570	0.0454	0.0411

Note: Percentage points of the asymptotic distribution are based on $n = 20,000$ replications using partial sums of $\varepsilon \sim iidN(0,1)$ random variables of $T = 2,000$ observations to approximate the Wiener process. λ denotes the break fraction and k is the number of stochastic regressors in the model.

obtained in the previous section no longer hold, because the estimation of the vector of cointegration is not efficient and the limiting distribution depends on nuisance parameters. However, several methods such as those of Phillips and Hansen (1990), Saikkonen (1991) and Stock and Watson (1993), which are asymptotically equivalent, can be applied to obtain an efficient estimation of the vector of cointegration. In this paper we use the Dynamic OLS estimator of Stock and Watson (1993).[2] The steps that have to be taken to test the null hypothesis of cointegration in this framework are:

1. Estimate
$$y_t = g_i(t) + x_t'\beta + \sum_{j=-k}^{k} \Delta x_t'\gamma_j + \varepsilon_t \tag{7}$$
if $i = \{An, A, B, C\}$, or
$$y_t = g_i(t) + x_t'\beta_1 + x_t'\beta_2 DU_t + \sum_{j=-k}^{k} \Delta x_t'\gamma_j + \varepsilon_t, \tag{8}$$
when $i = \{D, E\}$ and store the estimated residuals $\{\hat{e}_{i,t}\}$, $i = \{An, A, B, C, D, E\}$;

2. Compute the test statistic as:
$$SC_i^+(\lambda) = T^{-2}\hat{\omega}_1^{-2} \sum_{t=1}^{T} \left(S_{i,t}^+\right)^2,$$
where $\hat{\omega}_1^2$ is a consistent estimation of the long-run variance of $\{\varepsilon_t\}$ using $\hat{e}_{i,t}$ and $S_{i,t}^+ = \sum_{j=1}^{t} \hat{e}_{i,t}, i = \{An, A, B, C, D, E\}$.

The asymptotics concerning the test statistics are given in the following Theorem.

Theorem 2 *Let* $\{y_t\}_{t=1}^T$ *be generated by (1), (2) and (3), and let* $\zeta_t = \sum_{i=1}^t (\varepsilon_i, \varsigma_i')'$ *satisfy the multivariate invariance principle of Phillips and Durlauf (1986) with log-run variance matrix* Ω*. If* $T_b = \lambda T$ $(0 < \lambda < 1)$ *and as* $T \to \infty$, $T_b \to \infty$*, so* λ *remains constant, under the null hypothesis of cointegration:*

$$SC_i^+(\lambda) \Rightarrow \lambda^2 \int_0^1 V_{k,i}^2(b_1)\, db_1 + (1-\lambda)^2 \int_0^1 V_{k,i}^2(b_2)\, db_2,$$

[2] We also tried the Fully Modified estimator of Phillips and Hansen (1990) but yielded poor results in simulation experiments.

$i = \{An, A, B, C, D, E\}$, *being* $V_{k,i}(\cdot)$ *the same functions defined in Theorem 1.*

The proof can be found in Carrion-i-Silvestre and Sansó (2001). Theorem 2 states that, after endogeneity is accounted for the asymptotic distributions of the test statistics are the same as those assuming that x_t is strictly exogenous and, hence, the critical values presented in Tables 1 and 2 can be used.

4 Unknown date of the break

In previous sections we have implicitly assumed that the date of the break was known *a priori*. However, the break point often has to be estimated in empirical applications. The approach followed in this paper is based on the analysis in Bai (1994, 1997) and Kurozumi (2002), and estimates the break point as the date that minimises the sequence of the sum of squared residuals. In fact, this is also the approach in Bartley et al. (2001), since they select the break point as the argument that minimizes the sequence of the Bayesian Information Criterion (BIC) that is obtained when computing the test for all possible break points. Furthermore, the dynamic optimization algorithm in Bai and Perron (1998, 2003) can also be applied to minimise the sum of squared residuals. After the break point is estimated, we can proceed to the computation of the test assuming it as if it were known. Formally speaking,

$$\tilde{T}_b = \arg\min_{\lambda \in \Lambda} \left[SSR(\lambda)\right], \tag{9}$$

where $SSR(\lambda)$ is the sum of the squared residuals of (4) or (5), depending on the model, and Λ denotes a closed subset of the interval $(0, 1)$, that, in order to minimise the lost of information, we define it as $\Lambda = [2/T, (T-1)/T]$. We denote the tests that are computed using this strategy as $SC_i\left(\tilde{\lambda}\right)$ and $SC_i^+\left(\tilde{\lambda}\right)$, $i = \{An, A, B, C, D, E\}$. As shown in the following Theorem, the application of this approach leads to a consistent estimate of the break fraction parameter.

Theorem 3 *Let* $\{y_t\}_{t=1}^T$ *be generated by (1), (2) and (3) with* $\sigma_\eta^2 = 0$ *and either* $\theta \neq 0$ *or* $\gamma \neq 0$ *depending on the model, where* $\{\zeta_t\}_{t=1}^T$ *satisfies the multivariate invariance principle of Phillips* *and Durlauf (1986). If* $\tilde{T}_b = \arg\min_{t\in(1,T)}[SSR(\lambda)]$ *then, as* $T \to \infty$*,*

$$\tilde{\lambda} \xrightarrow{p} \lambda,$$

and

$$P\left(T\left|\tilde{\lambda}-\tau\right| \geq C\right) \leq \eta,$$

for arbitrary $\eta > 0$ *and* $C > 0$*,where* $\xrightarrow{p}$ *denotes convergence in probability.*

The proof is shown in the Appendix. Note that the result in Theorem 3 holds irrespective of the magnitude of the shifts. More interestingly, the estimation of the break fraction is T-consistent. Then, we can proceed to perform the hypothesis testing assuming the estimated break fraction as if it were known and, hence, using the critical values reported in Tables 1 and 2.

5 Consistency

The consistency of the test statistics against the alternative hypothesis of no cointegration is shown by deriving their asymptotic distribution under the alternative, and showing that the test statistics diverges. In the following theorem we state the limiting distribution of the tests under the alternative hypothesis.

Theorem 4 *Let* $\{y_t\}$ *be generated by 1, 2 and 3 with* $\sigma_\eta^2 > 0$*, and where* $\{\zeta_t\}_{t=1}^T$ *satisfies the multivariate invariance principle of Phillips and Durlauf (1986). If* $T_b = \lambda T$ $(0 < \lambda < 1)$ *and as* $T \to \infty$, $T_b \to \infty$*, so* λ *remains constant:*

$$SC_i(\lambda) = O_p(T/l),$$

$i = \{An, A, B, C, D, E\}$, *where* l *is the bandwidth of the spectral window used to estimate the long run variance of* ε_t*.*

The proof follows from Kwiatkowski et al. (1992) and is omitted in order to save space, although it can be found in Carrion-i-Silvestre and Sansó (2001). Theorem 4 shows that under the alternative hypothesis the tests diverge at rate $O_p(T/l)$. This conclusion has also been reached in Harris and Inder (1994) and Shin (1994) for cointegration test without structural breaks. Two remarks are in order. First, the results given in theorem 4 are valid regardless of the assumption of strictly exogenous stochastic regressors. Second, the rate of divergence of the statistics depends on the spectral bandwidth that is used, so we expect lag parameter selection to influence the power of the test statistics. Therefore, we conclude that the test is consistent when there is no cointegration.

6 Finite sample performance

The DGP that is used in this Section to assess the finite sample performance of the test statistic is similar to the ones specified by Banerjee, Dolado, Hendry and Smith (1986), Gonzalo (1994), Gregory and Hansen (1996a) and McCabe et al. (1997), among others, and is given by the triangular system representation of the CI(1, 1) process:

$$\begin{aligned} y_t - g(t) - \beta_t x_t &= z_t, \qquad z_t = \rho z_{t-1} + \alpha_t + e_{z_t}, \\ a_1 y_t - a_2 x_t &= w_t, \qquad w_t = w_{t-1} + e_{w_t}, \\ \alpha_t &= \alpha_{t-1} + \eta_t, \\ \begin{pmatrix} e_{z_t} \\ e_{w_t} \\ \eta_t \end{pmatrix} &\sim iid\ N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta\sigma & 0 \\ \delta\sigma & \sigma^2 & 0 \\ 0 & 0 & \sigma_\eta^2 \end{pmatrix}\right], \end{aligned}$$

with $|\rho| < 1$. This DGP allows us to analyse how the test statistics behave under different situations. First, under the null hypothesis of cointegration, $\sigma_\eta^2 = 0$, ρ accounts for the presence of correlation in the error of the cointegrating regression. Second, the DGP will be under the alternative of no cointegration when $\sigma_\eta^2 > 0$. Third, δ denotes the correlation between $\{e_{w_t}\}$ and $\{e_{z_t}\}$. Fourth, $g(t)$ represents the deterministic component of the long-run equilibrium relationship. In this simulation experiment we have considered the model D so that $g_D(t) = \alpha + \theta DU_t$. Finally, when $a_1 = 0$, x_t will be strictly exogenous. On the other hand, when $a_1 \neq 0$, x_t is no longer strictly exogenous. It should also be noticed that the vector of parameter associated to stochastic regressors depends on time, so $\beta_t = \beta_1$ for $t \leq T_b$ and $\beta_t = \beta_2$ for $t > T_b$.

We consider the parameter space $(a_1, a_2, \alpha, \theta, \beta_t, \lambda, \rho, \sigma, \delta, \sigma_\eta^2)$ where $a_1 =\{0, 1\}$, $a_2 = -1$, $\alpha = 1$, $\theta = \{0.5, 1\}$, $\beta_t = (\beta_1, \beta_2) = \{(2, 1), (2, 4)\}$, $\lambda =\{0.25, 0.5, 0.75\}$, $\rho =\{0, 0.5\}$, $\sigma =\{0.25, 0.33, 0.5, 1, 2\}$, $\delta =\{-0.5, 0, 0.5\}$ and $\sigma_\eta^2 =\{0, 0.001, 0.01, 0.1\}$. The sample size is set to $T = 200$ and an amount of $n = 1,000$ replications of the DGP are carried out for each parameterisation. Using this specification we have analysed the finite sample performance of our proposal in two stages. First, we focus on the situation that the break point is assumed to be known. Second, we address the estimation of the date of the break shading light on the ability of the procedure for dating the structural break. This is of special interest, because the cointegration testing requires a consistent estimate of the break point. At this point, we proceed to the analysis of the empirical size and power of the test using the estimation method of the date of the break that offers better results. That is the reason why we decided to carry out the simulations using the parametric correction instead of the non-parametric one. The number of leads and lags has been selected using the BIC information criterion with an initial value of 4 leads and lags. In addition, to avoid the inconsistency on the estimation of the long-run covariance matrix we have followed Kurozumi (2002), and estimate it using the Bartlett kernel with the modified automatic lag length selection Andrews' procedure. In brief, Kurozumi (2002) proposes estimating the bandwidth for the Bartlett kernel as:

$$\hat{l} = \min \left\{ 1.1447 \left(\frac{4\hat{a}^2 T}{(1+\hat{a})^2 (1-\hat{a})^2} \right)^{1/3}, 1.1447 \left(\frac{4k^2 T}{(1+k)^2 (1-k)^2} \right)^{1/3} \right\},$$

where $\hat{a}$ is the estimate of the autoregressive parameter that produces Andrews method. The simulations that he carries out suggest to use $k = 0.7$ or $k = 0.8$ as values that keep a compromise between the empirical size and power of the test –here we use $k = 0.8$. The GAUSS codes that have been used in the simulations are available upon request.

6.1 The break point is assumed to be known

The simulation results for $\lambda = 0.5$, strictly exogenous regressor ($a_1 = 0$), $\theta = 0.5$ and $\beta_t = (2, 1)$ are summarized in Table 3. Qualitatively similar results are obtained for the other values of λ, for $\beta_t = (2, 4)$ and $\theta = 1$, so they are omitted in order to save space, although the complete set of results is available from the authors upon request. Also, the rejections frequencies for $\delta = -0.5$ are similar to those of $\delta = 0.5$, so we only report the last ones.

When x_t is strictly exogenous ($a_1 = 0$) and there is no serial correlation of the error in the cointegrating regression ($\rho = 0$) there are no serious problems of size distortion, regardless of the break point position, the magnitude of the structural break and the covariance between disturbance terms of the model. When there is autocorrelation ($\rho = 0.5$) the tests suffers certain size distortion problems. Note also, that the oversize of $SC_D^+(\lambda)$ is lower than that of $SC_D(\lambda)$. As can be seen in Table 3, if x_t is no longer strictly exogenous ($a_1 \neq 0$) the discrepancy between empirical and nominal size increases, though as the signal-to-noise ratio grows the size distortion tends to decrease. As expected, in this situation the $SC_D^+(\lambda)$ test generally outperforms $SC_D(\lambda)$. Hence, the $SC_D^+(\lambda)$ statistic has better size than $SC_D(\lambda)$ in all the scenarios considered.

Table 3: Empirical size for test SC_D with known break date

		Panel A: strictly exog. reg. ($a_1 = 0$)				**Panel B: non strictly exog. reg.** ($a_1 = 1$)			
		$SC_D(\lambda)$		$SC_D^+(\lambda)$		$SC_D(\lambda)$		$SC_D^+(\lambda)$	
δ	$\sigma \backslash \rho$	0	0.5	0	0.5	0	0.5	0	0.5
0	0.25	0.054	0.087	0.052	0.082	0.965	0.783	0.137	0.191
	0.33	0.058	0.103	0.056	0.100	0.963	0.752	0.130	0.168
	0.5	0.048	0.086	0.045	0.081	0.935	0.676	0.121	0.155
	1	0.058	0.088	0.051	0.087	0.768	0.451	0.102	0.139
	2	0.072	0.095	0.066	0.094	0.493	0.270	0.077	0.104
0.5	0.25	0.086	0.138	0.056	0.100	0.965	0.759	0.124	0.139
	0.33	0.082	0.142	0.050	0.097	0.957	0.662	0.131	0.134
	0.5	0.090	0.154	0.058	0.092	0.895	0.502	0.099	0.100
	1	0.076	0.140	0.052	0.084	0.561	0.238	0.068	0.087
	2	0.086	0.146	0.048	0.097	0.214	0.127	0.073	0.082

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; \theta = 0.5; \beta_t = (2, 1); e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0, \sigma^2); E(e_{z_t} e_{w_t}) = \delta\sigma; T = 200, T_b = [\lambda T]; \lambda = 0.5; n = 1,000$ replications; 5% nominal size

Next, we consider the power of the statistics when the break date is known. The DGP is as the previous one, but with σ_η^2 ={0.001, 0.01, 0.1}. This set of values is the same as in McCabe et al. (1997). Table 4 reports the rejection frequencies for $\lambda = 0.5$. Similar results were obtained for other break fractions and values of the rest of the parameters. The $SC_D(\lambda)$ statistic has a very high power when $a_1 = 1$ which can be explained by the size distortion that shows in this situation. Except for this case, the power attained by the tests is slightly smaller than the one obtained by McCabe et al. (1997), were they assume that there is no break in the cointegration relationship, weak exogeneity and used a parametric approach to deal with the serial correlation.

Table 4: Empirical power for test SC_D with known break date

Panel A: strictly exogenous regressors $(a_1 = 0)$				
	$\theta = 0.5; \beta_t = (2, 1)$		$\theta = 1; \beta_t = (2, 4)$	
σ^2_η	$SC_D(\lambda)$	$SC^+_D(\lambda)$	$SC_D(\lambda)$	$SC^+_D(\lambda)$
0.001	0.169	0.162	0.161	0.153
0.01	0.600	0.588	0.612	0.595
0.1	0.830	0.82	0.788	0.773
Panel B: non strictly exogenous regressors $(a_1 = 1)$				
	$\theta = 0.5; \beta_t = (2, 1)$		$\theta = 1; \beta_t = (2, 4)$	
σ^2_η	$SC_D(\lambda)$	$SC^+_D(\lambda)$	$SC_D(\lambda)$	$SC^+_D(\lambda)$
0.001	0.778	0.178	0.794	0.323
0.01	0.858	0.506	0.886	0.625
0.1	0.828	0.619	0.821	0.652

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \alpha_t + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; \alpha_t = \alpha_{t-1} + \eta_t; g_D(t) = \alpha + \theta DU_t; \theta = 0.5; \beta_t = (2, 1); e_{z_t} \sim iidN(0, 1); e_{w_t} \sim iidN(0, \sigma^2); \eta_t \sim iidN(0, \sigma^2_\eta); E(e_{z_t} e_{w_t}) = \delta\sigma; \sigma^2 = 1; \delta = 1; T = 200, T_b = [\lambda T]; \lambda = 0.5; n = 1,000$ replications; 5% nominal size

6.2 The break point is estimated

In this subsection we analyse the finite sample performance of the test statistic when the break point is estimated using the procedure described in Section 4, see expression (9). We use the notation $\tilde{T}^+_b$ when the estimator is based on (8). We start considering the estimation of the break date. As shown in Table 5, the estimators based on the minimization of the sum of the squared residuals, $\tilde{T}_b$ and $\tilde{T}^+_b$, usually locate correctly the break fraction and the error, if any, tends to be small. A similar picture was obtained for other values of the parameters of the DGP.

Let us consider the size of the statistics. Panel A in Table 6 reports the results assuming strictly exogenous regressors, while panel B presents the empirical size when the stochastic regressors are not strictly exogenous. These rejection frequencies are very similar to those of Table 3 were the break fraction were known. Even the slight size distortion observed for the SC^+ statistic seems to be smaller in Table 6. Hence, we may conclude that estimation the break date with (9) and then using this break date to compute the test statistic has a very little impact on the size of the test. As in the case of known break date, the statistic SC^+_D outperforms SC_D in all the scenarios.

Table 5: Frequencies of correct detection of the unknown break date

Panel A: strictly exogenous regressors ($a_1 = 0$)				
	$\theta = 0.5; \beta_t = (2,1)$		$\theta = 1; \beta_t = (2,4)$	
λ	$\tilde{T}_b$	$\tilde{T}_b^+$	$\tilde{T}_b$	$\tilde{T}_b^+$
0.1	0.78 (0.03)	0.78 (0.04)	0.95 (0.00)	0.95 (0.00)
0.25	0.85 (0.02)	0.85 (0.02)	0.96 (0.00)	0.95 (0.00)
0.5	0.87 (0.02)	0.87 (0.02)	0.98 (0.00)	0.98 (0.00)
0.75	0.91 (0.01)	0.92 (0.01)	0.97 (0.00)	0.97 (0.00)
0.9	0.90 (0.01)	0.89 (0.01)	0.98 (0.00)	0.98 (0.00)
Panel B: non strictly exogenous regressors ($a_1 = 1$)				
	$\theta = 0.5; \beta_t = (2,1)$		$\theta = 1; \beta_t = (2,4)$	
λ	$\tilde{T}_b$	$\tilde{T}_b^+$	$\tilde{T}_b$	$\tilde{T}_b^+$
0.1	0.26 (0.53)	0.24 (0.32)	0.55 (0.28)	0.33 (0.09)
0.25	0.42 (0.34)	0.30 (0.24)	0.69 (0.16)	0.49 (0.07)
0.5	0.52 (0.26)	0.44 (0.15)	0.80 (0.09)	0.72 (0.04)
0.75	0.57 (0.27)	0.54 (0.16)	0.82 (0.10)	0.78 (0.05)
0.9	0.43 (0.44)	0.48 (0.18)	0.72 (0.19)	0.62 (0.07)

NOTE: Frequencies of correct location of the date break and frecuencies of location errors bigger than 5 between brackets. DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; \theta = 0.5; \beta_t = (2,1); e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0,\sigma^2); E(e_{z_t} e_{w_t}) = \delta\sigma; T = 200, T_b = [\lambda T]; n = 1,000$ replications;

Table 6: Empirical size for test SC_D with unknown break date

		Panel A: strictly exog. reg. ($a_1 = 0$)				**Panel B: non strictly exog. reg.** ($a_1 = 1$)			
		$SC_D\left(\tilde{\lambda}\right)$		$SC_D^+\left(\tilde{\lambda}\right)$		$SC_D\left(\tilde{\lambda}\right)$		$SC_D^+\left(\tilde{\lambda}\right)$	
δ	$\sigma \backslash \rho$	0	0.5	0	0.5	0	0.5	0	0.5
0	0.25	0.051	0.088	0.053	0.085	0.812	0.262	0.115	0.136
	0.33	0.056	0.082	0.053	0.078	0.855	0.624	0.115	0.123
	0.5	0.061	0.093	0.056	0.084	0.827	0.606	0.091	0.119
	1	0.063	0.079	0.056	0.077	0.737	0.557	0.101	0.097
	2	0.034	0.109	0.039	0.112	0.476	0.395	0.085	0.097
0.5	0.25	0.075	0.144	0.050	0.095	0.836	0.575	0.110	0.120
	0.33	0.085	0.135	0.044	0.085	0.849	0.576	0.106	0.128
	0.5	0.085	0.143	0.055	0.094	0.784	0.436	0.093	0.107
	1	0.106	0.122	0.059	0.072	0.507	0.238	0.065	0.078
	2	0.097	0.145	0.062	0.091	0.212	0.133	0.064	0.082

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; \theta = 0.5; \beta_t = (2,1); e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0,\sigma^2); E(e_{z_t} e_{w_t}) = \delta\sigma; T = 200, T_b = [\lambda T]; \lambda = 0.5; n = 1,000$ replications; 5% nominal size

Regarding the empirical power of the test, Table 7 shows the rejection frequencies in different situations. Comparing these results with those of Table 4, when the regressors are strictly exogenous ($a_1 = 0$, Panel A), the figures are very similar, whereas for non strictly exogenous regressors, the statistics based on the estimation of the break fraction has a slight reduction in power.

Table 7: Empirical power for test SC_D with unknown break date

Panel A: strictly exogenous regressors ($a_1 = 0$)				
	$\theta = 0.5; \beta_t = (2,1)$		$\theta = 1; \beta_t = (2,4)$	
σ^2_η	$SC_D\left(\widetilde{\lambda}\right)$	$SC^+_D\left(\widetilde{\lambda}\right)$	$SC_D\left(\widetilde{\lambda}\right)$	$SC^+_D\left(\widetilde{\lambda}\right)$
0.001	0.156	0.150	0.162	0.160
0.01	0.591	0.586	0.599	0.583
0.1	0.791	0.762	0.821	0.808
Panel B: non strictly exogenous regressors ($a_1 = 1$)				
	$\theta = 0.5; \beta_t = (2,1)$		$\theta = 1; \beta_t = (2,4)$	
σ^2_η	$SC_D\left(\widetilde{\lambda}\right)$	$SC^+_D\left(\widetilde{\lambda}\right)$	$SC_D\left(\widetilde{\lambda}\right)$	$SC^+_D\left(\widetilde{\lambda}\right)$
0.001	0.740	0.235	0.783	0.199
0.01	0.761	0.578	0.849	0.582
0.1	0.673	0.651	0.740	0.724

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \alpha_t + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; \alpha_t = \alpha_{t-1} + \eta_t; g_D(t) = \alpha + \theta DU_t; \theta = 0.5; \beta_t = (2,1); e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN\left(0,\sigma^2\right); \eta_t \sim iidN\left(0,\sigma^2_\eta\right); E\left(e_{z_t} e_{w_t}\right) = \delta\sigma; \sigma^2 = 1; \delta = 1; T = 200, T_b = [\lambda T]; \lambda = 0.5; n = 1{,}000$ replications; 5% nominal size

Finally, in simulations not reported here, we find that the size and the power of $SC_D\left(\widetilde{\lambda}\right)$ and $SC^+_D\left(\widetilde{\lambda}\right)$ were similar to those of the previous tables when there were no structural break, that is $\theta = 0$ and β constant for all the sample.

To sum up, the main conclusion that arises from these simulation experiments is that the estimation of the break date minimising the SSR has little effect on the performance of the test statistics. The statistics SC^+ have, in general, a correct size with small size distortions for correlated errors, which is a common feature of all the statistics of this class. In terms of size, this statistic outperforms SC in all the scenarios considered, especially when the regressors are not strictly exogenous. So, we recommend using the SC^+ statistic. The test has also an adequate power, which is only slightly reduced when the break date has to be estimated. Hence, the test statistics presented in this paper can be seen as complementary tools for no cointegration tests with structural breaks and their use is recommended in empirical research.

7 Conclusions

This paper has focused on testing the null hypothesis of cointegration allowing for one structural break. Our approach can be understood as a complementary method to ensure the presence of cointegration around a break-cointegrating relationship. The test statistic used in the paper is a multivariate extension of the KPSS test, which was first proposed in Harris and Inder (1994) and Shin (1994) to test the null hypothesis of cointegration. We considered six different possibilities of structural break and two different situations depending on the properties of the set of stochastic regressors: when they are strictly exogenous they are not. In this last case, we suggest using the DOLS estimation of Stock and Watson (1993).

The paper has distinguished two different situations depending on whether the break point is known or unknown. The estimation of the break point bases on the minimisation of the sum of the squared residuals, which provides T-consistent estimates of the break fraction. Monte Carlo experiments show that the statistics that correct for the presence of endogenous regressors has good size and power properties irrespective of whether they are strictly or non-strictly exogenous regressors. Hence, we recommend using this test in empirical applications as a complement of non-cointegration tests with structural breaks.

A Mathematical Appendix

In this Appendix we outline the proof of the Theorems that have been proposed during the paper, although a more detailed exposition can be found in the working paper of this article –see Carrion-i-Silvestre and Sansó (2001). Some intermediate results are required beforehand. We use the following lemmas throughout.

Lemma 1 *The assumption of multivariate invariance principle of Phillips and Durlauf (1986) states that in the limit* $\{\zeta_t\}_{t=1}^T$, $\zeta_t = \sum_{i=1}^t (\varepsilon_i, \varsigma_i')'$, *converges to:*

$$\begin{bmatrix} T^{-1/2}\omega_1^{-1}\sum_{i=1}^t \varepsilon_i \\ T_{22}^{-1/2}\Omega_{22}^{-1/2}\sum_{i=1}^t \varsigma_i \end{bmatrix} \Rightarrow \begin{bmatrix} W_{11}(r) \\ W_{2k}(r) \end{bmatrix} = W(r); \quad t/T \leq r < (t+1)/T$$

$t = 1, ..., T$, *where* $\Rightarrow$ *denotes weak convergence of the associated probability measures and* $W(r)$ *is the* $(k+1)$*-vector of standard Wiener processes defined on* $C[0,1]^{k+1}$ *with*

$$\Omega = \lim_{T\to\infty} T^{-1}E(\zeta_T\zeta_T') = \begin{bmatrix} \omega_1^2 & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{bmatrix}$$

as the covariance matrix.

Proof: *see Phillips and Durlauf (1986).*

Lemma 2 *Let* $\{x_t\}_{t=1}^T$ *be generated by (2) and* $\{\zeta_t\}_{t=1}^T$ *a* $(k+1)$*-vector of stochastic processes that satisfies the multivariate invariance principle of Phillips and Durlauf (1986). Therefore, it can be shown that* $T^{-3/2}\sum_{t=1}^T x_t \Rightarrow \Omega_{22}^{1/2}\int_0^1 W_{2k}(r)\,dr$; $T^{-3/2}\sum_{t=1}^T x_t DU_t \Rightarrow \Omega_{22}^{1/2}\int_\lambda^1 W_{2k}(r)\,dr$; $T^{-1}\sum_{t=1}^T x_t\varepsilon_t \Rightarrow \omega_1\Omega_{22}^{1/2}\int_0^1 W_{2k}(r)\,dW_{11}(r)+\delta_{21}$; $T^{-2}\sum_{t=1}^T x_t x_t' \Rightarrow \Omega_{22}^{-1/2}\int_0^1 W_{2k}(r)\,W_{2k}'(r)\,dr\,\Omega_{22}^{-1/2}$ *where* $W(r)$ *is the* $(k+1)$*-vector of unit Wiener processes defined in lemma 1 and* δ_{21} *denotes the correlation between the components of* ζ_t.

Proof: *see Perron (1989), Park and Phillips (1988), Phillips (1988) and Hansen (1992a).*

For short we denote the integrals involving Brownian motions such as $\int_\lambda^1 W_{2k}(r)\,dr$ by $\int_\lambda^1 W_{2k}$, and the stochastic integrals such as $\int_0^1 W_{2k}(r)\,dW_{11}(r)$ by $\int_0^1 W_{2k}dW_{11}$.

A.1 Proof of Theorem 1

Here we only sketch the proof of Theorem 1 for model E –the most general model considered in the paper– since the proof for the other models follows the same reasoning and are particular cases for it. The OLS estimated residuals of (5) can be computed as $\hat{e}_t = \varepsilon_t - z_t \left(z'z\right)^{-1} z'\varepsilon$, where $z_t = [1,\, DU_t,\, t,\, DT_t^*,\, x_t',\, x_t'DU_t]$ is the $(T \times (2k+4))$-matrix of regressors. Let $P = diag(T^{-1/2}$, $T^{-1/2}$, $T^{-3/2}$, $T^{-3/2}$, T^{-1}, ..., $T^{-1})$ and $A = diag(1,\, 1,\, 1,\, 1,\, \Omega_{22}^{-1/2},\, \Omega_{22}^{-1/2})$ be scaling $((2k+4)$ $\times$ $(2k+4))$-matrices. Hence, the partial sum processes $\hat{S}_t = \sum_{j=1}^{t} \hat{e}_j$ can be computed from:

$$T^{-1/2}\omega_1^{-1}\hat{S}_t = T^{-1/2}\omega_1^{-1}\sum_{j=1}^{t}\varepsilon_j - T^{-1/2}\omega_1^{-1}\sum_{j=1}^{t} z_j PA\left(A'Pz'zPA\right)^{-1}A'Pz'\varepsilon.$$

It is straightforward to see that, at the limit, the cross-product matrix $A'Pz'zPA$ converges in distribution to a (6×6)-symmetric matrix, $A'Pz'zPA \Rightarrow H$, with elements given by: $h_{11} = 1$, $h_{12} = h_{22} = (1-\lambda)$, $h_{13} = 1/2$, $h_{14} = h_{24} = (1-\lambda)^2/2$, $h_{15} = \int_0^1 W_{2k}'$, $h_{16} = h_{25} = h_{26} = \int_\lambda^1 W_{2k}'$, $h_{23} = \left(1-\lambda^2\right)/2$, $h_{33} = 1/3$, $h_{34} = \left(2-3\lambda+\lambda^3\right)/6$, $h_{35} = \int_0^1 rW_{2k}'$, $h_{36} = \int_\lambda^1 rW_{2k}'$, $h_{44} = (1-\lambda)^3/3$, $h_{45} = h_{46} = \int_\lambda^1 (r-\lambda)\, W_{2k}'$, $h_{55} = \int_0^1 W_{2k}W_{2k}'$, $h_{56} = h_{66} = \int_\lambda^1 W_{2k}W_{2k}'$. Moreover, the product between matrix regressors and disturbance term ε converges to the (6×1)-vector J, $Pz'\varepsilon \Rightarrow J$, with elements given by: $J_1 = \omega_1 W_{11}(1)$, $J_2 = \omega_1 \left(W_{11}(1) - W_{11}(\lambda)\right)$, $J_3 = \omega_1\left(\int_0^1 rdW_{11}\right)$, $J_4 = \omega_1\left(\int_\lambda^1 rdW_{11} - \lambda\left(W_{11}(1) - W_{11}(\lambda)\right)\right)$, $J_5 = \omega_1\Omega_{22}^{1/2}\int_0^1 W_{2k}dW_{11}$, $J_6 = \omega_1\Omega_{22}^{1/2}\int_\lambda^1 W_{2k}dW_{11}$. The partial sum process of matrix regressors converges to:

$$T^{-1/2}\sum_{j=1}^{t} z_j PA \;\Rightarrow \begin{cases} K_1 & \text{for } t \le T_b \\ K_2 & \text{for } t > T_b \end{cases},$$

where $K_1 = [r \quad 0 \quad \frac{r^2}{2} \quad 0 \quad \int_0^r W_{2k}' \quad 0]$ and $K_2 = [r \quad (r-\lambda) \quad \frac{r^2}{2} \quad \frac{(r-\lambda)^2}{2} \quad \int_0^r W_{2k}' \quad \int_\lambda^r W_{2k}']$. Using these results, for $t \le T_b$ we have that at the limit the partial sum process converges to:

$$T^{-1/2}\omega_1^{-1}\hat{S}_t \Rightarrow \begin{cases} L_{k,E1}(r,\lambda) & \text{for } t \le T_b \\ L_{k,E2}(r,\lambda) & \text{for } t > T_b \end{cases},$$

where $L_{k,E1}(r,\lambda) = W_{11}(r) - K_1 H^{-1} J$ and $L_{k,E2}(r,\lambda) = W_{11}(r) - K_2 H^{-1} J$. Appealing to the Continuous Mapping Theorem (CMT), we obtain:

$$\begin{aligned} SC_E(\lambda) &= T^{-2}\hat{\omega}_1^{-2}\left[\sum_{t=1}^{T_b}\left(\sum_{j=1}^{t}\hat{e}_j\right)^2 + \sum_{t=T_b+1}^{T}\left(\sum_{j=1}^{t}\hat{e}_j\right)^2\right] \\ &\Rightarrow \int_0^{\lambda} L_{k,E1}(r,\lambda)^2\,dr + \int_{\lambda}^{1} L_{k,E2}(r,\lambda)^2\,dr, \end{aligned}$$

where $\hat{\omega}_1^2$ is a consistent estimate of the long-run variance of $\{e_t\}$. The symmetry of the asymptotic distribution is shown in the companion working paper in Carrion-i-Silvestre and Sansó (2001). Thus, after some algebra manipulations the limiting distribution can be equivalently expressed as

$$\lambda^2 \int_0^1 V_{k,E}^2(b_1)\,db_1 + (1-\lambda)^2 \int_0^1 V_{k,E}^2(b_2)\,db_2,$$

where $V_{k,E}(\cdot)$ are the residuals from the projection of a standard Wiener process onto the subspace generated by the deterministic regressors (properly rescaled) and the $W_{2k}(\cdot)$, and $V_{k,E}(b_1)$ and $V_{k,E}(b_2)$ are independent.

A.2 Proof of Theorem 3

This proof follows similar arguments in Bai (1997), Bai and Perron (1998) and Perron and Zhu (2004). The OLS residuals can be written as:

$$\begin{aligned} \hat{e}_t &= \varepsilon_t - z_t(\lambda)\hat{\beta} + z_t(\tau)\beta \\ &= \varepsilon_t - z_t(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)\varepsilon \\ &\quad - \left(z_t(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)z(\tau) - z_t(\tau)\right)\beta, \end{aligned}$$

where $\tau \in (0,1)$ denotes the true break fraction and $\lambda \in (0,1)$. We can write

$$
\begin{aligned}
SSR(\lambda) - SSR(\tau) &= -\beta' d'(\lambda) z(\lambda) P \left(P z'(\lambda) z(\lambda) P \right)^{-1} P z'(\lambda) d(\lambda) \beta \\
&\quad + \beta' d'(\lambda) d(\lambda) \beta \\
&\quad + 2\varepsilon' z(\lambda) P \left(P z'(\lambda) z(\lambda) P \right)^{-1} P z'(\lambda) d(\lambda) - 2\varepsilon' d(\lambda) \\
&= \beta' d'(\lambda) M d(\lambda) \beta - 2\varepsilon' M d(\lambda) \qquad \text{(A.1)} \\
&\equiv Q(\lambda) - 2G(\lambda)
\end{aligned}
$$

where $d(\lambda) = z(\lambda) - z(\tau)$ and $M = I - z(\lambda) P \left(P z'(\lambda) z(\lambda) P \right)^{-1} P z'(\lambda)$.

Lemma 3 *Given that M is a idempotent matrix we have*

$$
\begin{aligned}
0 &\leq Q(\lambda) \leq \beta' d'(\lambda) d(\lambda) \beta \\
&= \begin{cases} |\lambda - \tau| \, O(T) & \text{for models An and A} \\ |\lambda - \tau| \, O\left(T^3\right) & \text{for models B,C and E} \\ O_p\left(T^2\right) & \text{for model D} \end{cases}
\end{aligned}
$$

and that the probability orders of $Q(\lambda)$ and $\beta' d'(\lambda) d(\lambda) \beta$ are the same. Morevover, for large samples

$$
G(\lambda) = \begin{cases} O_p\left(T^{1/2}\right) & \text{for models An and A} \\ O_p\left(T^{3/2}\right) & \text{for models B, C and E} \\ O_p(T) & \text{for model D} \end{cases} .
$$

Proof. The probability orders are obtained by direct calculations and presented in a summarised way as:

$$
\beta' d'(\lambda) d(\lambda) \beta = \begin{cases} \theta^2 |\lambda - \tau| \, T & \text{for models An and A} \\ \gamma^2 \frac{1}{3} T^3 |\lambda - \tau| \left(\lambda\tau + \lambda^2 + \tau^2 \right) + O_p\left(T^2\right) & \text{for models B, C and E} \\ \beta_2^2 \sum_{[\lambda T]}^{[\tau T]} x_t^2 + o_p\left(T^2\right) & \text{for model D} \end{cases} .
$$

■

Now we have the elements to proof Theorem 3. From (A.1) and Lemma 3

$$T^{-j}\left(SSR\left(\lambda\right)-SSR\left(\tau\right)\right)=T^{-j}Q\left(\lambda\right)+o_{p}\left(1\right). \tag{A.2}$$

where $j=1$ for models An and A, $j=3$ for models B, C and E and, $j=2$ for model D. Next, given the properties of projections,

$$T^{-j}SSR\left(\tilde{\lambda}\right)\leq T^{-j}SSR\left(\tau\right) \tag{A.3}$$

Let us suppose that $\tilde{\lambda}\nrightarrow_{p}\tau$. Then, according to (A.2) for large T we need

$$T^{-j}Q\left(\tilde{\lambda}\right)\leq 0$$

but, $T^{-j}Q\left(\tilde{\lambda}\right)>0$ when $\tilde{\lambda}\neq\tau$. Hence, there is a contradiction with inequality (A.3) so that the only way in which it can be satisfied is $\tilde{\lambda}\rightarrow_{p}\tau$.

To establish the convergence rate we first define the sets

$$V_{\epsilon}=\left\{T_{k}:\left|T_{k}-T_{b}\right|<\epsilon T\right\}$$

for $\epsilon\in(0,1)$ and

$$V_{\epsilon}\left(C\right)=\left\{T_{k}:\left|T_{k}-T_{b}\right|<\epsilon T,\left|T_{k}-T_{b}\right|>C\right\}$$

for $C>0$, so $V_{\epsilon}\left(C\right)\subset V_{\epsilon}$. Note that $SSR\left(\tilde{\lambda}\right)\leq SSR\left(\tau\right)$ with probability 1. Next, note that given results of Lemma

$$P\left(\min_{\lambda T\in V_{\epsilon}(C)}\frac{SSR\left(\lambda\right)-SSR\left(\tau\right)}{\left|\lambda-\tau\right|T}\leq 0\right)<\eta$$

for small $\eta>0$ provided that

$$\frac{SSR\left(\lambda\right)-SSR\left(\tau\right)}{\left|\lambda-\tau\right|T}>0$$

on $V_{\epsilon}\left(C\right)$ with large probability and that the term $Q\left(\lambda\right)$ in (A.1) dominates $G\left(\lambda\right)$ for large samples. This proves the Theorem.

References

Bai, J., 1994. Least Squares Estimation of a Shift in Linear Processes, Journal of Time Series Analysis, 15, 453–472.

____ , 1997. Estimation of a Change Point in Multiple Regression Models, Review of Economics and Statistics, pp. 551–563.

____ and P. Perron, 1998. Estimating and Testing Linear Models with Multiple Structural Changes, Econometrica, 66 1, 47–78.

____ and ____ , 2003. Computation and Analysis of Multiple Structural Change Models, Journal of Applied Econometrics, 18, 1–22.

Banerjee, A., J. J. Dolado, D. F. Hendry and G. W. Smith, 1986. Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence, Oxford Bulletin of Economics and Statistics, 48, 253–277.

Bartley, W. A., J. Lee and M. C. Strazicich, 2001. Testing the Null of Cointegration in the Presence of a Structural Break, Economics Letters, 73 3, 315–323.

Busetti, F., 2002. Testing for (Common) Stochastic Trends in the Presence of Structural Breaks, Journal of Forecasting, 21, 81–105.

Carrion-i-Silvestre, J. L. and A. Sansó, 2001. Testing the Null of Cointegration with Structural Breaks, Department of Econometrics, Statistics and Spanish Economy. University of Barcelona.

Engle, R. F. and C. W. J. Granger, 1987. Cointegration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 251–276.

____ and S. Yoo, Cointegrated Economic Time Series: An Overview with New Results, in R. F. Engle and C. W. J. Granger, eds., Long-Run Economic Relationships, Oxford University Press, 1991.

Gabriel, V., A. Da Silva and L. Nunes, 2003. Instability in Cointegration Regressions: A Brief Review with an Application to Money Demand in Portugal, Applied Economics, 35, 893–900.

Gonzalo, J., 1994. Five Alternative Methods of Estimating Long-Run Equilibrium Relationships, Journal of Econometrics, 60, 203–233.

Granger, C. W. J., 1981. Some Propierties of Time Series Data and their Use in Econometric Model Specification, Journal of Econometrics, 16, 121–130.

Gregory, A. W. and B. E. Hansen, 1996. Residual-based Tests for Cointegration in Models with Regime Shifts, Journal of Econometrics, 70, 99–126.

—— and ——, 1996. Tests for Cointegration in Models with Regime and Trend Shifts, Oxford Bulletin of Economics and Statistics, 58, 3, 555–560.

Hansen, B. E., 1992. Convergence to Stochastic Integrals for Dependent Heterogeneous Processes, Econometric Theory, 8, 489–500.

——, 1992. Testing for Parameter Instability in Regressions with I(1) Processes, Journal of Business & Economic Statistics, 10, 321–335.

Hao, K., 1996. Testing for Structural Change in Cointegrated Regression Models: some Comparisons and Generalisations, Econometric Reviews, 15, 4, 401–429.

Harris, D. and B. Inder, A Test of the Null Hypothesis of Cointegration, in Colin P. Hargreaves, ed., Nonstationary Time Series Analysis and Cointegration, Oxford University Press, 1994. chapter 5, pp. 133–152.

Kurozumi, E., 2002. Testing for Stationarity with a Break, Journal of Econometrics, 108, 63–99.

Kwiatkowski, D., P. C. B. Phillips, P. J. 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.

Lee, J. and M. Strazicich, 2001. Testing the Null of Stationarity in the Presence of One Structural Break, Applied Economics Letters, 8, 377–382.

Leybourne, S. J. and B. P. M. McCabe, 1993. A Simple Test for Cointegration, Oxford Bulletin of Economics and Statistics, 55, 2, 97–103.

Lütkepohl, H., C. Muller and P. Saikkonen, Unit Root Tests for Time Series with a Structural Break When the Break Point is Known, Technical Report SFB. 373 33/99, Institut für Statistik und Ökonometrie. Humboldt University 1999.

McCabe, B. P. M., S. J. Leybourne and Y. Shin, 1997. A Parametric Approach to Testing the Null of Cointegration, Journal of Time Series Analysis, 18, 4, 395–413.

Nyblom, J. and A. C. Harvey, 2000. Tests of Common Stochastic Trends, Econometric Theory, 16, 176–199.

Park, J. Y., Testing for Unit Roots and Cointegration by Variable Addition, in T. B. Fomby and G. F. Rhodes, eds., Advances in Econometrics: Co-Integration, Spurious Regressions and Unit Roots, Jai Press Inc, 1990.

—— and P. C. B. Phillips, 1988. Statistical Inference in Regressions with Integrated Processes: Part 1, Econometric Theory, 4, 468–498.

Perron, P., 1989. The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, Econometrica, 57, 6, 1361–1401.

——, 1990. Testing for a Unit Root in a Time Series with a Changing Mean, Journal of Business & Economic Statistics, 8, 2, 153–162.

—— and X. Zhu, 2004. Structural Breaks with Deterministic and Strochastic Trends, Journal of Econometrics, p. Forthcoming.

Phillips, P. C. B., 1988. Multiple Regression with Integrated Time Series, Contemporary Mathematics, 80, 79–105.

—— and B. Hansen, 1990. Statistical Inference in Instrumental Variables Regression with I(1) Processes, Review of Economic Studies, 57, 99–125.

——— and S. N. Durlauf, 1986. Multiple Time Series Regression with Integrated Processes, Review of Economic Studies, 53, 473–496.

——— and S. Ouliaris, 1990. Asymptotic Properties of Residual Based Tests for Cointegration, Econometrica, 58, 165–193.

Saikkonen, P., 1991. Asymptotically Efficient Estimation of Cointegration Regression, Econometric Theory, 7, 1–21.

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, 1993. A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, Econometrica, 61, 4, 783–820.