Cointegration and Structural Breaks

Josep Lluís Carrion-i-Silvestre[a†]
Andreu Sansó-i-Rosselló[b]
Manuel Artís Ortuño[a]

[a]Anàlisi Quantitativa Regional (AQR) Research Group
Department d'Econometria, Estadística i Economia Espanyola
Universitat de Barcelona

[b]Departament d'Economia i Empresa
Universitat de les Illes Balears

This draft: January 24th, 2002

Abstract

In this paper we propose a LM-Type statistic to test the null of cointegration allowing for the possibility of a structural break, both in the deterministic and the cointegration vector. The test can be used as a complement of the usual non-cointegration tests in order to get stronger evidence of cointegration. We consider both cases, when the break date is known and when it is not. In this last case, three different procedures to estimate the date of the break are analysed. We show that the usual ways to estimate the break date do not produce good results and propose a new procedure that works. Finally, the behaviour of the tests is studied through Monte Carlo experiments.

JEL classification: C12, C22
Keywords: cointegration, structural break, change in regime, KPSS test

1 Introduction

The concept of cointegration has supposed a great contribution in the field of economics giving raise to a huge number of theoretical and applied research. Several methods have been proposed in the econometric literature to test the stationarity of a linear combination of non-stationary time series since Granger (1981) and Engle and Granger (1987) defined the concept of cointegration. Engle and Granger (1987), Phillips and Ouliaris (1990), Stock (1987), Bossaerts (1988), Johansen (1988, 1991, 1992, 1994, 1995) , Ahn and Reinsel (1990), Johansen and Joselius (1990), Perron and Campbell (1993),

† Corresponding author: Departament d'Econometria, Estadística i Economia Espanyola. Universitat de Barcelona. Avd. Diagonal, 690. 08034 Barcelona. Spain. e-mail: carrion@eco.ub.es

Yap and Reinsel (1995), Saikkonen and Luukkonen (1997), Quintos (1998), Stock and Watson (1988), Hansen and Phillips (1990), Engle and Yoo (1991), Phillips (1991), Phillips and Loretan (1991), Saikkonen (1991), Bewley and Yang (1995), Yang and Bewley (1995) and Bierens (1997) are some of these proposals. The main characteristic that share all these procedures is that they 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 be tested should be the null hypothesis of cointegration instead of the null hypothesis of lack of cointegration. Papers like the ones of 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 cointegration property 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 be aware to have in mind that the larger sample period covering the time series the more probability of appearance of 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 to collect the effect of the structural breaks. In these terms, the definition of cointegration under consideration falls into the classification of deterministic and stochastic cointegration of Park (1990). The generalization of Perron's proposal to the cointegration tests of Engle and Granger (1987) and Phillips and Ouliaris (1990) is done by Gregory and Hansen (1996a, 1996b). On the other hand, Hao (1996) generalises 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) deal also with the KPSS test but now including a shift in the mean and in the trend of the deterministic elements of the cointegrating relationship.[1] 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 generalises the previous contributions and develops a test for cointegration around a break-cointegrating relationship. We remark the differences that can be established between these previous contributions and the ones proposed here. We also consider the endogenous selection of the time break, analysing up to three different methods to estimate it.

In section 2 we present the models and test statistics and derive their asymptotic distribution under two assumptions: (i) the stochastic regressors are strictly exogenous and (ii) the break point is assumed to be known. In section 3 we relax the first assumption and present the way to get efficient estimates 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

[1] We came across this last paper when our work was about to be completed.

samples. Finally, section 7 concludes. All the proofs of the Theorems are collected 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. Hereafter, we will refer as the *structural change models* to those models where only the deterministic component is allowed to shift and, following Gregory and Hansen (1996a, 1996b) , as the *change in regime models* to the ones where both the deterministic and stochastic components suffer a change at a particular time (T_b). The data generating process (DGP) is of the form:

$$\begin{aligned} y_t &= \alpha_t + \xi t + x_t'\beta_1 + \varepsilon_t; & (1)\\ x_t &= x_{t-1} + \varsigma_t, & (2)\\ \alpha_t &= f(t) + \alpha_{t-1} + \eta_t, & (3) \end{aligned}$$

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 structural change models of interest in this paper 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 specify that $\sigma_\eta^2 = 0$, under which the model given by (1), (2) and (3) turns out to be:

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

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. It has to be mentioned that 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). Regarding the change in regime models, 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:

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

$i = \{D, E\}$ 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. Notice that under the alternative hypothesis that $\sigma_\eta^2 > 0$, $(y_t, x_t')'$ will not be cointegrated.

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

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

a block diagonal matrix that 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 situation.

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 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 $(\sigma_\eta^2 = 0)$:*

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

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

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 (see expressions (A-5) and (A-6)). 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 these ones. 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 (1996) 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 depends both on k, the number of elements of x_t that involves the model, 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. In the first one the date of the break is assumed to be known and, hence, exogenous with respect to the model. Thus, in some cases, the date of the break can be assumed to be known so that the break fraction does not need to be estimated. This situation might be suitable, for instance, for many German macroeconomic time series for which the reunification process that took place in 1990 had provoked 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 nine values of the break fraction, $\lambda = \{0.1, 0.2, \ldots, 0.8, 0.9\}$, and for up to four stochastic regressors, $k = \{1, 2, 3, 4\}$, are collected in Tables 1 to 3. The observation of these critical values indicates that the asymptotic distributions of the tests show the symmetric behaviour around $\lambda = 0.5$ that has been derived analytically.

At this point we would like to introduce a comment regarding the proposal developed by Bartley et al. (2001). As mentioned above, these authors analyse the specifications that correspond with models A and C for which they compute the corresponding critical value sets. However, there is no mention anywhere about the fact that the asymptotic distribution of the test depends on the number of stochastic regressors that are present in the models. Actually, if we compare their critical values sets with the ones computed here we see that theirs are almost equivalent to the ones computed here for $k = 1$. But as pointed out in Theorem 1, the asymptotic distributions of the tests relay on the number of stochastic regressors that are present in the cointegrating relationship and, hence, the percentage points of the asymptotic distributions depend on k. In fact, it is well known that the more regressors that are included in the static regression the more stationary tend to resemble the residuals of the long-run relationship. This become apparent if we compare the percentage points of the asymptotic distributions collected in Tables 1 to 3 for the different values of k. Notice that, for a given λ and percentage point, the critical values diminishes as k increases. Consequently, a size distortion problem (underrejection of the null hypothesis) is expected when using the critical values provided by Bartley et al. (2001) for those models with $k > 1$.

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 clear enough. This situation gives rise to the

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

Model An									
	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$	$\lambda=0.6$	$\lambda=0.7$	$\lambda=0.8$	$\lambda=0.9$
					$k=1$				
90%	0.1932	0.1583	0.1395	0.1281	0.1256	0.1272	0.1405	0.1603	0.1912
95%	0.2582	0.2087	0.1855	0.1632	0.1553	0.1594	0.1828	0.2148	0.2551
97.5%	0.3367	0.2676	0.2341	0.1991	0.1855	0.1933	0.2306	0.2731	0.3229
99%	0.4546	0.3543	0.2948	0.2503	0.2287	0.2435	0.2940	0.3628	0.4360
					$k=2$				
90%	0.1336	0.1157	0.1079	0.1020	0.1029	0.1033	0.1075	0.1179	0.1342
95%	0.1796	0.1557	0.1400	0.1306	0.1292	0.1297	0.1397	0.1566	0.1815
97.5%	0.2325	0.2007	0.1759	0.1622	0.1557	0.1600	0.1772	0.2014	0.2341
99%	0.3116	0.2631	0.2259	0.2035	0.1903	0.1998	0.2306	0.2643	0.3091
					$k=3$				
90%	0.1007	0.0907	0.0856	0.0847	0.0840	0.0852	0.0853	0.0911	0.1015
95%	0.1319	0.1179	0.1094	0.1063	0.1051	0.1067	0.1081	0.1174	0.1337
97.5%	0.1670	0.1490	0.1338	0.1276	0.1271	0.1301	0.1331	0.1463	0.1707
99%	0.2238	0.1989	0.1773	0.1602	0.1594	0.1624	0.1757	0.1956	0.2330
					$k=4$				
90%	0.0799	0.0738	0.0712	0.0704	0.0706	0.0711	0.0719	0.0731	0.0800
95%	0.1037	0.0924	0.0873	0.0878	0.0874	0.0874	0.0902	0.0927	0.1036
97.5%	0.1304	0.1151	0.1091	0.1073	0.1056	0.1060	0.1094	0.1164	0.1328
99%	0.1754	0.1502	0.1385	0.1365	0.1350	0.1355	0.1398	0.1487	0.1717

Model A									
	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$	$\lambda=0.6$	$\lambda=0.7$	$\lambda=0.8$	$\lambda=0.9$
					$k=1$				
90%	0.0827	0.0736	0.0747	0.0821	0.0840	0.0808	0.0749	0.0736	0.0819
95%	0.1028	0.0885	0.0907	0.1021	0.1060	0.0994	0.0890	0.0893	0.1011
97.5%	0.1228	0.1054	0.1062	0.1229	0.1315	0.1195	0.1035	0.1066	0.1197
99%	0.1537	0.1305	0.1251	0.1508	0.1642	0.1486	0.1219	0.1301	0.1463
					$k=2$				
90%	0.0700	0.0630	0.0650	0.0690	0.0693	0.0676	0.0647	0.0638	0.0696
95%	0.0865	0.0759	0.0774	0.0852	0.0858	0.0841	0.0777	0.0778	0.0855
97.5%	0.1033	0.0891	0.0909	0.1023	0.1037	0.1023	0.0927	0.0918	0.1022
99%	0.1273	0.1095	0.1083	0.1254	0.1348	0.1255	0.1103	0.1119	0.1244
					$k=3$				
90%	0.0594	0.0554	0.0571	0.0581	0.0584	0.0581	0.0566	0.0556	0.0593
95%	0.0728	0.0670	0.0692	0.0712	0.0725	0.0710	0.0677	0.0667	0.0728
97.5%	0.0871	0.0784	0.0803	0.0843	0.0877	0.0851	0.0788	0.0780	0.0873
99%	0.1064	0.0941	0.0971	0.1035	0.1103	0.1044	0.0961	0.0949	0.1074
					$k=4$				
90%	0.0507	0.0490	0.0501	0.0509	0.0510	0.0502	0.0495	0.0489	0.0512
95%	0.0616	0.0588	0.0606	0.0617	0.0621	0.0614	0.0598	0.0592	0.0623
97.5%	0.0729	0.0691	0.0724	0.0728	0.0741	0.0738	0.0702	0.0699	0.0749
99%	0.0898	0.0846	0.0864	0.0886	0.0938	0.0916	0.0852	0.0845	0.0920

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 B and C

Model B

	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$	$\lambda=0.6$	$\lambda=0.7$	$\lambda=0.8$	$\lambda=0.9$
					$k=1$				
90%	0.0842	0.0747	0.0663	0.0614	0.0604	0.0619	0.0670	0.0746	0.0849
95%	0.1059	0.0919	0.0809	0.0730	0.0729	0.0753	0.0824	0.0921	0.1061
97.5%	0.1280	0.1102	0.0973	0.0863	0.0844	0.0883	0.0986	0.1098	0.1289
99%	0.1578	0.1313	0.1197	0.1039	0.1013	0.1086	0.1195	0.1369	0.1580
					$k=2$				
90%	0.0723	0.0632	0.0579	0.0542	0.0533	0.0542	0.0575	0.0631	0.0714
95%	0.0892	0.0775	0.0694	0.0651	0.0639	0.0646	0.0706	0.0771	0.0884
97.5%	0.1069	0.0925	0.0825	0.0761	0.0752	0.0756	0.0842	0.0912	0.1065
99%	0.1314	0.1161	0.1019	0.0940	0.0903	0.0915	0.1021	0.1118	0.1335
					$k=3$				
90%	0.0602	0.0536	0.0507	0.0475	0.0470	0.0479	0.0503	0.0539	0.0593
95%	0.0740	0.0657	0.0613	0.0575	0.0561	0.0571	0.0608	0.0660	0.0728
97.5%	0.0884	0.0784	0.0724	0.0675	0.0663	0.0675	0.0723	0.0782	0.0875
99%	0.1106	0.0969	0.0888	0.0805	0.0788	0.0816	0.0881	0.0950	0.1076
					$k=4$				
90%	0.0523	0.0472	0.0443	0.0429	0.0421	0.0423	0.0443	0.0470	0.0520
95%	0.0638	0.0574	0.0529	0.0511	0.0498	0.0506	0.0531	0.0567	0.0637
97.5%	0.0757	0.0684	0.0626	0.0596	0.0578	0.0592	0.0622	0.0672	0.0758
99%	0.0921	0.0834	0.0756	0.0711	0.0699	0.0712	0.0769	0.0819	0.0928

Model C

	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.3$	$\lambda=0.4$	$\lambda=0.5$	$\lambda=0.6$	$\lambda=0.7$	$\lambda=0.8$	$\lambda=0.9$
					$k=1$				
90%	0.0802	0.0661	0.0559	0.0493	0.0484	0.0493	0.0561	0.0664	0.0802
95%	0.1000	0.0813	0.0673	0.0579	0.0562	0.0581	0.0681	0.0826	0.0994
97.5%	0.1200	0.0973	0.0802	0.0669	0.0645	0.0684	0.0804	0.0987	0.1204
99%	0.1470	0.1208	0.0967	0.0799	0.0746	0.0817	0.0979	0.1182	0.1518
					$k=2$				
90%	0.0667	0.0567	0.0487	0.0450	0.0430	0.0450	0.0488	0.0557	0.0665
95%	0.0828	0.0696	0.0583	0.0530	0.0498	0.0533	0.0592	0.0691	0.0822
97.5%	0.0998	0.0824	0.0687	0.0614	0.0574	0.0609	0.0703	0.0820	0.0995
99%	0.1235	0.1010	0.0843	0.0731	0.0680	0.0720	0.0838	0.1014	0.1229
					$k=3$				
90%	0.0563	0.0493	0.0430	0.0403	0.0393	0.0405	0.0432	0.0488	0.0567
95%	0.0688	0.0602	0.0518	0.0476	0.0463	0.0478	0.0518	0.0595	0.0695
97.5%	0.0819	0.0716	0.0616	0.0551	0.0527	0.0551	0.0606	0.0706	0.0837
99%	0.1023	0.0887	0.0754	0.0657	0.0618	0.0657	0.0745	0.0883	0.1036
					$k=4$				
90%	0.0491	0.0428	0.0389	0.0368	0.0359	0.0367	0.0387	0.0427	0.0491
95%	0.0603	0.0510	0.0464	0.0431	0.0415	0.0430	0.0460	0.0516	0.0600
97.5%	0.0714	0.0614	0.0544	0.0495	0.0474	0.0497	0.0542	0.0615	0.0715
99%	0.0871	0.0748	0.0659	0.0592	0.0559	0.0582	0.0661	0.0745	0.0894

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 3: Asymptotic critical values for the models D and E

Model D	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
					$k = 1$				
90%	0.1908	0.1547	0.1265	0.1098	0.1044	0.1087	0.1276	0.1502	0.1898
95%	0.2560	0.2067	0.1670	0.1395	0.1309	0.1392	0.1682	0.2041	0.2571
97.5%	0.3295	0.2631	0.2098	0.1729	0.1603	0.1724	0.2176	0.2657	0.3341
99%	0.4463	0.3449	0.2699	0.2224	0.1941	0.2145	0.2862	0.3563	0.4449
					$k = 2$				
90%	0.1319	0.1087	0.0885	0.0760	0.0735	0.0765	0.0878	0.1064	0.1351
95%	0.1759	0.1459	0.1163	0.0969	0.0922	0.0988	0.1141	0.1423	0.1810
97.5%	0.2288	0.1873	0.1485	0.1198	0.1123	0.1224	0.1464	0.1853	0.2349
99%	0.3068	0.2510	0.1942	0.1578	0.1419	0.1565	0.1950	0.2482	0.3261
					$k = 3$				
90%	0.0983	0.0803	0.0664	0.0572	0.0542	0.0562	0.0648	0.0793	0.0973
95%	0.1286	0.1049	0.0851	0.0721	0.0672	0.0715	0.0824	0.1037	0.1291
97.5%	0.1638	0.1363	0.1079	0.0883	0.0819	0.0894	0.1043	0.1317	0.1651
99%	0.2307	0.1816	0.1425	0.1145	0.1039	0.1127	0.1367	0.1757	0.2165
					$k = 4$				
90%	0.0772	0.0616	0.0512	0.0451	0.0423	0.0445	0.0507	0.0620	0.0771
95%	0.0981	0.0791	0.0648	0.0548	0.0514	0.0540	0.0646	0.0799	0.0986
97.5%	0.1225	0.1002	0.0806	0.0658	0.0613	0.0660	0.0804	0.1007	0.1230
99%	0.1579	0.1312	0.1048	0.0852	0.0766	0.0837	0.1028	0.1364	0.1623

Model E	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
					$k = 1$				
90%	0.0808	0.0654	0.0538	0.0463	0.0436	0.0462	0.0538	0.0648	0.0801
95%	0.1004	0.0804	0.0659	0.0552	0.0512	0.0551	0.0650	0.0799	0.0981
97.5%	0.1205	0.0974	0.0784	0.0645	0.0587	0.0633	0.0780	0.0959	0.1185
99%	0.1480	0.1223	0.0960	0.0763	0.0681	0.0765	0.0938	0.1196	0.1469
					$k = 2$				
90%	0.0671	0.0540	0.0448	0.0387	0.0363	0.0386	0.0448	0.0536	0.0671
95%	0.0832	0.0661	0.0544	0.0462	0.0423	0.0455	0.0548	0.0668	0.0836
97.5%	0.0994	0.0790	0.0639	0.0534	0.0488	0.0535	0.0658	0.0787	0.1011
99%	0.1218	0.0980	0.0795	0.0641	0.0574	0.0636	0.0815	0.1002	0.1276
					$k = 3$				
90%	0.0561	0.0457	0.0375	0.0323	0.0309	0.0324	0.0377	0.0458	0.0562
95%	0.0696	0.0559	0.0454	0.0379	0.0360	0.0386	0.0456	0.0566	0.0690
97.5%	0.0828	0.0658	0.0542	0.0444	0.0406	0.0448	0.0542	0.0684	0.0837
99%	0.1040	0.0821	0.0660	0.0529	0.0474	0.0541	0.0666	0.0840	0.1054
					$k = 4$				
90%	0.0484	0.0391	0.0326	0.0282	0.0266	0.0280	0.0324	0.0397	0.0484
90%	0.0597	0.0476	0.0393	0.0329	0.0308	0.0333	0.0388	0.0483	0.0590
97.5%	0.0719	0.0572	0.0463	0.0379	0.0353	0.0385	0.0462	0.0571	0.0712
99%	0.0899	0.0703	0.0570	0.0454	0.0411	0.0462	0.0557	0.0707	0.0866

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.

second way of dealing with the dependency of the asymptotic distributions on the nuisance parameter λ. Thus, if there is not a strong knowledge about the position of the structural break, analysts would prefer to consider the date of the break, and hence, the break fraction, as a parameter 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 conditioned to the assumption that x_t is strictly exogenous, although this assumption might be very restrictive in practice. Instead, we may be interested in allowing for correlation between the disturbance terms of (1) and (2) so that the variance matrix of $\psi_t = (\varepsilon_t, \varsigma_t')'$ ought to be specified as:

$$\Sigma = E\left(\psi_t'\psi_t\right) = \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{array}\right],$$

with the long-run variance matrix:

$$\Omega = \lim_{T\to\infty} T^{-1}E\left(\zeta_T\zeta_T'\right) = \left[\begin{array}{cc} \omega_1^2 & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{array}\right],$$

which can be suitable decomposed as $\Omega \equiv \Sigma + \Lambda + \Lambda' \equiv \Delta + \Lambda'$ where:

$$\Lambda = \sum_{j=1}^{\infty} E\left(\psi_t\psi_{t-j}'\right) = \left[\begin{array}{cc} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \Lambda_{22} \end{array}\right].$$

When this is the case, the asymptotic results that have been obtained in the previous section do not longer hold because the estimation of the vector of cointegration is not efficient. However, several methods such the ones of Phillips and Hansen (1990), Saikkonen (1991) and Stock and Watson (1993) can be applied to obtain an efficient estimation of the vector of cointegration and we are going to deal with the first one, although they are asymptotically equivalent. Following Harris and Inder (1994), the steps that have to be given to test the null of cointegration in this framework are:

1. Estimate (4) or (5), depending on the model, and store the estimated residuals $\{\hat{e}_{i,t}\}, i = \{An, A, B, C, D, E\}$;
2. Compute $\hat{\psi}_t = (\hat{e}_t, \Delta x_t')'$ from which is possible to get a consistent estimation of the long-run variance and covariance matrix:

$$\hat{\Omega} = \left[\begin{array}{cc} \hat{\omega}_1^2 & \hat{\omega}_{12} \\ \hat{\omega}_{21} & \hat{\Omega}_{22} \end{array}\right] = T^{-1}\sum_{t=1}^{T}\hat{\psi}_t\hat{\psi}_t' + T^{-1}\sum_{s=1}^{l} w(s,l) \sum_{t=s+1}^{T}\left(\hat{\psi}_{t-s}\hat{\psi}_t' + \hat{\psi}_t\hat{\psi}_{t-s}'\right),$$

$$\hat{\Delta} = \left[\begin{array}{cc} \hat{\delta}_{11} & \hat{\delta}_{12} \\ \hat{\delta}_{21} & \hat{\Delta}_{22} \end{array}\right] = T^{-1}\sum_{s=0}^{l} w(s,l) \sum_{t=s+1}^{T}\hat{\psi}_t\hat{\psi}_{t-s}';$$

3. Transform y_t as $y_t^+ = y_t - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\Delta x_t$ and $\hat{\delta}_{21}$ as $\hat{\delta}_{21}^+ = \hat{\delta}_{21} - \hat{\Delta}_{22}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21}$;[2]

[2] For models D and E, Ω and Δ are functions of the estimates of the previous step.

4. Compute the fully-modified OLS estimation of the cointegration vector from:

$$\hat{\beta}^{+} = (z'z)^{-1}\left(z'y^{+} - e_K T\hat{\delta}_{21}^{+}\right),$$

where z is the $(T \times K)$ matrix of regressors that are present in (4) or (5), $e_K = [0, \iota_m]'$ is a $(K \times 1)$-vector of zeros and ones, and K is the total number of regressors in (4) or (5) and $\iota_m = (1, ..., 1)'$ being a vector which dimension equals the number of stochastic regressors, that is, $m = k$ for models An to C and $m = 2k$ for models D and E;

5. Get the fully-modified residuals $\hat{e}^{+} = y^{+} - z\,\hat{\beta}^{+}$ from which test statistic is computed as:

$$SC_i^{+}(\lambda) = T^{-2}\hat{\omega}_{1.2}^{-2}\sum_{t=1}^{T}\left(S_{i,t}^{+}\right)^2,$$

where $\hat{\omega}_{1.2}^2 = \hat{\omega}_1^2 - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\omega}_{21}$ is a consistent estimation of the long-run variance of $\{\varepsilon_t\}$ conditioned to $\{\varsigma_t\}$ and $S_{i,t}^{+} = \sum_{j=1}^{t}\hat{e}_{i,j}^{+}$, $i = \{An, A, B, C, D, E\}$.

It can be shown that the method of Phillips and Hansen (1990) is equivalent to the estimation of:

$$y^{+} = g_i(t) + x_t'\beta + \left(\varepsilon_t - \hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\varsigma_t\right), \tag{7}$$

that is to say, (7) is a transformation of (4) where $\left(\hat{\omega}_{12}\hat{\Omega}_{22}^{-1}\varsigma_t\right)$ is removed from each side of the equality. The same applies for (5). The asymptotics concerning these 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:*

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

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

Thus, following Gregory and Hansen (1996b) and making a notacional change, if we denote by:

$$\hat{\Omega}^{*} = \begin{bmatrix} \hat{\omega}_1^{*2} & \hat{\omega}_{12}^{*} \\ \hat{\omega}_{21}^{*} & \hat{\Omega}_{22}^{*} \end{bmatrix}$$

the long-run variance and covariance matrix estimation that has been obtained in the previous step, the matrix that has to be used in the correction is:

$$\hat{\Omega} = \begin{bmatrix} \hat{\omega}_1^{*2} & \hat{\omega}_{12}^{*} & (1-\lambda)\,\hat{\omega}_{12}^{*} \\ \hat{\omega}_{21}^{*} & \hat{\Omega}_{22}^{*} & (1-\lambda)\,\hat{\Omega}_{22}^{*} \\ (1-\lambda)\,\hat{\omega}_{21}^{*} & (1-\lambda)\,\hat{\Omega}_{22}^{*} & (1-\lambda)\,\hat{\Omega}_{22}^{*} \end{bmatrix}.$$

The same transformation proceeds for $\hat{\Delta}$.

The proof is given in the Appendix. Theorem 2 states that, after the correction imposed by the method of Phillips and Hansen (1990) is applied, 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 to 3 can also be used to test the null hypothesis of cointegration in this situation.

4 Unknown date of the break

In previous sections we have implicitly assumed that the date of the break was known *a priory*. However, far to be known, the break point has often to be estimated in the empirical applications. There are several procedures in the unit root and cointegration literature that might be applied to our proposal in order to estimate the date of the break. Provided that the test statistic we are dealing with is an LM-type test and that the null hypothesis of stationarity is rejected for large values of the statistic, all the proposals we consider in this paper relay on the use of the minimum functional as a way to identify the date that most favours the null hypothesis of cointegration with a structural break. Thus, if the null is rejected using these procedures then we should conclude that there is strong evidence against cointegration since the conclusion is based on the break point that most favours the null hypothesis. From now on we are going to focus on three different methods to estimate the date of the break. Which of these strategies is the best one is the question that addresses Section 6, where the procedures will compete in the arena. These procedures are described as follows.

The first strategy is based on the proposal of Lee (1996) for the univariate KPSS test with one structural break. Lee applies the minimum functional to the sequence of the test computed for all possible break points and chooses the argument that minimizes such sequence as the estimate of the break point. That is,

$$\hat{\lambda} = \arg\min_{\lambda \in \Lambda} \left[SC_i \left(\lambda \right) \right].$$

Then, he assumes the estimated break point as if it was known and test the null hypothesis using the suitable critical values. Thus, the proposal assumes that there has been a structural break that has shifted the cointegrating relationship and, therefore, that the estimated break point equals the true break point with probability one. Under these assumptions the minimum functional will provide a consistent estimation of the true date of the break. Moreover, the test statistics converge to the same asymptotic distribution that has been derived when the break point is assumed to be known due to the fact that the break fraction is a relevant parameter under the null hypothesis of cointegration, that is to say, the distribution conditioned to the fact that there is a structural break. Hence, the asymptotic critical values collected in Tables 1 to 3 can still be applied. We denote the tests that are computed using this first strategy as $SC_i\left(\hat{\lambda}\right)$ and $SC_i^+\left(\hat{\lambda}\right)$, $i = \{An, A, B, C, D, E\}$.

The second strategy follows, to some extent, the first one, since the estimation of the break point is given by the argument that minimizes the sequence of tests as in the first one. The difference between the two approaches relays on the degree of analyst's uncertainty about the presence of the break. Sometimes analyst's believes about the occurrence of a structural break might be hesitant. This lack of confidence on the presence of one structural break affecting the cointegrating relationship should be taken into account when carrying out the test because such indecision changes the asymptotic distribution of the test. We denote the tests that are computed using this second strategy as $SC_i\left(\tilde{\lambda}\right)$ and $SC_i^+\left(\tilde{\lambda}\right)$, $i = \{An, A, B, C, D, E\}$. The following Theorem collects the limiting distribution of the test statistics for this strategy.

Theorem 3 *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:*

$$SC_i^+\left(\tilde{\lambda}\right) \Rightarrow \inf_{\lambda \in \Lambda} \left[\lambda^2 \int_0^1 V_{k,i}^2(b_1)\, db_1 + (1-\lambda)^2 \int_0^1 V_{k,i}^2(b_2)\, db_2\right],$$

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

The proof of this Theorem, which is outlined in the Appendix, only needs to use the Continuous Mapping Theorem as in Zivot and Andrews (1992). However, if Ω is a block diagonal matrix, so that x_t is strictly exogenous, it follows that:

$$SC_i\left(\tilde{\lambda}\right) \Rightarrow \inf_{\lambda \in \Lambda} \left[\lambda^2 \int_0^1 V_{k,i}^2(b_1)\, db_1 + (1-\lambda)^2 \int_0^1 V_{k,i}^2(b_2)\, db_2\right],$$

$i = \{An, A, B, C, D, E\}$.

Contrary to the first strategy, in this approach the existence of one structural break is not imposed, that is, the analysis is carried out without assuming that there is one structural break. If one believes that this is the suitable situation under which focus the analysis, it will be necessary to apply the critical values that have been computed assuming that there might be one structural break. This second approach was followed in Hao (1996) where critical values for the model An were computed. The asymptotic critical values that correspond with this approach are collected in Table 4.

Finally, the third approach is based on Bai and Perron (1998). These authors use dynamic minimisation techniques to develop an algorithm that estimates the dates of the break through the global minimisation of the sum of squared residuals. This paper shows that the estimates of the break points that are obtained using such algorithm are consistent. Thus, if we assume that there has been a structural break we can select, as an estimate of the break point, the date that minimizes the sequence of the sum of squared residuals. Then, we can use the estimated break point to compute the test statistic. Formally speaking,

$$\ddot{\lambda} = \arg\min_{\lambda \in \Lambda} \left[SSR(\lambda)\right],$$

Table 4: Asymptotic critical values for the $SC\left(\tilde{\lambda}\right)$ tests

	Model An				Model A			
	90%	95%	97.5%	99%	90%	95%	97.5%	99%
$k=1$	0.0594	0.0725	0.0878	0.1075	0.0356	0.0409	0.0465	0.0548
$k=2$	0.0471	0.0564	0.0650	0.0781	0.0311	0.0354	0.0400	0.0471
$k=3$	0.0389	0.0454	0.0524	0.0606	0.0280	0.0321	0.0363	0.0420
$k=4$	0.0334	0.0398	0.0451	0.0526	0.0255	0.0289	0.0323	0.0374
	Model B				Model C			
	90%	95%	97.5%	99%	90%	95%	97.5%	99%
$k=1$	0.0451	0.0548	0.0645	0.0753	0.0287	0.0324	0.0364	0.0420
$k=2$	0.0391	0.0458	0.0520	0.0616	0.0258	0.0292	0.0329	0.0369
$k=3$	0.0340	0.0394	0.0461	0.0539	0.0231	0.0261	0.0292	0.0338
$k=4$	0.0306	0.0352	0.0403	0.0469	0.0214	0.0238	0.0265	0.0301
	Model D				Model E			
	90%	95%	97.5%	99%	90%	95%	97.5%	99%
$k=1$	0.0525	0.0635	0.0770	0.0939	0.0267	0.0309	0.0349	0.0411
$k=2$	0.0380	0.0456	0.0533	0.0655	0.0229	0.0261	0.0293	0.0328
$k=3$	0.0302	0.0356	0.0408	0.0486	0.0197	0.0224	0.0248	0.0280
$k=4$	0.0245	0.0279	0.0323	0.0380	0.0172	0.0194	0.0215	0.0241

Note: k denotes the number of stochastic regressors that are considered in the cointegrating regression. Percentage points of the asymptotic distribution are based on $n = 5,000$ replications using partial sums of $\varepsilon \sim iidN(0,1)$ random variables of $T = 1,000$ observations to approximate the Wiener process.

where $SSR(\lambda)$ denotes the sum of the squared residuals of (4) or (5), depending on the model. We denote the tests that are computed using this strategy as $SC_i\left(\ddot{\lambda}\right)$ and $SC_i^+\left(\ddot{\lambda}\right)$, $i = \{An, A, B, C, D, E\}$. Notice that the first and this third proposals assume that there has been a structural break but they apply the minimum functional on two different sequences of values to obtain the estimation of the break point.

At this point we should introduce a comment regarding the method that use Bartley et al. (2001) to estimate the break point. These authors select the break point as the argument that minimizes the sequence of the Bayesian Information Criterion (BIC) that is obtained when computing the test statistic for all possible break points. Formally speaking,

$$\breve{\lambda} = \arg\min_{\lambda\in\Lambda}\left[BIC(\lambda)\right].$$

Actually, this criteria is equivalent to the third one proposed in this paper. Thus, the use of whichever information criterion is suitable when selecting between different models since they incorporates both the precision of the estimate and a weight for the rule of parsimony of the model. But notice that this is not the problem here, since when computing the BIC criterion for all the possible values of λ, the model does not change either the number of regressors and the number of time periods. Consequently, the penalty for the degrees of freedom lost does not need to be applied. Therefore, the minimization of the BIC information criterion and the minimization of the global sum of the squared residuals are two equivalent procedures.

The consistency of the estimated break point is an opened question to be addressed. The use of the minimum functional on the sequence of tests can only be supported by the fact that it provides a consistent estimate of the date of the break. Theorem 4 analyses this point.

Theorem 4 *Let $\{y_t\}_{t=1}^T$ be generated by (1), (2) and (3), where $\{\zeta_t\}_{t=1}^T$ satisfies the multivariate invariance principle of Phillips and Durlauf (1986). Also let $T_b = \lambda T$ $(0 < \lambda < 1)$ be the estimated break point and $T_b^p = \tau T$ $(0 < \tau < 1)$ the true break point, with $\lambda \neq \tau$. As $T \to \infty$, $T_b \to \infty$, $T_b^p \to \infty$, so that λ and τ remain constant, it can be shown that under the null hypothesis the test statistics are of order:*

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

$i = \{An, A, B, C, D, E\}$.

The proof is collected in the appendix. As can be seen, Theorem 4 shows that the test statistic diverges when there is a misspecification error in the estimation of the date of the break. Thus, the application of the minimum functional will provide with a consistent estimation of the break point. This result constrasts with the one given in Lee (1996) where it is argued that the use of the minimum functionals to compute the asymptotic distribution of the KPSS test with one structural break produces an inconsistent estimation of the date of the break.

5 Consistency

The consistency of the test statistics against the alternative hypothesis of no cointegration is shown through the derivation of the asymptotic distribution of the test statistics under the alternative. The goal is to show that, under the alternative hypothesis, the asymptotic distribution of the test statistics diverges, so they are consistent. We state in the following theorem the limiting distribution of the tests under the alternative hypothesis.

Theorem 5 *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:*

$$(T/l)^{-1} SC_i(\lambda) \Rightarrow \frac{\left[\int_0^\lambda K_{1,i}^2(\lambda, b)\, db + \int_\lambda^1 K_{2,i}^2(\lambda, b)\, db\right]}{K\left[\int_0^\lambda \underline{W}(s)_{1,i}^2\, db + \int_\lambda^1 \underline{W}(s)_{2,i}^2\, db\right]},$$

$i = \{An, A, B, C, D, E\}$ *, where l is the bandwidth of the spectral window used to estimate the long run variance of* ε_t*,* $K_{k,i}(\cdot)$ *and* $\underline{W}(s)_{k,i}$ $(k = 1, 2)$ *are functions of standard Brownian motions and of the break fraction that are shown in the appendix.*

Theorem 5 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 5 are valid independent 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 that specified by Banerjee, Dolado, Hendry and Smith (1986), Engle and Granger (1987), Hansen and Phillips (1990), Gonzalo (1994), Fernández and Peruga (1996) and Gregory and Hansen (1996a), among others, and is given by the triangular system representation of the CI(1, 1) process:

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

This DGP allow us to analyse how the test statistic behaves under different situations. First, under the null of cointegration, $|\rho| < 1$, 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 $\rho = 1$. Third, notice that the variance of $\{e_{z_t}\}$ has been normalised to 1 so that σ^2 can be interpreted as the signal-to-noise ratio. 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 has also to 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 $t > T_b$.

6.1 The break point is assumed to be known

We have considered the parameter space (a_1, a_2, α, θ, β_t, λ, ρ, σ, δ) where a_1 ={0, 1}, $a_2 = -1$, $\alpha = 1$, θ = {0.5, 1}, $\beta_t = (\beta_1, \beta_2)$ = {(2, 1), (2, 4)}, λ ={0.1, 0.25, 0.5, 0.75, 0.9}, ρ ={0, 0.5, 0.9}, σ ={0.25, 0.33, 0.5, 1, 2} and δ ={-0.5, 0, 0.5}. The sample size has been set to $T = 200$ and an amount of $n = 1,000$ replications of the DGP have been carried out for each parametrization. Also, for the

Table 5: Empirical size for test SC_D with strictly exogenous regressors ($a_1 = 0$)

Panel A: strictly exogenous regressors $SC_D(\lambda)$							
		$\theta = 0.5; \beta_t = (2,1)$			$\theta = 1; \beta_t = (2,4)$		
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0	0.25	0.049	0.075	0.140	0.041	0.073	0.126
	0.33	0.051	0.066	0.140	0.059	0.071	0.126
	0.5	0.054	0.065	0.141	0.062	0.070	0.126
	1	0.049	0.089	0.127	0.056	0.068	0.122
	2	0.060	0.086	0.144	0.059	0.071	0.128
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0.5	0.25	0.052	0.075	0.140	0.061	0.086	0.145
	0.33	0.066	0.081	0.144	0.061	0.091	0.155
	0.5	0.057	0.098	0.139	0.064	0.099	0.160
	1	0.066	0.110	0.162	0.083	0.122	0.171
	2	0.084	0.126	0.190	0.091	0.141	0.171
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
-0.5	0.25	0.051	0.090	0.152	0.059	0.093	0.147
	0.33	0.067	0.080	0.149	0.065	0.082	0.176
	0.5	0.070	0.085	0.153	0.056	0.094	0.154
	1	0.093	0.122	0.163	0.077	0.115	0.183
	2	0.094	0.141	0.195	0.098	0.132	0.180
Panel B: strictly exogenous regressors $SC_D^+(\lambda)$							
		$\theta = 0.5; \beta_t = (2,1)$			$\theta = 1; \beta_t = (2,4)$		
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0	0.25	0.044	0.065	0.110	0.040	0.058	0.104
	0.33	0.047	0.055	0.139	0.058	0.059	0.092
	0.5	0.048	0.055	0.111	0.056	0.063	0.103
	1	0.044	0.071	0.110	0.052	0.057	0.109
	2	0.059	0.068	0.115	0.052	0.060	0.103
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0.5	0.25	0.044	0.062	0.117	0.046	0.057	0.119
	0.33	0.043	0.045	0.105	0.040	0.076	0.111
	0.5	0.040	0.060	0.119	0.051	0.068	0.124
	1	0.045	0.041	0.130	0.046	0.062	0.125
	2	0.047	0.047	0.133	0.033	0.049	0.125
δ	σ	0	0.5	0.9	0	0.5	0.9
-0.5	0.25	0.044	0.061	0.101	0.049	0.068	0.111
	0.33	0.045	0.053	0.120	0.053	0.062	0.116
	0.5	0.047	0.055	0.108	0.037	0.055	0.124
	1	0.041	0.063	0.139	0.045	0.066	0.147
	2	0.038	0.048	0.146	0.044	0.047	0.140

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0, \sigma^2); Covar(e_{z_t}, e_{w_t}) = \delta\sigma; T = 200, T_b = [\lambda T]; n =$ 1.000 replications; 5% nominal size

Table 6: Empirical size for test SC_D with non-strictly exogenous regressors ($a_1 = 1$)

		Panel A: non-strictly exogenous regressors $SC_D(\lambda)$					
		$\theta = 0.5; \beta_t = (2,1)$			$\theta = 1; \beta_t = (2,4)$		
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0	0.25	0.966	0.764	0.481	0.967	0.781	0.490
	0.33	0.954	0.696	0.442	0.962	0.691	0.457
	0.5	0.947	0.611	0.373	0.926	0.587	0.364
	1	0.773	0.412	0.287	0.730	0.423	0.299
	2	0.480	0.228	0.210	0.477	0.250	0.231
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0.5	0.25	0.959	0.767	0.516	0.959	0.781	0.513
	0.33	0.962	0.743	0.475	0.969	0.761	0.482
	0.5	0.952	0.617	0.428	0.944	0.644	0.404
	1	0.822	0.364	0.243	0.834	0.412	0.255
	2	0.486	0.176	0.193	0.462	0.175	0.169
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
-0.5	0.25	0.959	0.804	0.543	0.953	0.809	0.559
	0.33	0.959	0.804	0.528	0.966	0.793	0.505
	0.5	0.959	0.778	0.494	0.947	0.745	0.490
	1	0.957	0.705	0.450	0.954	0.720	0.455
	2	0.919	0.626	0.435	0.923	0.645	0.461
		Panel B: non-strictly exogenous regressors $SC_D^+(\lambda)$					
		$\theta = 0.5; \beta_t = (2,1)$			$\theta = 1; \beta_t = (2,4)$		
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0	0.25	0.164	0.112	0.365	0.169	0.216	0.531
	0.33	0.145	0.920	0.341	0.147	0.171	0.467
	0.5	0.118	0.078	0.270	0.118	0.145	0.415
	1	0.091	0.072	0.222	0.092	0.123	0.367
	2	0.056	0.048	0.167	0.069	0.103	0.283
δ	$\sigma \backslash \rho$	0	0.5	0.9	0	0.5	0.9
0.5	0.25	0.215	0.122	0.416	0.334	0.326	0.548
	0.33	0.170	0.120	0.374	0.295	0.285	0.491
	0.5	0.134	0.094	0.309	0.270	0.280	0.454
	1	0.091	0.065	0.207	0.204	0.271	0.405
	2	0.061	0.062	0.185	0.123	0.259	0.320
δ	σ	0	0.5	0.9	0	0.5	0.9
-0.5	0.25	0.230	0.210	0.474	0.378	0.400	0.617
	0.33	0.183	0.152	0.494	0.342	0.347	0.632
	0.5	0.144	0.122	0.408	0.234	0.294	0.582
	1	0.128	0.112	0.380	0.231	0.277	0.571
	2	0.100	0.098	0.373	0.214	0.284	0.617

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0, \sigma^2); Covar(e_{z_t}, e_{w_t}) = \delta\sigma; T = 200, T_b = [\lambda T]; n =$ 1.000 replications; 5% nominal size long-run covariance matrix estimation has been used the quadratic spectral window with automatic bandwidth selection and initial value $l = 4$.

The simulation results for $\lambda = 0.5$ and strictly exogenous regressor ($a_1 = 0$) are summarized in Table 5. Similar results were obtained for the other values of λ so that they have been omitted in order to save space, although the complete set of results are available from the authors upon request. When x_t is strictly exogenous ($a_1 = 0$) and there is not serial correlation of the error in the cointegrating regression –$\rho = 0$– there are no serious problems of size distortion, independently of the break point position, of the magnitude of the structural break and of the covariance between disturbance terms of the model. However, as ρ approaches 1 the test suffer from some size distortion problems. It has to be noticed that the values of size distortion are similar to the power that shows the ADFRC test statistic. As it can be seen in Table 6, 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, 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.

6.2 The break point is estimated

In this subsection we analyse the finite sample performance of the test statistic for model D when the break point is estimated using the three procedures described in Section 4. Because of the time of computation that is needed to carry out the simulation experiment, we have specified the DGP's given by $\theta = \{0.5, 1\}$, $\beta_t = \{(2,1), (2,4)\}$ and $\delta = \rho = 0$ with $a_1 = 0$ –strictly exogenous regressors– and $a_1 = 1$ –non strictly exogenous regressors. As before, the quadratic spectral window with automatic bandwidth selection and $l = 4$ initial lags is used. The sample size has been set equal to $T = 200$ observations and $n = 1,000$ replications has been carried out.

The panel A of Table 7 presents the empirical size for the $SC_D\left(\hat{\lambda}\right)$, $SC_D\left(\tilde{\lambda}\right)$ and $SC_D\left(\ddot{\lambda}\right)$ tests, respectively, when the stochastic regressors are strictly exogenous, whereas the panel B of this table offers the empirical size when the stochastic regressors are not strictly exogenous. Let us first consider the panel A of Table 7. As can be seen, the $SC_D\left(\hat{\lambda}\right)$ test presents a serious size distortion that implies an under rejection of the null hypothesis –it almost never rejects the null hypothesis. The size of $SC_D\left(\tilde{\lambda}\right)$ is also distorted but on the opposite direction. Thus, the second procedure to estimate the date of the break defines a test that over rejects the null hypothesis. This might be understood as the price we have to pay if we are not sure about the presence of a structural break. However, a positive fact is that the size distortion tends to diminish as the magnitude of the structural break increases. Finally, the simulations indicate that the best procedure to test the null hypothesis is the third one since the empirical size of the $SC_D\left(\ddot{\lambda}\right)$ test is close to the nominal size, irrespective of the magnitude of the break's effect.

The presence of non-strictly exogenous regressors in the model changes the properties of the tests in terms of the empirical size. Panel B of Table 7 shows that for some situations the empirical size of the $SC_D^+\left(\hat{\lambda}\right)$ test is close the nominal one, although the prevailing tendency is to under reject the null hypothesis of cointegration. Hence, it would seems that in some situations the presence of non-strictly exogenous regressors drives the empirical size of the test to approach the nominal size. We will turn back on this point below. On the other hand, the size distortion of the $SC_D\left(\tilde{\lambda}\right)$ test get worse when the stochastic regressors are non-strictly exogenous pointing to a huge over rejection of the null hypothesis. Better results are obtained for the $SC_D^+\left(\ddot{\lambda}\right)$ test, although the size distortion appears for the lowest values of the signal-to-ratio and for the break points located at the extremes of the time period. If one wants to establish a comparison among the three tests it would be concluded that the size distortion is mild for the $SC_D^+\left(\ddot{\lambda}\right)$ test when the signal-to-ratio takes high values.

The properties of the test can also be analysed in terms of its ability to estimate the break point. Tables 8 and 9 show the mean and the standard deviation of the estimates of the date of the break when the null hypothesis of cointegration is not rejected for the DGP's with strictly and non-strictly exogenous regressors, respectively. Let us first focus on the results collected in Table 8. If we compare the three tests using these statistical tools we see that the $SC_D\left(\ddot{\lambda}\right)$ test is the best one. Not only the mean of the estimates of the break point tends to be around the true date, but also the accuracy on the estimation of the break point increases with the magnitude of the break and with the signal-to-noise ratio. Notice that for the higher magnitude of the break the mean of the estimates equals the population parameter. The results for the $SC_D\left(\tilde{\lambda}\right)$ test indicate that the estimation of the break point might be understood as reasonably good only for those situations for which the true break point is far from the extremes of the time period. Moreover, the performance of the procedure when estimating the break point does not seem to depend on the magnitude of the structural break and on the signal-to-noise ratio since the results are similar for the two situations that have been considered. Similar results are obtained for the $SC_D\left(\hat{\lambda}\right)$ test although the precision on the estimation of the break point is the lowest in all situations. The consideration of non-strictly exogenous regressors damages the accuracy of the procedures when estimating the date of the break. As before, the best results are obtained for the $SC_D^+\left(\ddot{\lambda}\right)$ test for those situations where the size distortion is mild. Thus, it can seen that the mean of the estimates of the break point collapses around the true date when the signal-to-noise ratio is high and when the magnitude of the break is also high. The accuracy on the estimation also follows this pattern. Finally, we showed that the empirical size of the $SC_D^+\left(\hat{\lambda}\right)$ test equalled the nominal size for some situations. However, we wish to emphasize that the good results in terms of empirical size for the $SC_D^+\left(\hat{\lambda}\right)$ test does not correspond with the performance showed by this test when estimating the break

Table 7: Empirical size for the SC_D test

Panel A: strictly exogenous regressors

$\theta = 0.5; \beta_t = (2,1); \delta = 0; a_1 = 0$

	$SC_D(\hat{\lambda})$					$SC_D(\tilde{\lambda})$					$SC_D(\ddot{\lambda})$				
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	0.01	0.00	0.00	0.01	0.01	0.24	0.20	0.15	0.21	0.19	0.07	0.04	0.05	0.06	0.05
0.33	0.00	0.00	0.00	0.00	0.00	0.22	0.18	0.15	0.19	0.16	0.05	0.04	0.06	0.07	0.05
0.5	0.01	0.00	0.00	0.00	0.00	0.19	0.18	0.13	0.18	0.12	0.06	0.05	0.05	0.07	0.05
1	0.00	0.00	0.00	0.00	0.00	0.16	0.12	0.13	0.15	0.08	0.06	0.04	0.05	0.06	0.05
2	0.00	0.00	0.00	0.00	0.00	0.08	0.12	0.13	0.13	0.06	0.06	0.04	0.05	0.06	0.05

$\theta = 1; \beta_t = (2,4); \delta = 0; a_1 = 0$

	$SC_D(\hat{\lambda})$					$SC_D(\tilde{\lambda})$					$SC_D(\ddot{\lambda})$				
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	0.00	0.00	0.00	0.00	0.00	0.05	0.08	0.11	0.12	0.03	0.06	0.05	0.05	0.07	0.05
0.33	0.00	0.00	0.00	0.00	0.00	0.05	0.09	0.12	0.11	0.02	0.05	0.04	0.06	0.08	0.06
0.5	0.00	0.00	0.00	0.00	0.00	0.03	0.10	0.12	0.11	0.01	0.07	0.04	0.05	0.05	0.05
1	0.00	0.00	0.01	0.00	0.00	0.02	0.10	0.13	0.10	0.01	0.06	0.04	0.05	0.06	0.05
2	0.00	0.00	0.01	0.00	0.00	0.01	0.10	0.15	0.11	0.00	0.05	0.04	0.05	0.06	0.05

Panel B: non-strictly exogenous regressors

$\theta = 0.5; \beta_t = (2,1); \delta = 0; a_1 = 1$

	$SC_D^+(\hat{\lambda})$					$SC_D^+(\tilde{\lambda})$					$SC_D^+(\ddot{\lambda})$				
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	0.02	0.03	0.05	0.04	0.04	0.41	0.46	0.43	0.46	0.55	0.41	0.22	0.11	0.22	0.32
0.33	0.02	0.02	0.02	0.04	0.04	0.36	0.40	0.33	0.47	0.48	0.31	0.18	0.08	0.20	0.29
0.5	0.01	0.02	0.02	0.02	0.02	0.37	0.35	0.29	0.39	0.44	0.23	0.14	0.07	0.15	0.20
1	0.01	0.01	0.02	0.01	0.01	0.32	0.30	0.23	0.33	0.35	0.17	0.10	0.08	0.10	0.12
2	0.01	0.01	0.01	0.01	0.01	0.29	0.27	0.19	0.26	0.25	0.11	0.08	0.07	0.09	0.08

$\theta = 1; \beta_t = (2,4); \delta = 0; a_1 = 1$

	$SC_D^+(\hat{\lambda})$					$SC_D^+(\tilde{\lambda})$					$SC_D^+(\ddot{\lambda})$				
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	0.02	0.03	0.04	0.03	0.05	0.55	0.47	0.41	0.53	0.56	0.41	0.25	0.15	0.25	0.31
0.33	0.01	0.02	0.03	0.02	0.03	0.47	0.43	0.36	0.46	0.51	0.33	0.21	0.14	0.20	0.26
0.5	0.01	0.01	0.01	0.01	0.01	0.41	0.35	0.32	0.37	0.44	0.27	0.14	0.14	0.15	0.20
1	0.00	0.01	0.01	0.00	0.01	0.26	0.26	0.25	0.26	0.30	0.12	0.07	0.10	0.12	0.11
2	0.01	0.00	0.00	0.00	0.00	0.18	0.21	0.18	0.18	0.15	0.11	0.07	0.07	0.09	0.08

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0,\sigma^2); Covar(e_{z_t}, e_{w_t}) = \delta\sigma; T = 100, T_b = [\lambda T]; n = 2.000$ replications; 5% nominal size

Table 8: Breakpoint estimation for the SC_D test with strictly exogenous regressors

							$\theta = 0.5; \beta_t = (2,1); \delta = 0; a_1 = 0$								
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	45	56	100	145	165	45	54	100	146	167	24	50	100	150	179
	(37.61)	(18.43)	(9.57)	(16.35)	(24.49)	(37.97)	(15.67)	(8.93)	(15.10)	(24.06)	(20.17)	(4.90)	(3.28)	(3.36)	(10.32)
0.33	40	54	100	146	166	39	53	100	147	168	21	50	100	150	179
	(33.61)	(15.12)	(8.87)	(15.48)	(23.63)	(33.26)	(14.28)	(8.40)	(12.87)	(22.19)	(12.30)	(3.85)	(2.91)	(2.14)	(8.97)
0.5	40	55	100	146	167	39	54	100	147	168	21	50	100	150	180
	(33.22)	(15.79)	(9.48)	(12.84)	(23.48)	(33.88)	(13.89)	(8.72)	(11.50)	(22.99)	(8.30)	(2.65)	(1.98)	(2.01)	(1.73)
1	36	54	99	146	169	36	54	99	147	170	20	50	100	150	180
	(28.03)	(12.88)	(11.13)	(13.06)	(20.14)	(28.47)	(11.28)	(8.45)	(10.67)	(19.17)	(3.81)	(1.25)	(1.42)	(1.49)	(1.70)
2	36	55	100	146	169	35	55	100	146	170	20	50	100	150	180
	(26.21)	(14.69)	(10.89)	(12.74)	(19.66)	(25.23)	(14.24)	(9.48)	(11.45)	(18.25)	(5.10)	(1.05)	(0.92)	(0.73)	(0.79)

							$\theta = 1; \beta_t = (2,4); \delta = 0; a_1 = 0$								
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	33	55	101	146	171	32	55	100	146	172	20	50	100	150	180
	(21.66)	(15.03)	(9.47)	(12.19)	(15.42)	(21.56)	(14.87)	(9.27)	(11.73)	(15.01)	(0.99)	(0.80)	(0.50)	(0.57)	(0.62)
0.33	33	56	100	145	171	33	56	100	145	172	20	50	100	150	180
	(21.86)	(15.94)	(11.57)	(13.31)	(16.46)	(21.78)	(15.77)	(11.21)	(12.04)	(16.00)	(0.85)	(0.66)	(0.44)	(0.40)	(0.44)
0.5	31	55	100	146	173	31	55	100	146	173	20	50	100	150	180
	(19.85)	(13.01)	(9.72)	(11.18)	(13.38)	(19.73)	(12.21)	(9.60)	(11.17)	(12.45)	(0.66)	(0.49)	(0.35)	(0.35)	(0.35)
1	29	54	100	146	174	29	54	100	147	174	20	50	100	150	180
	(16.84)	(12.40)	(10.78)	(10.69)	(10.85)	(16.70)	(12.11)	(10.31)	(10.33)	(10.83)	(0.40)	(0.25)	(0.21)	(0.33)	(0.18)
2	27	54	100	146	175	26	54	100	147	175	20	50	100	150	180
	(10.37)	(13.19)	(9.79)	(11.01)	(8.56)	(10.31)	(13.08)	(9.64)	(10.40)	(8.56)	(0.24)	(0.24)	(0.17)	(0.19)	(0.19)

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0, \sigma^2); Covar(e_{z_t}, e_{w_t}) = \delta\sigma; T = 100, T_b = [\lambda T]; n = 2.000$ replications; 5% nominal size

Table 9: Breakpoint estimation for the SC_D^+ test with non-strictly exogenous regressors

								$\theta = 0.5; \beta_t = (2,1); \delta = 0; a_1 = 1$							
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	81	73	103	131	137	81	75	103	128	127	71	80	104	125	126
	(44.89)	(36.07)	(21.15)	(33.59)	(45.77)	(44.14)	(36.53)	(21.40)	(36.62)	(49.46)	(51.51)	(40.98)	(33.99)	(43.08)	(54.38)
0.33	78	70	104	134	140	79	71	105	134	135	70	75	107	127	136
	(46.65)	(32.62)	(20.89)	(30.94)	(42.98)	(45.40)	(31.91)	(21.02)	(32.56)	(45.95)	(49.34)	(37.21)	(31.50)	(42.77)	(50.47)
0.5	68	67	103	140	149	69	67	103	141	147	58	71	108	137	144
	(45.28)	(30.05)	(17.93)	(24.82)	(39.87)	(45.47)	(29.71)	(17.81)	(24.48)	(41.88)	(45.40)	(33.43)	(25.79)	(37.26)	(49.75)
1	58	62	102	143	155	58	62	102	144	155	42	62	105	145	157
	(43.76)	(26.60)	(14.13)	(20.40)	(35.75)	(43.80)	(25.47)	(13.48)	(19.69)	(37.61)	(37.20)	(25.22)	(16.87)	(27.78)	(43.92)
2	49	59	100	145	160	49	57	101	146	161	31	55	103	150	169
	(38.77)	(21.15)	(12.02)	(16.95)	(32.56)	(39.78)	(18.65)	(11.54)	(15.39)	(33.73)	(25.41)	(13.29)	(10.75)	(14.93)	(34.60)

								$\theta = 1; \beta_t = (2,4); \delta = 0; a_1 = 1$							
$\sigma^2 \backslash \lambda$	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
0.25	64	66	100	132	146	64	65	102	133	147	44	65	108	138	144
	(43.12)	(30.71)	(22.96)	(34.33)	(40.98)	(41.06)	(27.94)	(20.97)	(34.27)	(42.51)	(34.28)	(26.68)	(22.53)	(35.59)	(47.07)
0.33	60	66	100	136	146	60	64	102	137	146	37	63	107	143	151
	(41.68)	(30.63)	(19.99)	(29.02)	(40.15)	(40.96)	(25.60)	(19.03)	(28.98)	(41.99)	(27.02)	(24.11)	(16.55)	(30.19)	(43.55)
0.5	54	63	99	138	151	55	61	101	139	153	30	57	104	148	160
	(38.36)	(28.69)	(18.10)	(26.30)	(37.94)	(39.00)	(23.40)	(15.80)	(26.34)	(36.28)	(17.67)	(15.88)	(12.92)	(20.43)	(38.45)
1	49	61	101	141	157	48	60	101	142	158	25	53	102	151	173
	(36.51)	(26.86)	(17.80)	(21.05)	(32.84)	(35.62)	(25.09)	(14.85)	(19.86)	(31.87)	(11.15)	(7.67)	(6.95)	(8.74)	(26.98)
2	46	59	99	143	161	45	59	100	143	161	22	51	101	151	178
	(36.42)	(24.65)	(16.05)	(18.64)	(27.09)	(34.64)	(23.11)	(15.15)	(18.35)	(27.13)	(4.74)	(3.22)	(3.36)	(4.14)	(13.87)

DGP: $y_t - g_D(t) - \beta_t x_t = z_t; z_t = \rho z_{t-1} + e_{z_t}; a_1 y_t - a_2 x_t = w_t; w_t = w_{t-1} + e_{w_t}; g_D(t) = \alpha + \theta DU_t; e_{z_t} \sim iidN(0,1); e_{w_t} \sim iidN(0,\sigma^2); Covar(e_{z_t}, e_{w_t}) = \delta\sigma; T = 100, T_b = [\lambda T]; n = 2.000$ replications; 5% nominal size date, since the mean of the estimates is far from the true date of the break for these cases. This contradiction indicates that the use of the $SC_D^+\left(\hat{\lambda}\right)$ test is inadvisable in applied research.

Therefore, the main conclusion that arise from these simulation experiments is that the best procedure to perform the tests when the date of the break is unknown is the third one, that is, the one that defines the $SC_D\left(\ddot{\lambda}\right)$ test and which is based on the minimization of the sequence of the sum of squared residuals. It has been shown that the $SC_D\left(\hat{\lambda}\right)$ test, which picks out the argument that minimizes the sequence of tests and applies the critical values that were computed assuming the date of the break as known, almost never rejects the null hypothesis. Finally, the size distortion of the $SC_D\left(\tilde{\lambda}\right)$ test might be due to the uncertainty on the presence of a structural break.

7 Conclusions

This paper has focused on testing the null hypothesis of cointegration allowing for one structural break. Since previous proposals in the literature have only considered structural breaks in cointegration tests based on the unit root test, our approach can be understood as a complementary way to ensure the presence of cointegration around a break-cointegrating relationship. In order to provide analysts with a complete tool set to test the null of cointegration with one structural break, the paper has focused on six different possibilities of structural break. The test statistic that is 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. As shown in the paper, this test is not invariant to the presence of a structural break affecting the deterministic and/or the stochastic components since the asymptotic distribution depends on the specification of the model under the null hypothesis and on a nuisance parameter, the break fraction parameter λ. It has also been shown that the asymptotic distributions of the tests are symmetric around $\lambda = 0.5$, that is, the middle point of the range allowed for this parameter. As a consequence, a different set of critical values has to be used when a break is suspected to be occurred and we want to test the null hypothesis of cointegration.

The computation of these critical values and, consequently, the application of the test, is conditioned on the assumption that can be made about the value of λ. Thus, a first set of critical values for which the date of the structural break is assumed to be known are offered in the paper. However, for those situations for which practitioners are not confident on the date of the break we recommend to proceed to the estimation of the date of the structural break using three different procedures, all of them based on the application of the minimum functional. The use of such functional defines a set of new tests with asymptotic distributions free of any nuisance parameter. The paper also offers the critical values for this tests. The consistency of the procedures when estimating the break date follows after it has been proved that the test diverges when there is a misspecification error in the date of the break.

Having defined the general framework, the paper has accounted for two different situations depending on the properties of x_t, the set of stochastic regressors. First, the analysis has focused on the assumption that x_t is strictly exogenous. However, when this is not the fact, OLS estimation of the cointegrating vector is not efficient. Hence, the paper has also focused on the situation where x_t is no longer strictly exogenous and has outlined the corrections that have to be done to get an efficient estimation of the cointegrating vector. The correction is given by the fully-modified OLS estimation procedure of Phillips and Hansen (1990).

The simulations that have been conducted to assess the finite sample size of the test suggest that there might be some size distortion problems when x_t is no longer strictly exogenous. Notwithstanding, for those specifications with strictly exogenous regressors the empirical size of the test approaches the nominal one. This might be understood as the general conclusion. These results suggest that further research on alternative methods to estimate the cointegrating vector when x_t is no longer strictly exogenous. But the Monte Carlo experiments have given rise to some important facts. The first one is that among the different procedures that have been proposed in the literature in order to estimate the date of the break, the one that, on the first stage, estimates the date of the break through the minimization of the sequence of the sum of squared residuals and, on the second stage, performs the test assuming such estimate as the true break point is the best one. As a consequence, the procedures that estimates the break point minimising the sequence of tests should not be applied in applied research provided the size distortion that induce.

Appendix A. Proofs of the Theorems

In this appendix we carry out the proof of the Theorems that have been proposed along the paper. Some intermediate results are needed in order to proceed to such proofs. Throughout, we use the following lemmas.

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*

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

as the covariance matrix.

Proof: *see Herrndorf (1984) and 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:*

$$\begin{aligned}
T^{-3/2}\sum_{t=1}^T x_t &\Rightarrow \Omega_{22}^{1/2}\int_0^1 W_{2k}(r)\,dr; \\
T^{-3/2}\sum_{t=1}^T x_t DU_t &\Rightarrow \Omega_{22}^{1/2}\int_\lambda^1 W_{2k}(r)\,dr; \\
T^{-1}\sum_{t=1}^T x_t\varepsilon_t &\Rightarrow \omega_1\Omega_{22}^{1/2}\int_0^1 W_{2k}(r)\,dW_{11}(r) + \delta_{21}; \\
T^{-2}\sum_{t=1}^T x_t x_t' &\Rightarrow \Omega_{22}^{-1/2}\int_0^1 W_{2k}(r)\,W_{2k}'(r)\,dr\;\Omega_{22}^{-1/2},
\end{aligned}$$

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 like $\int_\lambda^1 W_{2k}(r)\,dr$ by $\int_\lambda^1 W_{2k}$, and the stochastic integrals like $\int_0^1 W_{2k}(r)\,dW_{11}(r)$ by $\int_0^1 W_{2k}dW_{11}$.

Lemma 3 *Let $\lambda = T_b^*/T$ be the break fraction at the estimated break point and $\tau = T_b/T$ the break fraction associated to the true break point. Thus, if $\lambda > \tau$:*
(a) $\sum_{t=1}^{T} (DU_t(\lambda) - DU_t(\tau)) = (-\lambda + \tau) T$
(b) $\sum_{t=1}^{T} (DT_t^*(\lambda) - DT_t^*(\tau)) = (-1 + \lambda)(\lambda - \tau) T^2$
(c) $\sum_{t=1}^{T} (x_t' DU_t(\lambda)(DU_t(\lambda) - DU_t(\tau))) = 0$
(d) $\sum_{t=1}^{T} (DU_t(\lambda)(DU_t(\lambda) - DU_t(\tau))) = 0$
(e) $\sum_{t=1}^{T} (DU_t(\lambda)(DT_t^*(\lambda) - DT_t^*(\tau))) = (-1 + \lambda)(\lambda - \tau) T^2$
(f) $\sum_{t=1}^{T} (DU_t(\lambda)(x_t'(DU_t(\lambda) - DU_t(\tau)))) = 0$
(g) $\sum_{t=1}^{T} (t(DU_t(\lambda) - DU_t(\tau))) = -\frac{1}{2} T(\lambda - \tau)(\lambda T + 1 + \tau T)$
(h) $\sum_{t=1}^{T} (t(DT_t^*(\lambda) - DT_t^*(\tau))) = -\frac{1}{2} T^2 (\lambda - \tau)(T + 1)$
(i) $\sum_{t=1}^{T} (t x_t'(DU_t(\lambda) - DU_t(\tau))) = -\sum_{t=T\tau+1}^{T\lambda} t x_t'$
(j) $\sum_{t=1}^{T} (DT_t^*(\lambda)(DU_t(\lambda) - DU_t(\tau))) = 0$
(k) $\sum_{t=1}^{T} (DT_t^*(\lambda)(DT_t^*(\lambda) - DT_t^*(\tau))) = -\frac{1}{2} T^2 (-1 + \lambda)(\lambda - \tau)(\lambda T - T - 1)$
(l) $\sum_{t=1}^{T} (DT_t^*(\lambda) x_t'(DU_t(\lambda) - DU_t(\tau))) = 0$
(m) $\sum_{t=1}^{T} x_t (DU_t(\lambda) - DU_t(\tau)) = -\sum_{t=\tau T+1}^{T\lambda} x_t$
(n) $\sum_{t=1}^{T} x_t (DT_t^*(\lambda) - DT_t^*(\tau)) = -\sum_{t=\tau T+1}^{T\lambda} t x_t - T\lambda \sum_{t=T\lambda+1}^{T} x_t + \tau T \sum_{t=T\tau+1}^{T} x_t$
(o) $\sum_{t=1}^{T} (x_t x_t'(DU_t(\lambda) - DU_t(\tau))) = -\sum_{t=\tau T+1}^{T\lambda} x_t x_t'$
(p) $\sum_{t=1}^{T} x_t DU_t(\lambda)(DU_t(\lambda) - DU_t(\tau)) = 0$
(q) $\sum_{t=1}^{T} (x_t x_t' DU_t(\lambda)(DU_t(\lambda) - DU_t(\tau))) = 0$

Proof: *These results were obtained by direct calculations and we are going to sketch the steps that were followed in order to obtain them. Thus, it can be shown that these statements are computed as:*

(a) $\sum_{t=1}^{T} (DU_t(\lambda) - DU_t(\tau)) = (1 - \lambda) T - (1 - \tau) T = (-\lambda + \tau) T$

(b) $\sum_{t=1}^{T} (DT_t^*(\lambda) - DT_t^*(\tau)) = \sum_{t=\lambda T+1}^{T} ((t - \lambda T) - (t - \tau T)) = (-1 + \lambda)(\lambda - \tau) T^2$;

(c), (d) and (f) $\sum_{t=1}^{T} (DU_t(\lambda)(DU_t(\lambda) - DU_t(\tau))) = \sum_{t=1}^{T} DU_t(\lambda) - \sum_{t=1}^{T} DU_t(\lambda) = 0$;

(e)

$$\begin{aligned}\sum_{t=1}^{T} (DU_t(\lambda)(DT_t^*(\lambda) - DT_t^*(\tau))) &= \sum_{t=\lambda T+1}^{T} ((t - \lambda T) - (t - \tau T)) \\ &= T^2 (-1 + \lambda)(\lambda - \tau);\end{aligned}$$

(g)

$$\begin{aligned}\sum_{t=1}^{T} (t(DU_t(\lambda) - DU_t(\tau))) &= (1/2)(T(T+1) - \lambda T(\lambda T + 1)) \\ &\quad - (1/2)(T(T+1) - \tau T(\tau T + 1)) \\ &= -\frac{1}{2} T(\lambda - \tau)(\lambda T + 1 + \tau T);\end{aligned}$$

(h)

$$\begin{aligned}\sum_{t=1}^{T}\left(t\left(DT_t^*\left(\lambda\right)-DT_t^*\left(\tau\right)\right)\right) &= \sum_{t=1}^{T}\left(t\left(\left(t-\lambda T\right)-\left(t-\tau T\right)\right)\right)\\ &= -\frac{1}{2}T^2\left(\lambda-\tau\right)\left(T+1\right);\end{aligned}$$

(i) $\sum_{t=1}^{T}\left(tx_t'\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{T}tx_t'-\sum_{t=T\tau+1}^{T}tx_t'=-\sum_{t=T\tau+1}^{T\lambda}tx_t'$;

(j) and (l) $\sum_{t=1}^{T}\left(DT_t^*\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=1}^{T}DT_t^*\left(\lambda\right)-\sum_{t=1}^{T}DT_t^*\left(\lambda\right)=0$;

(k)

$$\begin{aligned}\sum_{t=1}^{T}\left(DT_t^*\left(\lambda\right)\left(DT_t^*\left(\lambda\right)-DT_t^*\left(\tau\right)\right)\right) &= \sum_{t=T\lambda+1}^{T}\Big(\left(t-\lambda T\right)^2-\left(t-\lambda T\right)\left(t-\tau T\right)\Big)\\ &= -\frac{1}{2}T^2\left(-1+\lambda\right)\left(\lambda-\tau\right)\left(\lambda T-T-1\right);\end{aligned}$$

(m) $\sum_{t=1}^{T}x_t\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)=\sum_{t=T\lambda+1}^{T}x_t-\sum_{t=\tau T+1}^{T}x_t=-\sum_{t=\tau T+1}^{T\lambda}x_t$

(n)

$$\begin{aligned}\sum_{t=1}^{T}x_t\left(DT_t^*\left(\lambda\right)-DT_t^*\left(\tau\right)\right) &= \sum_{t=T\lambda+1}^{T}x_t\left(t-\lambda T\right)-\sum_{t=T\tau+1}^{T}x_t\left(t-\tau T\right)\\ &= -\sum_{t=\tau T+1}^{T\lambda}tx_t-T\lambda\sum_{t=T\lambda+1}^{T}x_t+\tau T\sum_{t=T\tau+1}^{T}x_t;\end{aligned}$$

(o) $\sum_{t=1}^{T}\left(x_tx_t'\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{T}x_tx_t'-\sum_{t=T\tau+1}^{T}x_tx_t'=-\sum_{t=\tau T+1}^{T\lambda}x_tx_t'$;

(p) $\sum_{t=1}^{T}x_tDU_t\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)=\sum_{t=1}^{T}x_t\left(DU_t\left(\lambda\right)-DU_t\left(\lambda\right)\right)=0$;

(q) $\sum_{t=1}^{T}\left(x_tx_t'DU_t\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=1}^{T}x_tx_t'\left(DU_t\left(\lambda\right)-DU_t\left(\lambda\right)\right)=0$.

Lemma 3 has taken into account the situation in which $\lambda>\tau$, although the reverse situation in which $\lambda<\tau$ can also be encountered. We have only collected in Lemma 4 the differences that arise when we reverse the inequality. In order to make easy comparisons, the numeration of the statements presented in Lemma 4 follows the one in Lemma 3.

Lemma 4 *Let $\lambda=T_b^*/T$ be the break fraction at the estimated break point and $\tau=T_b/T$ the break fraction associated to the true break point. Thus, if $\lambda<\tau$:*
(c) $\sum_{t=1}^{T}\left(x_t'DU_t\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{\tau T}x_t'$
(d) $\sum_{t=1}^{T}\left(DU_t\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\left(-\lambda+\tau\right)T$
(f) $\sum_{t=1}^{T}\left(DU_t\left(\lambda\right)\left(x_t'\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)\right)=\sum_{t=T\lambda+1}^{\tau T}x_t'$

(i) $\sum_{t=1}^{T}\left(tx_t'\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{T\tau}tx_t'$
(j) $\sum_{t=1}^{T}\left(DT_t^*\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{\tau T}DT_t^*\left(\lambda\right)$
(l) $\sum_{t=1}^{T}\left(DT_t^*\left(\lambda\right)x_t'\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{\tau T}DT_t^*\left(\lambda\right)x_t'$
(m) $\sum_{t=1}^{T}x_t\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)=\sum_{t=T\lambda+1}^{\tau T}x_t$
(n) $\sum_{t=1}^{T}x_t\left(DT_t^*\left(\lambda\right)-DT_t^*\left(\tau\right)\right)=\sum_{t=T\lambda+1}^{\tau T}tx_t-T\lambda\sum_{t=T\lambda+1}^{T}x_t+\tau T\sum_{t=T\tau+1}^{T}x_t$
(o) $\sum_{t=1}^{T}\left(x_tx_t'\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{\tau T}x_tx_t'$
(p) $\sum_{t=1}^{T}x_tDU_t\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)=\sum_{t=T\lambda+1}^{\tau T}x_t$
(q) $\sum_{t=1}^{T}\left(x_tx_t'DU_t\left(\lambda\right)\left(DU_t\left(\lambda\right)-DU_t\left(\tau\right)\right)\right)=\sum_{t=T\lambda+1}^{\tau T}x_tx_t'$

Proof: *These results were obtained by direct calculations following the steps sketched in Lemma 3. Consequently, the proof is omitted in order to save space.*

Lemma 5 *Let us denote $0<r<\lambda$ as the index that defines the standard Brownian motion. Thus, by the properties of the Brownian motions it can be shown that:*
(a) $W\left(r\right)=\sqrt{\lambda}W\left(b_1\right),$
(b) $\int_0^{\lambda}W_{2k}'\left(s\right)ds=\lambda^{3/2}\int_0^1W_{2k}'\left(s_1\right)ds_1,$
(c) $\int_0^{\lambda}sW_{2k}'\left(s\right)ds=\lambda^{5/2}\int_0^1s_1W_{2k}'\left(s_1\right)ds_1,$
(d) $\int_0^{\lambda}W_{2k}\left(s\right)W_{2k}'\left(s\right)ds=\lambda^2\int_0^1W_{2k}\left(s_1\right)W_{2k}'\left(s_1\right)ds_1,$
(e) $\int_0^{\lambda}sdW_{11}\left(s\right)=\lambda^{3/2}\int_0^1s_1dW_{11}\left(s_1\right),$
(f) $\int_0^{\lambda}W_{2k}\left(s\right)dW_{11}\left(s\right)=\lambda\int_0^1W_{2k}\left(s_1\right)dW_{11}\left(s_1\right),$
with $b_1=r/\lambda$, $0<b_1<1$.

Proof: *The first statement follows from the properties of the Brownian motions. Thus, for the statement (a):*

$$W\left(r\right)=\sqrt{\lambda}W\left(\frac{r}{\lambda}\right)=\sqrt{\lambda}W\left(b_1\right),$$

for (b):

$$\begin{aligned}\int_0^{\lambda}W_{2k}'\left(s\right)ds &= \int_{0/\lambda}^{\lambda/\lambda}\lambda^{1/2}W_{2k}'\left(\frac{s}{\lambda}\right)d\left(\lambda\left(\frac{s}{\lambda}\right)\right)\\ &= \lambda^{3/2}\int_0^1W_{2k}'\left(s_1\right)ds_1,\end{aligned}$$

for (c):

$$\begin{aligned}\int_0^{\lambda}sW_{2k}'\left(s\right)ds &= \int_{0/\lambda}^{\lambda/\lambda}\lambda\left(\frac{s}{\lambda}\right)\lambda^{1/2}W_{2k}'\left(\frac{s}{\lambda}\right)d\left(\lambda\left(\frac{s}{\lambda}\right)\right)\\ &= \lambda^{5/2}\int_0^1s_1W_{2k}'\left(s_1\right)ds_1,\end{aligned}$$

for (d):

$$\begin{aligned}\int_0^{\lambda} W_{2k}(s) W_{2k}'(s)\, ds &= \int_{0/\lambda}^{\lambda/\lambda} \lambda^{1/2} W_{2k}\left(\frac{s}{\lambda}\right) \lambda^{1/2} W_{2k}'\left(\frac{s}{\lambda}\right) d\left(\lambda\left(\frac{s}{\lambda}\right)\right) \\ &= \lambda^2 \int_0^1 W_{2k}(s_1) W_{2k}'(s_1)\, ds_1,\end{aligned}$$

for (e):

$$\begin{aligned}\int_0^{\lambda} s dW_{11}(s) &= \int_{0/\lambda}^{\lambda/\lambda} \lambda\left(\frac{s}{\lambda}\right) d\left(\lambda^{1/2} W_{11}\left(\frac{s}{\lambda}\right)\right) \\ &= \lambda^{3/2} \int_0^1 s_1 dW_{11}(s_1),\end{aligned}$$

for (f):

$$\begin{aligned}\int_0^{\lambda} W_{2k}(s)\, dW_{11}(s) &= \int_{0/\lambda}^{\lambda/\lambda} \lambda^{1/2} W_{2k}\left(\frac{s}{\lambda}\right) d\left(\lambda^{1/2} W_{11}\left(\frac{s}{\lambda}\right)\right) \\ &= \lambda \int_0^1 W_{2k}(s_1)\, dW_{11}(s_1).\end{aligned}$$

Thus, Lemma 5 has been proved. ■

Lemma 6 *Let us denote $\lambda < r < 1$ as the index that defines the standard Brownian motion. Thus, by the properties of the Brownian motions it can be shown that*
(a) $W(r) = \sqrt{1-\lambda} W(b_2)$,
(b) $\int_{\lambda}^{r} W_{2k}'(s)\, ds = (1-\lambda)^{3/2} \int_0^{b_2} W_{2k}'(s_2)\, ds_2$,
(c) $\int_{\lambda}^{1} s W_{2k}'(s)\, ds = (1-\lambda)^{5/2} \int_0^1 s_2 W_{2k}'(s_2)\, ds_2 + \lambda(1-\lambda)^{3/2} \int_0^1 W_{2k}'(s_2)\, ds_2$,
(d) $\int_{\lambda}^{1} W_{2k}(s) W_{2k}'(s)\, ds = (1-\lambda)^2 \int_0^1 W_{2k}(s_2) W_{2k}'(s_2)\, ds_2$,
(e) $W_{11}(1) - W_{11}(\lambda) = (1-\lambda)^{1/2} W_{11}(1)$,
(f) $\int_{\lambda}^{1} s dW_{11}(s) = (1-\lambda)^{3/2} \int_0^1 s_2 dW_{11}(s_2) + \lambda(1-\lambda)^{1/2} W_{11}(1)$,
(g) $\int_{\lambda}^{1} W_{2k}(s)\, dW_{11}(s) = (1-\lambda) \int_0^1 W_{2k}(s_2)\, dW_{11}(s_2)$,
(h) $\int_0^{\lambda} W_{2k}'(s)\, ds = \int_0^1 W_{2k}'(s)\, ds - (1-\lambda)^{3/2} \int_0^1 W_{2k}'(s_2)\, ds_2$,
(i) $\int_0^{\lambda} s W_{2k}'(s)\, ds = \int_0^1 s W_{2k}'(s)\, ds - \Big((1-\lambda)^{5/2} \int_0^1 s_2 W_{2k}'(s_2)\, ds_2$
$+\lambda(1-\lambda)^{3/2} \int_0^1 W_{2k}'(s_2)\, ds_2\Big)$,
(j) $\int_0^{\lambda} W_{2k}(s) W_{2k}'(s)\, ds = \int_0^1 W_{2k}(s) W_{2k}'(s)\, ds - (1-\lambda)^2 \int_0^1 W_{2k}(s_2) W_{2k}'(s_2)\, ds_2$,
(k) $W_{11}(\lambda) = 0$,
(l) $\int_0^{\lambda} s dW_{11}(s) = -\int_0^1 W_{11}(s)\, ds + (1-\lambda)^{3/2} \int_0^1 W_{11}(s_2)\, ds_2$,
(m) $\int_0^{\lambda} W_{2k}(s)\, dW_{11}(s) = \int_0^1 W_{2k}(s)\, dW_{11}(s) - (1-\lambda) \int_0^1 W_{2k}(s_2)\, dW_{11}(s_2)$,
with $b_2 = (r-\lambda)/(1-\lambda)$, $0 < b_2 < 1$.

Proof: *The first statement follows from the properties of the Brownian motions. Thus, for the statement (a):*

$$W(r) = \sqrt{1-\lambda} W\left(\frac{r-\lambda}{1-\lambda}\right) = \sqrt{1-\lambda} W(b_2),$$

for (b):

$$\begin{aligned}\int_\lambda^r W'_{2k}(s)\,ds &= \int_{\frac{\lambda-\lambda}{1-\lambda}}^{\frac{r-\lambda}{1-\lambda}} (1-\lambda)^{1/2} W'_{2k}\left(\frac{s-\lambda}{1-\lambda}\right) d\left(\frac{s-\lambda}{1-\lambda}\right)(1-\lambda) \\ &= (1-\lambda)^{3/2} \int_0^{b_2} W'_{2k}(s_2)\,ds_2,\end{aligned}$$

for (c):

$$\begin{aligned}\int_\lambda^1 sW'_{2k}(s)\,ds &= \int_\lambda^1 (s-\lambda+\lambda) W'_{2k}(s)\,ds \\ &= \int_\lambda^1 (s-\lambda) W'_{2k}(s)\,ds + \lambda \int_\lambda^1 W'_{2k}(s)\,ds \\ &= (1-\lambda)^{5/2} \int_0^1 s_2 W'_{2k}(s_2)\,ds_2 + \lambda (1-\lambda)^{3/2} \int_0^1 W'_{2k}(s_2)\,ds_2,\end{aligned}$$

for (d):

$$\begin{aligned}\int_\lambda^1 W_{2k}(s) W'_{2k}(s)\,ds &= \int_{\frac{\lambda-\lambda}{1-\lambda}}^{\frac{r-\lambda}{1-\lambda}} (1-\lambda)^{1/2} W_{2k}\left(\frac{s-\lambda}{1-\lambda}\right)(1-\lambda)^{1/2} W'_{2k}\left(\frac{s-\lambda}{1-\lambda}\right) d\left((1-\lambda)\left(\frac{s-\lambda}{1-\lambda}\right)\right) \\ &= (1-\lambda)^2 \int_0^1 W_{2k}(s_2) W'_{2k}(s_2)\,ds_2,\end{aligned}$$

for (e):

$$\begin{aligned}W_{11}(1) - W_{11}(\lambda) &= (1-\lambda)^{1/2} W_{11}\left(\frac{1-\lambda}{1-\lambda}\right) - (1-\lambda)^{1/2} W_{11}\left(\frac{\lambda-\lambda}{1-\lambda}\right) \\ &= (1-\lambda)^{1/2} W_{11}(1),\end{aligned}$$

for (f):

$$\begin{aligned}\int_\lambda^1 s\,dW_{11}(s) &= \int_\lambda^1 (s-\lambda+\lambda)\,dW_{11}(s) \\ &= \int_\lambda^1 (s-\lambda)\,dW_{11}(s) + \lambda \int_\lambda^1 dW_{11}(s) \\ &= (1-\lambda)^{3/2} \int_0^1 s_2\,dW_{11}(s_2) + \lambda \int_\lambda^1 dW_{11}(s) \\ &= (1-\lambda)^{3/2} \int_0^1 s_2\,dW_{11}(s_2) + \lambda (1-\lambda)^{1/2} W_{11}(1),\end{aligned}$$

for (g):

$$\int_{\lambda}^{1} W_{2k}(s)\,dW_{11}(s) = \int_{\frac{\lambda-\lambda}{1-\lambda}}^{\frac{r-\lambda}{1-\lambda}} (1-\lambda)^{1/2} W_{2k}\left(\frac{s-\lambda}{1-\lambda}\right) d\left((1-\lambda)^{1/2} W_{11}\left(\frac{s-\lambda}{1-\lambda}\right)\right)$$
$$= (1-\lambda)\int_{0}^{1} W_{2k}(s_2)\,dW_{11}(s_2),$$

for (h):

$$\int_{0}^{\lambda} W'_{2k}(s)\,ds = \int_{0}^{1} W'_{2k}(s)\,ds - \int_{\lambda}^{1} W'_{2k}(s)\,ds$$
$$= \int_{0}^{1} W'_{2k}(s)\,ds - (1-\lambda)^{3/2}\int_{0}^{1} W'_{2k}(s_2)\,ds_2,$$

for (i):

$$\int_{0}^{\lambda} sW'_{2k}(s)\,ds = \int_{0}^{1} sW'_{2k}(s)\,ds - \int_{\lambda}^{1} sW'_{2k}(s)\,ds$$
$$= \int_{0}^{1} sW'_{2k}(s)\,ds - \left((1-\lambda)^{5/2}\int_{0}^{1} s_2 W'_{2k}(s_2)\,ds_2 + \lambda(1-\lambda)^{3/2}\int_{0}^{1} W'_{2k}(s_2)\,ds_2\right),$$

for (j):

$$\int_{0}^{\lambda} W_{2k}(s)\,W'_{2k}(s)\,ds = \int_{0}^{1} W_{2k}(s)\,W'_{2k}(s)\,ds - \int_{\lambda}^{1} W_{2k}(s)\,W'_{2k}(s)\,ds$$
$$= \int_{0}^{1} W_{2k}(s)\,W'_{2k}(s)\,ds - (1-\lambda)^{2}\int_{0}^{1} W_{2k}(s_2)\,W'_{2k}(s_2)\,ds_2,$$

for (k):

$$W_{11}(\lambda) = (1-\lambda)^{1/2} W_{11}\left(\frac{\lambda-\lambda}{(1-\lambda)}\right)$$
$$= 0,$$

for (l):

$$\int_{0}^{\lambda} s\,dW_{11}(s) = \int_{0}^{1} s\,dW_{11}(s) - \int_{\lambda}^{1} s\,dW_{11}(s)$$
$$= \left(W_{11}(1) - \int_{0}^{1} W_{11}(s)\,ds\right) - \left((1-\lambda)^{1/2} W_{11}(1) - (1-\lambda)^{3/2}\int_{0}^{1} W_{11}(s_2)\,ds_2\right)$$
$$= -\int_{0}^{1} W_{11}(s)\,ds + (1-\lambda)^{3/2}\int_{0}^{1} W_{11}(s_2)\,ds_2,$$

for (m):

$$\int_{0}^{\lambda} W_{2k}(s)\,dW_{11}(s) = \int_{0}^{1} W_{2k}(s)\,dW_{11}(s) - \int_{\lambda}^{1} W_{2k}(s)\,dW_{11}(s)$$
$$= \int_{0}^{1} W_{2k}(s)\,dW_{11}(s) - (1-\lambda)\int_{0}^{1} W_{2k}(s_2)\,dW_{11}(s_2).$$

Thus, Lemma 6 has been proved. ■

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– provided that the proof for the other models follows the same reasoning and are particular cases of that one. 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:

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

Using lemmas 1 and 2 is straightforward to see that, in 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: $h_{11} = 1$, $h_{12} = h_{22} = (1-\lambda)$, $h_{13} = 1/2$, $h_{14} = h_{24} = (1-\lambda)^2/2$, $h_{15} = \int_0^1 W_{2k}'$, $h_{16} = h_{25} = h_{26} = \int_\lambda^1 W_{2k}'$, $h_{23} = \left(1-\lambda^2\right)/2$, $h_{33} = 1/3$, $h_{34} = \left(2 - 3\lambda + \lambda^3\right)/6$, $h_{35} = \int_0^1 rW_{2k}'$, $h_{36} = \int_\lambda^1 rW_{2k}'$, $h_{44} = (1-\lambda)^3/3$, $h_{45} = h_{46} = \int_\lambda^1 (r-\lambda) W_{2k}'$, $h_{55} = \int_0^1 W_{2k}W_{2k}'$, $h_{56} = h_{66} = \int_\lambda^1 W_{2k}W_{2k}'$. Also, it is easy to see that, in the limit, the product between matrix regressors and disturbance term ε converges to the (6×1)-vector J, $Pz'\varepsilon \Rightarrow J$, with elements given by: $J_1 = \omega_1 W_{11}(1)$, $J_2 = \omega_1 (W_{11}(1) - W_{11}(\lambda))$, $J_3 = \omega_1 \left(\int_0^1 rdW_{11}\right)$, $J_4 = \omega_1 \left(\int_\lambda^1 rdW_{11} - \lambda (W_{11}(1) - W_{11}(\lambda))\right)$, $J_5 = \omega_1 \Omega_{22}^{1/2} \int_0^1 W_{2k}dW_{11}$, $J_6 = \omega_1 \Omega_{22}^{1/2} \int_\lambda^1 W_{2k}dW_{11}$. The partial sum process of matrix regressors for $t \leq T_b$ converges to:

$$\begin{aligned} T^{-1/2}\sum_{j=1}^{t} z_j PA &\Rightarrow \begin{bmatrix} r & 0 & \frac{r^2}{2} & 0 & \int_0^r W_{2k}' & 0 \end{bmatrix} \\ &= K_1, \end{aligned} \tag{A-1}$$

whereas for $t > T_b$:

$$\begin{aligned} T^{-1/2}\sum_{j=1}^{t} z_j PA &\Rightarrow \begin{bmatrix} r & (r-\lambda) & \frac{r^2}{2} & \frac{(r-\lambda)^2}{2} & \int_0^r W_{2k}' & \int_\lambda^r W_{2k}' \end{bmatrix} \\ &= K_2. \end{aligned} \tag{A-2}$$

Using these results, for $t \leq T_b$ we have that in the limit partial sum process converges to:

$$T^{-1/2}\omega_1^{-1}\hat{S}_t \Rightarrow W_{11}(r) - \begin{bmatrix} r & 0 & \frac{r^2}{2} & 0 & \int_0^r W_{2k}' & 0 \end{bmatrix} H^{-1}J$$

$$
\begin{aligned}
&= W_{11}(r) - K_1 H^{-1} J \\
&\equiv L_{k,E1}(r,\lambda). \qquad \text{(A-3)}
\end{aligned}
$$

On the other hand, we have that for $t > T_b$ partial sum process converges to:

$$
\begin{aligned}
T^{-1/2}\omega_1^{-1}\hat{S}_t &\Rightarrow W_{11}(r) - \left[\begin{array}{cccccc} r & (r-\lambda) & \frac{r^2}{2} & \frac{(r-\lambda)^2}{2} & \int_0^r W_{2k}' & \int_\lambda^r W_{2k}' \end{array}\right] H^{-1} J \\
&= W_{11}(r) - K_2 H^{-1} J \\
&\equiv L_{k,E2}(r,\lambda) \qquad \text{(A-4)}
\end{aligned}
$$

Appealing to the Continuous Mapping Theorem (CMT), we obtain:

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

where $L_{k,E1}(r,\lambda)$ and $L_{k,E2}(r,\lambda)$ have been defined above and $\hat{\omega}_1^2$ is a consistent estimate of the long-run variance of $\{e_t\}$. The symmetry of the asymptotic distribution needs some additional algebra calculations. To show the symmetry of the asymptotic distribution, we write the equation that is used to compute the test statistic as:

$$
\begin{aligned}
y_t &= \alpha_1 D_{1t} + \alpha_2 D_{2t} + \beta_1 DT_{1t} + \beta_2 DT_{2t} + \gamma_1 x_t' D_{1t} + \gamma_2 x_t' D_{2t} + \varepsilon_t \\
&= \alpha_1 D_{1t} + \beta_1 DT_{1t} + \gamma_1 x_t' D_{1t} + \alpha_2 D_{2t} + \beta_2 DT_{2t} + \gamma_2 x_t' D_{2t} + \varepsilon_t,
\end{aligned}
$$

where $D_{1t} = 1$ and $DT_{1t} = t$ for $t \leq T_b$ and 0 otherwise, $D_{2t} = 1$ and $DT_{2t} = (t - T_b)$ for $t > T_b$ and 0 otherwise. Notice that now the model can be expressed in terms of a block matrix of the regressors:

$$
y = z\delta + \varepsilon,
$$

with

$$
z = \left[\begin{array}{cc} z_1 & 0 \\ 0 & z_2 \end{array}\right],
$$

z_1 being a $(T_b \times (2+k))$-matrix with row vectors given by $z_{1t} = (1, t, x_t')$, $1 < t \leq T_b$, and z_2 being a $((T - T_b) \times (2+k))$-matrix with row vectors given by $z_{2t} = (1, t, x_t')$, $T_b < t \leq T$. In these terms, the partial sum processes for $t \leq T_b$ are given by:

$$
T^{-1/2}\omega_1^{-1}\sum_{j=1}^{t}\hat{e}_j \Rightarrow W_{11}(r) - \left[\begin{array}{cccccc} r & \frac{r^2}{2} & \int_0^r W_{2k}'(s)\,ds & 0 & 0 & 0 \end{array}\right] \times
$$

$$\begin{bmatrix} l(z_1'z_1) & 0 \\ 0 & l(z_2'z_2) \end{bmatrix}^{-1} \begin{bmatrix} W_{11}(\lambda) \\ \int_0^{\lambda} s dW_{11}(s) \\ \int_0^{\lambda} W_{2k}(s) dW_{11}(s) \\ W_{11}(1) - W_{11}(\lambda) \\ \int_{\lambda}^{1} s dW_{11}(s) \\ \int_{\lambda}^{1} W_{2k}(s) dW_{11}(s) \end{bmatrix},$$

with

$$\begin{aligned} l(z_1'z_1) &= \begin{bmatrix} \lambda & \lambda^2/2 & \int_0^{\lambda} W_{2k}'(s) ds \\ \lambda^2/2 & \lambda^3/3 & \int_0^{\lambda} sW_{2k}'(s) ds \\ \int_0^{\lambda} W_{2k}(s) ds & \int_0^{\lambda} sW_{2k}(s) ds & \int_0^{\lambda} W_{2k}(s) W_{2k}'(r) ds \end{bmatrix} \\ l(z_2'z_2) &= \begin{bmatrix} (1-\lambda) & (1-\lambda^2)/2 & \int_{\lambda}^{1} W_{2k}'(s) ds \\ (1-\lambda^2)/2 & (1-\lambda^3)/3 & \int_{\lambda}^{1} sW_{2k}'(s) ds \\ \int_{\lambda}^{1} W_{2k}(s) ds & \int_{\lambda}^{1} sW_{2k}(s) ds & \int_{\lambda}^{1} W_{2k}(s) W_{2k}'(s) ds \end{bmatrix}. \end{aligned}$$

Using the results from Lemma 5, we can write this expression in terms of $b_1 = r/\lambda$ and reduce the partial sum processes to:

$$\begin{aligned} T^{-1/2}\omega_1^{-1} \sum_{j=1}^{t} \hat{e}_j &\Rightarrow \sqrt{\lambda} W_{11}(b_1) - \begin{bmatrix} \lambda b_1 & \lambda^2 \frac{b_1^2}{2} & \lambda^{3/2} \int_0^{b_1} W_{2k}'(s_1) ds_1 \end{bmatrix} \\ & \quad \times \left[(l z_1' z_1)_{s_1} \right]^{-1} \begin{bmatrix} \sqrt{\lambda} W_{11}(1) \\ \lambda^{3/2} \int_0^1 s_1 dW_{11}(s_1) \\ \lambda \int_0^1 W_{2k}(s_1) dW_{11}(s_1) \end{bmatrix}, \end{aligned}$$

$0 < b_1 < 1$ and $0 < s_1 < 1$ and where $(l z_1' z_1)_{s_1}$ denotes the matrix $l z_1' z_1$ written using $s_1 = s/\lambda$ as the index of the Brownian motions as shown in Lemma 5. After some calculations, it can be established that the partial sum processes converges in probability to:

$$\begin{aligned} T^{-1/2}\omega_1^{-1} \sum_{j=1}^{t} \hat{e}_j &\Rightarrow \sqrt{\lambda} \left(W_{11}(b_1) - \frac{1}{\Delta(b_1)} f(b_1) \right) \\ &= \sqrt{\lambda} V_{k,E}(b_1) \end{aligned} \quad \text{(A-5)}$$

where

$$\begin{aligned} \Delta(b_1) &= \int_0^1 W_{2k}(s_1) W_{2k}'(s_1) ds_1 - 12 \int_0^1 s_1 W_{2k}'(s_1) ds_1 \int_0^1 s_1 W_{2k}(s_1) ds_1 \\ & \quad + 6 \int_0^1 W_{2k}'(s_1) ds_1 \int_0^1 s_1 W_{2k}(s_1) ds_1 + 6 \int_0^1 W_{2k}(s_1) ds_1 \int_0^1 s_1 W_{2k}'(s_1) ds_1 \\ & \quad - 4 \int_0^1 W_{2k}(s_1) ds_1 \int_0^1 W_{2k}'(s_1) ds_1 \end{aligned}$$

and

$$
\begin{aligned}
f\left(b_1\right) = & \left(\left(-3\int_0^1 W_{2k}\left(s_1\right) W_{2k}'\left(s_1\right) ds_1 + 6\int_0^1 W_{2k}\left(s_1\right) ds_1 \int_0^1 s_1 W_{2k}'\left(s_1\right) ds_1\right) b_1^2\right. \\
& + \left(4\int_0^1 W_{2k}\left(s_1\right) W_{2k}'\left(s_1\right) ds_1 - 12\int_0^1 s_1 W_{2k}'\left(s_1\right) ds_1 \int_0^1 s_1 W_{2k}\left(s_1\right) ds_1\right) b_1 \\
& \left. + \int_0^{b_1} W_{2k}'\left(s_1\right) ds_1 \left(6\int_0^1 s_1 W_{2k}\left(s_1\right) ds_1 - 4\int_0^1 W_{2k}\left(s_1\right) ds_1\right)\right) W_{11}\left(b_1\right) \\
& + \left(6\int_0^1 s_1 dW_{11}\left(s_1\right) \int_0^1 W_{2k}\left(s_1\right) W_{2k}'\left(s_1\right) ds_1\right. \\
& -6\int_0^1 s_1 dW_{11}\left(s_1\right) \int_0^1 W_{2k}\left(s_1\right) ds_1 \int_0^1 W_{2k}'\left(s_1\right) ds_1 \\
& -6\int_0^1 W_{2k}\left(s_1\right) dW_{11}\left(s_1\right) \int_0^1 s_1 W_{2k}'\left(s_1\right) ds_1 \\
& \left. +3\int_0^1 W_{2k}\left(s_1\right) dW_{11}\left(s_1\right) \int_0^1 W_{2k}'\left(s_1\right) ds_1\right) b_1^2 \\
& + \left(-4\int_0^1 W_{2k}\left(s_1\right) dW_{11}\left(s_1\right) \int_0^1 W_{2k}'\left(s_1\right) ds_1\right. \\
& -6\int_0^1 s_1 dW_{11}\left(s_1\right) \int_0^1 W_{2k}\left(s_1\right) W_{2k}'\left(s_1\right) ds_1 \\
& +12\int_0^1 s_1 dW_{11}\left(s_1\right) \int_0^1 W_{2k}'\left(s_1\right) ds_1 \int_0^1 s_1 W_{2k}\left(s_1\right) ds_1 \\
& \left. +6\int_0^1 W_{2k}\left(s_1\right) dW_{11}\left(s_1\right) \int_0^1 s_1 W_{2k}'\left(s_1\right) ds_1\right) b_1 \\
& -12\int_0^1 s_1 dW_{11}\left(s_1\right) \int_0^{b_1} W_{2k}'\left(s_1\right) ds_1 \int_0^1 s_1 W_{2k}\left(s_1\right) ds_1 \\
& +6\int_0^1 s_1 dW_{11}\left(s_1\right) \int_0^{b_1} W_{2k}'\left(s_1\right) ds_1 \int_0^1 W_{2k}\left(s_1\right) ds_1 \\
& + \int_0^1 W_{2k}\left(s_1\right) dW_{11}\left(s_1\right) \int_0^{b_1} W_{2k}'\left(s_1\right) ds_1 .
\end{aligned}
$$

Notice that $V_{k,E}\left(b_1\right)$ is the residual from a projection of a standard Wiener process $W_{11}\left(b_1\right)$ onto the subspace generated by the deterministic regressors (properly rescaled) and the $W_{2k}\left(r\right)$ defined in $0 < b_1 < 1$. The expression that can be obtained for the partial sum processes when $t > T_b$ is similar to the previous one. Now we use the transformation given by $b_2 = \left(r - \lambda\right) / \left(1 - \lambda\right)$, transformation that allow us to compute the partial sum processes as:

$$
\begin{aligned}
T^{-1/2}\omega_1^{-1}\sum_{j=1}^{t}\hat{e}_j &\Rightarrow \sqrt{1-\lambda}W_{11}(b_2) \\
&- \left[\begin{array}{ccc} (1-\lambda)b_2 & \frac{(1-\lambda)^2 b_2^2}{2}+\left(\lambda-\lambda^2\right)b_2 & (1-\lambda)^{3/2}\int_0^{b_2}W_{2k}'(s_2)\,ds_2 \end{array}\right] \\
&\times\left[(lz_2'z_2)_{s_2}\right]^{-1}\left[\begin{array}{c} (1-\lambda)^{1/2}W_{11}(1) \\ (1-\lambda)^{3/2}\int_0^1 s_2 dW_{11}(s_2)+\lambda(1-\lambda)^{1/2}W_{11}(1) \\ (1-\lambda)\int_0^1 W_{2k}(s_2)\,dW_{11}(s_2) \end{array}\right],
\end{aligned}
$$

where we have applied the results collected in Lemma 6. After some algebra manipulations, we can achieve the following expression for the partial sum processes for $t > T_b$ are given by:

$$
\begin{aligned}
T^{-1/2}\omega_1^{-1}\sum_{j=1}^{t}\hat{e}_j &\Rightarrow \sqrt{1-\lambda}\left(W_{11}(b_2)-\frac{1}{\Delta(b_2)}f(b_2)\right) \qquad \text{(A-6)} \\
&= \sqrt{1-\lambda}V_{k,E}(b_2),
\end{aligned}
$$

with $\Delta(b_2)$ and $f(b_2)$ as before but with b_2 instead of b_1 and s_2 instead of s_1. As before, $V_{k,E}(b_2)$ is a residual projection.

The test statistic $SC_E(\lambda)$ is given by:

$$
\begin{aligned}
SC_E(\lambda) &= T^{-2}\hat{\omega}_1^{-2}\sum_{t=1}^{T_b}S_{E,t}^2+T^{-2}\hat{\omega}_1^{-2}\sum_{t=T_b+1}^{T}S_{E,t}^2 \\
&\Rightarrow \int_0^{\lambda}\left(\sqrt{\lambda}V_{k,E}(b_1)\right)^2 dr+\int_{\lambda}^{1}\left(\sqrt{1-\lambda}V_{k,E}(b_2)\right)^2 dr \\
&= \lambda\int_0^1\left(\sqrt{\lambda}V_{k,E}(b_1)\right)^2 db_1+(1-\lambda)\int_0^1\left(\sqrt{1-\lambda}V_{k,E}(b_2)\right)^2 db_2 \\
&= \lambda^2\int_0^1 V_{k,E}^2(b_1)\,db_1+(1-\lambda)^2\int_0^1 V_{k,E}^2(b_2)\,db_2,
\end{aligned}
$$

that is, the asymptotic distribution of the test statistic can be expressed as a weighted sum of two independent detrended Brownian motions. Then, Theorem 1 has been proved. ■

A.2 Proof of Theorem 2

As for the proof of the Theorem 1 we concentrate on the model E. The transformation proposed by Phillips and Hansen (1990) on the DGP given by (1), (2) and (3) under the null hypothesis reduces to (7) that, in matrix notation, can be expressed as $y_t^+ =$

$z_t\beta + \varepsilon_t^+$, where, for the model E, $z_t = [1,\, DU_t,\, t,\, DT_t^*,\, x_t',\, x_t' DU_t]$. The fully-modified OLS residuals can be computed through:

$$\begin{aligned} \hat{e}_t^+ &= \left(z_t\beta + \varepsilon_t^+\right) - z_t\left(z'z\right)^{-1} z' \left(y^+ - e_K T\hat{\delta}_{21}^+\right) \\ &= \varepsilon_t^+ - z_t\left(z'z\right)^{-1} z' \left(\varepsilon^+ - e_K T\hat{\delta}_{21}^+\right). \end{aligned}$$

Provided that:

$$\begin{aligned} T^{-1/2}\sum_{j=1}^{t}\varepsilon_j^+ &\Rightarrow \omega_{1.2}W_{1.2}\left(r\right); \\ T^{-1}\sum_{t=1}^{T}x_t\varepsilon_t^+ &\Rightarrow \omega_{1.2}\Omega_{22}^{1/2}\int_0^1 W_{2k}dW_{1.2} + \delta_{21}^+, \end{aligned}$$

where $W_{1.2}\left(r\right) = W_{11}\left(r\right) - \omega_{21}'\Omega_{22}^{-1}W_{2k}\left(r\right)$ is a standard Brownian motion independent of the vector Brownian motion $W_{2k}\left(r\right)$ and where $\delta_{21}^+ = \delta_{21} - \Delta_{22}\Omega_{22}^{-1}\omega_{21}$, it can be established that:

$$A'Pz'\left(\varepsilon^+ - e_K T\hat{\delta}_{21}^+\right) \Rightarrow \begin{bmatrix} W_{1.2}\left(1\right) \\ \left(W_{1.2}\left(1\right) - W_{1.2}\left(\lambda\right)\right) \\ \left(W_{1.2}\left(1\right) - \int_0^1 W_{1.2}\right) \\ \left(\left(1-\lambda\right)\left(W_{1.2}\left(1\right) - W_{1.2}\left(\lambda\right)\right) - \int_0^1 W_{1.2}\right) \\ \int_0^1 W_{2k}dW_{1.2} \\ \int_\lambda^1 W_{2k}dW_{1.2} \end{bmatrix},$$

so that asymptotic distributions that have been derived under the strictly exogenous stochastic regressors assumption still hold. The only thing we have to modify is the notation. Thus, instead of $W_{11}\left(r\right)$ we have to write $W_{1.2}\left(r\right)$ and instead of ω_1^2 we have to write $\omega_{1.2}^2$. Hence, Theorem 2 has been proved. ■

A.3 Proof of Theorem 3

To proof this Theorem we have followed the developments in Zivot and Andrews (1992). The strategy expresses the test statistic as a function of composite functions. Therefore, the continuity of a composition of continuous functions allow the application of the Continuous Mapping Theorem (CMT). The asymptotic distribution of the test given by 3 can be reexpressed as the composite functional $g\left(T^{-1/2}\omega_1^{-1}S_t,\, zPA,\, A'Pz'\varepsilon,\, \omega_1^2\right)$ defined by:

$$g\left(T^{-1/2}\omega_1^{-1}S_t, zPA, A'Pz'\varepsilon, \omega_1^2\right) = h^*\left(h\left[H_1\left[\bullet\right], H_2\left[\bullet\right]\right]\right),$$

$\left(\bullet\right) = \left(T^{-1/2}\omega_1^{-1}S_t, zPA, A'Pz'\varepsilon, \omega_1^2\right)$, with the functionals $h^*\left(m\right) = \inf_{\lambda\in\Lambda} m\left(\lambda\right)$

and

$$h\left(m_1,m_2\right)=T^{-1}\sum_{t=1}^{T_b}m_{1t}\left(\lambda\right)+T^{-1}\sum_{t=T_b+1}^{T}m_{2t}\left(\lambda\right).$$

The functionals H_1 and H_2 are defined as the square of the partial sum process computed using the OLS estimated residuals of (5) for $0<t<T_b$ and for $T_b\leq t<T$, respectively:

$$H_i\left[T^{-1/2}\omega_1^{-1}S_t,zPA,A'Pz'\varepsilon,\omega_1^2\right]=\left(T^{-1/2}\hat{\omega}_1^{-1}\hat{S}_t\right)^2,$$

$i=\{1,2\}$. Notice that as shown by Zivot and Andrews (1992) and Perron (1997), the elements that involve H_1 and H_2 are continuous at $\left(W_{11}\left(r\right),K_1,J,\omega_1^2\right)$ and $\left(W_{11}\left(r\right),K_2,J,\omega_1^2\right)$, with K_1 defined by (A-1) and K_2 defined by (A-2), since their elements are bounded over $[0,1]$ with W-probability 1. The continuity of $g\left(\bullet\right)$ is provided by the fact that $g\left(\bullet\right)$ is a composition of continuous functions. Finally, Lemma A.4 of Zivot and Andrews (1992) shows that the function h^* is continuous at all functions m on Λ, the closed subset of $[0,1]$ to which λ belongs. Therefore, the result collected in Theorem 3 follows from Theorems 1 and 2. Thus, Theorem 3 has been proved. ■

A.4 Proof of Theorem 4

The OLS disturbance term can be written as:

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

where τ denotes the true break fraction and λ the estimated break fraction. Using this formulation, the partial sum process of the numerator of the test is computed as:

$$\begin{aligned}\hat{S}_t &= \sum_{j=1}^{t}\varepsilon_j-\sum_{j=1}^{t}z_j\left(\lambda\right)\left(z'\left(\lambda\right)z\left(\lambda\right)\right)^{-1}z'\left(\lambda\right)\varepsilon\\ &\quad -\left(\sum_{j=1}^{t}z_j\left(\lambda\right)\left(z'\left(\lambda\right)z\left(\lambda\right)\right)^{-1}z'\left(\lambda\right)z\left(\tau\right)-\sum_{j=1}^{t}z_j\left(\tau\right)\right)\beta.\end{aligned}\tag{A-7}$$

Let us define the difference between the matrix of regressors using the estimated break fraction parameter and the matrix computed using the true break fraction as $d\left(\lambda\right)=z\left(\lambda\right)-z\left(\tau\right)$. Thus, the partial sum process given by (A-7) can be alternatively expressed as:

$$\hat{S}_t = \sum_{j=1}^{t}\varepsilon_j-\sum_{j=1}^{t}z_j\left(\lambda\right)\left(z'\left(\lambda\right)z\left(\lambda\right)\right)^{-1}z'\left(\lambda\right)\varepsilon$$

$$+\left(\sum_{j=1}^{t} z_j(\lambda)\left(z'(\lambda)\, z(\lambda)\right)^{-1} z'(\lambda)\, d(\lambda) - \sum_{j=1}^{t} d_j(\lambda)\right)\beta.$$

If we rescale this partial sum process as:

$$\begin{aligned} T^{-1/2}\hat{S}_t &= T^{-1/2}S_t \\ &-T^{-1/2}\sum_{j=1}^{t} z_j(\lambda)\, PA\left(A'Pz'(\lambda)\, z(\lambda)\, PA\right)^{-1} A'Pz'(\lambda)\,\varepsilon \\ &+T^{-1/2}\left(\sum_{j=1}^{t} z_j(\lambda)\, PA\left(A'Pz'(\lambda)\, z(\lambda)\, PA\right)^{-1} A'Pz'(\lambda)\, d(\lambda)\right. \\ &\left.-\sum_{j=1}^{t} d_j(\lambda)\right)\beta, \end{aligned} \tag{A-8}$$

we get an expression where the convergence or divergence can be analysed. Notice that the elements associated to ε converge at the suitable rate so that the divergence is due to the elements related to $d(\lambda)$. The order of probability of the partial sum process depends on the model we are dealing with. Consider, for instance, the specification that corresponds to the model E, that is, the most general model we have specified in this paper. It can be established that for the model E the difference between the true and estimated matrix of regressors is equal to $d(\lambda) = [0,\ DU(\lambda) - DU(\tau),\ 0,\ DT^*(\lambda) - DT^*(\tau),\ 0,\ x(DU(\lambda) - DU(\tau))]$. Our analysis has to distinguish between two relevant situations. The first one arises when $\lambda > \tau$, whereas the second one is given by $\lambda < \tau$.

Let us first consider that $\lambda > \tau$. Using the statements collected in Lemma 3, it can be shown that $Pz'(\lambda)\, d(\lambda)$ is a (6×6) matrix with the first, third and fifth columns of zeros and the rest of the columns defined by the following elements: $Pz'(\lambda)\, d(\lambda)[1,2] = (-\lambda+\tau)\, T^{1/2}$, $Pz'(\lambda)\, d(\lambda)[2,2] = 0$, $Pz'(\lambda)\, d(\lambda)[3,2] = -\frac{1}{2}T^{1/2}(\lambda-\tau)(\lambda T + 1 + \tau T)$, $Pz'(\lambda)\, d(\lambda)[4,2] = 0$, $Pz'(\lambda)\, d(\lambda)[5,2] = -T^{-1/2}\sum_{t=\tau T+1}^{T\lambda} x_t$, $Pz'(\lambda)\, d(\lambda)[6,2] = 0$, $Pz'(\lambda)\, d(\lambda)[1,4] = (-1+\lambda)(\lambda-\tau)\, T^{1/2}$, $Pz'(\lambda)\, d(\lambda)[2,4] = (-1+\lambda)(\lambda-\tau)\, T^{1/2}$, $Pz'(\lambda)\, d(\lambda)[3,4] = -\frac{1}{2}T^{1/2}(\lambda-\tau)(T+1)$, $Pz'(\lambda)\, d(\lambda)[4,4] = -\frac{1}{2}T^{1/2}(-1+\lambda)(\lambda-\tau)(\lambda T - T - 1)$, $Pz'(\lambda)\, d(\lambda)[5,4] = -T^{-3/2}\sum_{t=\tau T+1}^{T\lambda} t x_t - T^{-1/2}\lambda\sum_{t=T\lambda+1}^{T} x_t + \tau T^{-1/2}\sum_{t=T\tau+1}^{T} x_t$, $Pz'(\lambda)\, d(\lambda)[6,4] = (-\lambda+\tau)\, T^{-1/2}\sum_{t=T\lambda+1}^{T} x_t$, $Pz'(\lambda)\, d(\lambda)[1,6] = Pz'(\lambda)\, d(\lambda)[2,6] = 0$, $Pz'(\lambda)\, d(\lambda)[3,6] = -T^{-1}\sum_{t=T\tau+1}^{T\lambda} t x_t$, $Pz'(\lambda)\, d(\lambda)[4,6] = 0$, $Pz'(\lambda)\, d(\lambda)[5,6] = -T^{-1}\sum_{t=\tau T+1}^{T\lambda} x_t x_t'$, $Pz'(\lambda)\, d(\lambda)[6,6] = 0$. It is easy to see that for the model E: $Pz'(\lambda)\, d(\lambda) = O_p\left(T^{3/2}\right)$. Notice that depending on the model we are interested in and, consequently, on the set of regressors included in the estimation, the rate of divergence of the matrix $Pz'(\lambda)\, d(\lambda)$ can change. Thus, it is easy to see that for the models An and A: $Pz'\left(\hat{\lambda}\right) d(\lambda) = O_p\left(T^{1/2}\right)$; for the model D: $Pz'\left(\hat{\lambda}\right) d(\lambda) = O_p(T)$ and, finally, for the models B, C and E: $Pz'\left(\hat{\lambda}\right) d(\lambda) =$

$O_p\left(T^{3/2}\right)$.

Let us now deal with the situation in which $\lambda < \tau$. It has to be mentioned that the differences between the statements in Lemmas 3 and 4 do not change the order of the probability of the matrix $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$. If we consider the specification given by the model E, is straightforward to see the second column of the matrix $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$ is $O_p\left(T^{1/2}\right)$, the fourth column is $O_p\left(T^{3/2}\right)$ and the sixth column is $O_p\left(T\right)$. Hence, $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$ is $O_p\left(T^{3/2}\right)$ for the model E. For the models An and A: $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$ is $O_p\left(T^{1/2}\right)$; for the model D: $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$ is $O_p\left(T^{1}\right)$ and, finally, for the models B and C: $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$ is $O_p\left(T^{3/2}\right)$. Consequently, we have the orders of probability of $Pz'\left(\hat{\lambda}\right)d\left(\lambda\right)$ are the same irrespective if $\lambda > \tau$ or $\lambda < \tau$.

On the other hand, we can see that for $\tau \neq \lambda$ and $\min\{\lambda, \tau\} < t/T$, the partial sum process $T^{-1/2}\sum_{j=1}^{t} d_j\left(\lambda\right)$ is defined by the following elements:

$$
\begin{aligned}
T^{-1/2}\sum_{j=1}^{t}\left(DU_j\left(\lambda\right) - DU_j\left(\tau\right)\right) &= T^{-1/2}\left(\left(t-\lambda T\right) - \left(t - \tau T\right)\right) \\
&= O\left(T^{1/2}\right), \\
T^{-1/2}\sum_{j=1}^{t}\left(DT_j^*\left(\lambda\right) - DT_j^*\left(\tau\right)\right) &= T^{-1/2}\left(\left(\frac{1}{2}t^2 + \frac{1}{2}t - \lambda T t\right) - \left(\frac{1}{2}t^2 + \frac{1}{2}t - \tau T t\right)\right) \\
&= O\left(T^{3/2}\right), \\
T^{-1/2}\sum_{j=1}^{t} x_j'\left(DU_j\left(\lambda\right) - DU_j\left(\tau\right)\right) &= O_p\left(T\right).
\end{aligned}
$$

Thus, for the models An and A we get that $T^{-1/2}\hat{S}_t = O_p\left(T^{1/2}\right)$, for the model D we have that $T^{-1/2}\hat{S}_t = O_p\left(T\right)$, whereas for the models B, C and E we have that $T^{-1/2}\hat{S}_t = O_p\left(T^{3/2}\right)$.

Provided that the numerator of the test statistics involves the squared sum of the partial sum processes, we can conclude that the numerator of the test diverges at a rate of $O_p\left(T\right)$ for the models An and A, at a rate of $O_p\left(T^2\right)$ for the model D and at a rate of $O_p\left(T^3\right)$ for the models B, C and E.

Now we are going to fix our attention in the denominator of the test statistic. The estimation of the long-run variance that is frequently used in applied research is a function of the variance and covariances of the disturbance term. From previous derivations, we have that the variance of the disturbance term is equal to:

$$
\hat{\sigma}_e^2 = T^{-1}\left(\varepsilon - z\left(\lambda\right)\left(z'\left(\lambda\right)z\left(\lambda\right)\right)^{-1}z'\left(\lambda\right)\varepsilon - \left(z\left(\lambda\right)\left(z'\left(\lambda\right)z\left(\lambda\right)\right)^{-1}z'\left(\lambda\right)z\left(\tau\right) - z\left(\tau\right)\right)\beta\right)'
$$

$$
\begin{aligned}
& \left(\varepsilon - z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)\varepsilon - \left(z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)z(\tau) - z(\tau)\right)\beta\right) \\
= \;& T^{-1}\left(\varepsilon'\varepsilon - \varepsilon' z(\lambda)P\left(Pz'(\lambda)z(\lambda)P\right)^{-1} Pz'(\lambda)\varepsilon \right.\\
& -\beta' d'(\lambda)z(\lambda)P\left(Pz'(\lambda)z(\lambda)P\right)^{-1} Pz'(\lambda)d(\lambda)\beta + \beta' d'(\lambda)d(\lambda)\beta \\
& \left. +2\varepsilon' z(\lambda)P\left(Pz'(\lambda)z(\lambda)P\right)^{-1} Pz'(\lambda)d(\lambda) - 2\varepsilon' d(\lambda)\right).
\end{aligned}
$$

The order of probability of the variance depends on the model. Let us first consider the model An for which the order of probability of the variance is given by the third and fourth element of the last equality. It has been shown that for the model An the matrix $Pz'(\lambda)d(\lambda)$ is $O_p\left(T^{1/2}\right)$ and, hence, the third element of the last equality is $O_p(T)$. Since this term is scaled by T^{-1} we have that it converges to an expression that is function of the parameters that involves the model. This is also true for the fourth element of the variance. Thus, $\hat{\sigma}_e^2$ is $O_p(1)$ for the model An. The same argument can be applied for the model A since the structural change only affects the independent term of the relationship. Therefore, $\hat{\sigma}_e^2$ is also $O_p(1)$ for the model A.

For the models that allow for a shift in the trend of the relationship, that is, for the models B, C and E, the third and fourth elements are $O_p\left(T^3\right)$, and after rescaling by T^{-1}, they provoke that $\hat{\sigma}_e^2$ is $O_p\left(T^2\right)$. Finally, for the model D we can see that both elements of the previous expression are $O_p\left(T^2\right)$ so that $\hat{\sigma}_e^2$ is $O_p(T)$.

The covariance of order $s \geq 1$ of the disturbance term can be computed from:

$$
\begin{aligned}
Cov\left(\hat{e}_t\hat{e}_{t-s}\right) &= T^{-1}\sum_{t=s+1}^{T}\left(\left(\varepsilon_t - z_t(\lambda)\hat{\beta} + z_t(\tau)\beta\right)'\left(\varepsilon_{t-l} - z_{t-s}(\lambda)\hat{\beta} + z_{t-s}(\tau)\beta\right)\right) \\
&= T^{-1}\sum_{t=s+1}^{T}\left(\varepsilon_t\varepsilon_{t-s} - \varepsilon_t z_{t-s}(\lambda)\hat{\beta} + \varepsilon_t z_{t-s}(\tau)\beta - \hat{\beta}' z_t'(\lambda)\varepsilon_{t-s}\right. \\
&\quad +\hat{\beta}' z_t'(\lambda)z_{t-s}(\lambda)\hat{\beta} - \hat{\beta}' z_t'(\lambda)z_{t-s}(\tau)\beta \\
&\quad \left. +\beta' z_t'(\tau)\varepsilon_{t-s} - \beta' z_t'(\tau)z_{t-s}(\lambda)\hat{\beta} + \beta' z_t'(\tau)z_{t-s}(\tau)\beta\right) \\
&= T^{-1}\sum_{t=s+1}^{T}\left(\varepsilon_t\varepsilon_{t-s} - \varepsilon_t z_{t-s}(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)\left(z(\tau)\beta + \varepsilon\right)\right. \\
&\quad +\varepsilon_t z_{t-s}(\tau)\beta - \beta' z'(\tau)z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z_t'(\lambda)\varepsilon_{t-s} \\
&\quad -\varepsilon' z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z_t'(\lambda)\varepsilon_{t-s} \\
&\quad +\beta' z'(\tau)z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z_t'(\lambda)z_{t-s}(\lambda) \\
&\quad \left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)\left(z(\tau)\beta + \varepsilon\right) \\
&\quad +\varepsilon' z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z_t'(\lambda)z_{t-s}(\lambda) \\
&\quad \left(z'(\lambda)z(\lambda)\right)^{-1} z'(\lambda)\left(z(\tau)\beta + \varepsilon\right) \\
&\quad -\beta' z'(\tau)z(\lambda)\left(z'(\lambda)z(\lambda)\right)^{-1} z_t'(\lambda)z_{t-s}(\tau)\beta
\end{aligned}
$$

$$
\begin{aligned}
&-\varepsilon' z(\lambda)\left(z'(\lambda) z(\lambda)\right)^{-1} z_t'(\lambda) z_{t-s}(\tau)\beta \\
&+\beta' z_t'(\tau)\varepsilon_{t-s} \\
&-\beta' z_t'(\tau) z_{t-s}(\lambda)\left(z'(\lambda) z(\lambda)\right)^{-1} z'(\lambda) z(\tau)\beta \\
&-\beta' z_t'(\tau) z_{t-s}(\lambda)\left(z'(\lambda) z(\lambda)\right)^{-1} z'(\lambda)\varepsilon \\
&\left.+\beta' z_t'(\tau) z_{t-s}(\tau)\beta\right).
\end{aligned}
$$

Following the reasoning adopted for the variance, we can see that it is the last element of the previous expression the one that gives the rate of divergency depending on the model under consideration. Thus, for the models that do not include a shift in the deterministic trend, the models An and A, this element diverges at a rate $O_p(1)$. For those models that allow for a shift in the trend, that is, for the models B, C and E the covariances diverges at a rate $O_p\left(T^2\right)$. For the model D the covariances diverges at a rate $O_p(T)$. Therefore, taking into account these and the previous results it has been proved that the estimation of the long-run variance diverges at $O_p(1)$ for the models An and A, at a rate $O_p(T)$ for the models D, and at a rate $O_p\left(T^2\right)$ for the models B, C and E.

Using these results it can be established that the test statistic behaves according to the following statements:

$$
\begin{aligned}
SC_{An}(\lambda) &= \frac{O_p(T)}{O_p(1)} = O_p(T); \\
SC_{A}(\lambda) &= \frac{O_p(T)}{O_p(1)} = O_p(T); \\
SC_{B}(\lambda) &= \frac{O_p\left(T^3\right)}{O_p\left(T^2\right)} = O_p(T); \\
SC_{C}(\lambda) &= \frac{O_p\left(T^3\right)}{O_p\left(T^2\right)} = O_p(T); \\
SC_{D}(\lambda) &= \frac{O_p\left(T^2\right)}{O_p(T)} = O_p(T); \\
SC_{E}(\lambda) &= \frac{O_p\left(T^3\right)}{O_p\left(T^2\right)} = O_p(T),
\end{aligned}
$$

that is to say, the test diverge when the break point is incorrectly estimated at a rate $O_p(T)$. Therefore, this shows that using the minimum functional provides a consistent estimation of the break point. Thus, Theorem 4 has been proved. ■

A.5 Proof of Theorem 5

The model defined by (1), (2) and (3) can be alternatively expressed in the following way:

$$
y_t = g_i(t) + x_t'\beta + \alpha_t + \varepsilon_t; \quad \text{(A-9)}
$$

$$\begin{aligned} x_t &= x_{t-1} + \varsigma_t, \\ \alpha_t &= \alpha_{t-1} + \eta_t, \end{aligned}$$

with $g_i(t), i = \{An, A, B, C, D, E\}$, given as in section 2. Estimated residuals of (A-9) can be computed from:

$$\hat{v}_t = y_t - \hat{y}_t = \alpha_t - z_t (z'z)^{-1} z'\alpha + \varepsilon_t - z_t (z'z)^{-1} z'\varepsilon,$$

where $v_t = r_t + \varepsilon_t$. Once the suitable rescaling matrix has been defined, the limit distribution of partial sum process only depends on the terms involving $\{\alpha_t\}$. Thus, is straightforward to see that:

$$T^{-3/2}\hat{S}_t = T^{-3/2}\sum_{j=1}^{[bT]}\alpha_j - T^{-3/2}\sum_{j=1}^{[bT]} z_j PA (A'Pz'zPA)^{-1} A'Pz'\alpha + o_p(1),$$

where $\hat{S}_t = \sum_{j=1}^{[bT]}\hat{v}_j$. Notice that since we are under the alternative hypothesis $T^{-3/2}\sum_{j=1}^{[bT]}\alpha_j \Rightarrow \sigma_\eta \int_0^b W(s)\,ds$. For the model E and for those moments of time such that $0 < b < \lambda$, the partial sum process of the matrix of regressors tends to:

$$T^{-1/2}\sum_{j=1}^{[bT]} z_j PA \Rightarrow \begin{bmatrix} b & 0 & b^2/2 & 0 & \int_0^b W_{2k}'(s)\,ds & 0 \end{bmatrix},$$

whereas for $\lambda < b < 1$:

$$T^{-1/2}\sum_{j=1}^{[bT]} z_j PA \Rightarrow \begin{bmatrix} b & (b-\lambda) & b^2/2 & (b-\lambda)^2\big/2 & \int_0^b W_{2k}'(s)\,ds & \int_\lambda^b W_{2k}'(s)\,ds \end{bmatrix}.$$

Notice that in the limit, $T^{-1}A'Pz'\alpha$ converges to the (6×1)-vector noted as J, that is, $T^{-1}A'Pz'\alpha \Rightarrow J$, now with elements defined by $J_1 = \sigma_\eta \int_0^1 W(s)\,ds$, $J_2 = \sigma_\eta \int_\lambda^1 W(s)\,ds$, $J_3 = \sigma_\eta \int_0^1 sW(s)\,ds$, $J_4 = \sigma_\eta \int_\lambda^1 (s-\lambda) W(s)\,ds$, $J_5 = \sigma_\eta \int_0^1 W_{2k}(s) W(s)\,ds$ and $J_6 = \sigma_\eta \int_\lambda^1 W_{2k}(s) W(s)\,ds$, provided that $\{\varsigma_t\}$ and $\{\eta_t\}$ are mutually independent. Using these results, we have that the partial sum processes when $0 < b < \lambda$ converge to:

$$\begin{aligned} T^{-3/2}\hat{S}_t &\Rightarrow \sigma_\eta \left(\int_0^b W(s)\,ds - \begin{bmatrix} b & 0 & b^2/2 & 0 & \int_0^b W_{2k}'(s)\,ds \end{bmatrix} H^{-1} J \right) \\ &= \sigma_\eta K_1(\lambda, b), \end{aligned}$$

whereas for $\lambda < b < 1$:

$$\begin{aligned} T^{-3/2}\hat{S}_t &\Rightarrow \sigma_\eta \left(\int_0^b W(s)\,ds - \begin{bmatrix} b & (b-\lambda) & b^2/2 & (b-\lambda)^2\big/2 & \int_0^b W_{2k}'(s)\,ds \end{bmatrix} H^{-1} J \right) \\ &= \sigma_\eta K_2(\lambda, b). \end{aligned}$$

Now is easy to see that under the alternative hypothesis:

$$T^{-4} \sum_{t=1}^{T} \hat{S}_t^2 \Rightarrow \sigma_\eta^2 \left[\int_0^\lambda K_1^2 (\lambda, b)\, db + \int_\lambda^1 K_2^2 (\lambda, b)\, db \right],$$

so that the numerator of (6) is $O_p\left(T^2\right)$. Consider now the denominator of (6). As shown by Kwiatkowski et al. (1992), the long-run variance estimator under the alternative hypothesis is $O_p\left(Tl\right)$ provided that:

$$(lT)^{-1} \hat{\omega}_1{}^2 \Rightarrow K\sigma_\eta^2 \left[\int_0^\lambda \underline{W}(s)_1^2\, db + \int_\lambda^1 \underline{W}(s)_3^2\, db \right],$$

where

$$\begin{aligned} \underline{W}(s)_1 &= W(s) - \left[\begin{array}{ccccc} b & 0 & b^2/2 & 0 & \int_0^b W_{2k}'(s)\, ds \end{array} \right] H^{-1} J; \\ \underline{W}(s)_2 &= W(s) - \left[\begin{array}{ccccc} b & (b-\lambda) & b^2/2 & (b-\lambda)^2 \big/ 2 & \int_0^b W_{2k}'(s)\, ds \end{array} \right] H^{-1} J, \end{aligned}$$

and $K = \int_{-1}^{1} k(s)\, ds$. Hence, under the alternative hypothesis (6) converges to the expression given in Theorem 5. ■

References

Ahn, S. K. and G. C. Reinsel, "Estimation for Patially Non-Stationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 1990, (85), 813–823.

Bai, J. and P. Perron, "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 1998, *66* (1), 47–78.

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

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

Bewley, R. and M. Yang, "Tests for Cointegration Based on Canonical Correlation Analysis," *Journal of the American Statistical Association*, 1995, (90), 990–996.

Bierens, H., "Nonparametric Cointegration Analysis," *Journal of Econometrics*, 1997, (77), 379–404.

Bossaerts, P., "Common Nonstationary Components of Asset Prices," *Journal of Economic Dynamics and Control*, 1988, (12), 347–364.

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

Fernández, J. L. and R. Peruga, "Análisis Comparativo de Tres Estadísticos para la Contrastación de Inestabilidad Paramétrica en Relaciones de Cointegración," Technical Report, Departamento de Economía Cuantitativa. Universidad Complutense de Madrid 1996.

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

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

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

___ **and** ___ , "Tests for Cointegration in Models with Regime and Trend Shifts," *Oxford Bulletin of Economics and Statistics*, 1996, (58, 3), 555–560.

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

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

___ **and P. C. B. Phillips**, "Estimation and Inference in Models of Cointegration: A

Simulation Study," *Advances in Econometrics*, 1990, (8), 225–248.

Hao, K., "Testing for Structural Change in Cointegrated Regression Models: some Comparisons and Generalisations," *Econometric Reviews*, 1996, (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.

Herrndorf, N., "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 1984, (12), 141–153.

Johansen, S., "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 1988, (12), 231–254.

——, "Estimation and Hypothesis Testing on Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 1991, (60), 1551–1580.

——, "Determination of Cointegration Rank in the Presence of a Linear Trend," *Oxford Bulletin of Economics and Statistics*, 1992, (54), 383–397.

——, "The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Time Series," *Econometric Reviews*, 1994, (13), 205–231.

——, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, 1995.

—— **and K. Joselius**, "Maximum Likelihood Estimation and Inference in Cointegration -with Applications to the Demand of Money," *Oxford Bulletin of Economics and Statistics*, 1990, (52), 169–210.

Kwiatkowski, D., P. C. B. Phillips, P. J. Schmidt, and Y. Shin, "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*, 1992, (54), 159–178.

Lee, J., "Minimum Statistics Testing for Stationarity in the Presence of a Structural Break," Technical Report 97-W03, Department of Economics and Business Administration. Vanderbilt University 1996.

Leybourne, S. J. and B. P. M. McCabe, "A Simple Test for Cointegration," *Oxford Bulletin of Economics and Statistics*, 1993, (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, "A Parametric Approach to Testing the Null of Cointegration," *Journal of Time Series Analysis*, 1997, (18, 4), 395–413.

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**, "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 1988, (4), 468–498.

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

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

___ , "Further Evidence on Breaking Trend Functions in Macroeconomic Variables," *Journal of Econometrics*, 1997, (80), 355–385.

___ **and J. Campbell**, "A Note on Johansen's Cointegration Procedure when Trends are Present," *Empirical Economics*, 1993, (18), 777–789.

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

___ , "Spectral Regression for Cointegrated Time Series," in W. Barnett, ed., *Nonparametric and Semiparametric Methods in Economics and Statistics*, Cambridge University Press, 1991.

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

___ **and M. Loretan**, "Estimating Long Run Economic Equilibria," *Review of Economic Studies*, 1991, (58), 407–436.

___ **and S. N. Durlauf**, "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 1986, (53), 473–496.

___ **and S. Ouliaris**, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 1990, (58), 165–193.

Quintos, C., "Fully Modified Vector Autoregressive Inference in Partially Nonstationary Models," *Journal of the American Statistical Association*, 1998, (93, 442), 783–795.

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

___ **and R. Luukkonen**, "Testing Cointegration in Infinite Order Vector Autoregressive Processes," *Journal of Econometrics*, 1997, (81), 93–129.

Shin, Y., "A Residual-based Test of the Null of Cointegration against the Alternative of no Cointegration," *Econometric Theory*, 1994, (10), 91–115.

Stock, J. H., "Asymptotic Propierties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 1987, (55), 1035–1056.

___ **and M. W. Watson**, "Testing for Common Trends," *Journal of the American Statistical Association*, 1988, (83), 1097–1107.

___ **and** ___ , "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 1993, (61, 4), 783–820.

Yang, M. and R. Bewley, "On Cointegration Tests for VAR Models with Drift," *Economics Letters*, 1995, (51), 45–50.

Yap, S. and G. C. Reinsel, "Estimation and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model," *Journal of the American Statistical Association*, 1995, (90), 253–267.

Zivot, E. and D. W. K. Andrews, "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit-Root Hypothesis," *Journal of Business & Economic Statistics*, 1992, (10, 3), 251–270.