Cointegration testing in dependent panels with breaks

Di Iorio, Francesca and Fachin, Stefano

University of Rome "La Sapienza"

9 May 2007

Cointegration Testing in Dependent Panels with Breaks

Francesca Di Iorio
University of Naples Federico II

Stefano Fachin
University of Rome "La Sapienza"

Abstract

In this paper we propose panel cointegration tests allowing for breaks and cross-section dependence based on the Continuos-Path Block bootstrap. Simulation evidence shows that the proposed panel tests have satisfactory size and power properties, hence improving considerably on asymptotic tests applied to individual series. As an empirical illustration we examine investment and saving for a panel of European countries over the 1960-2002 period, finding, contrary to the results of most individual tests, that the hypothesis of a long-run relationship with breaks is compatible with the data.

Keywords: Panel cointegration, continuos-path block bootstrap, breaks, Feldstein-Horioka Puzzle.
JEL codes: C23, C15

Contacts:
Francesca Di Iorio: fdiiorio@unina.it;
Stefano Fachin: s.fachin@caspur.it.

January 2007

1 Introduction[1]

Since Engle and Granger's seminal paper, the concept of cointegration has spurred an enormous amount of both theoretical results and applied work. However, the two strands of literature are typically divided by a serious gap, with the theory assuming large sample sizes very rarely available in practice. One way to fill this gap is to develop small sample methods (*inter alia*, Li and Maddala, 1997, Johansen, 2000, 2002, Omtizgt and Fachin, 2006). The other, obviously, is to increase the information content of the datasets analysed. Since increasing the sampling frequency is of little help for the goal of uncovering long-run structures (Shiller and Perron, 1985) this implies either waiting for years to pass or adding an extra dimension to the data: hence the growing interest in non-stationary panels and the new concept of panel cointegration (for a thorough discussion of the latter see Pedroni, 2004). While helping to solve one empirical problem this solution however created a new theoretical one, namely how to do inference on panel statistics computed from dependent units. First generation panel cointegration procedures simply ignored the issue, which is the focus of the most recent contributions to the debate (for a recent review see Breitung and Pesaran, 2006). In this vein, the aim of this paper is to extend to dependent panels the Gregory and Hansen (1996), henceforth GH, cointegration tests with structural breaks in the cointegrating relationship. Structural breaks procedures (which require recursive estimation) are very demanding in terms of information set, so that they can perform particularly poorly in small samples: according to the simulation results reported by GH, with 50 observations both positive size bias and very low power have to be expected. Hence, the properties of panel versions of this type of tests are a topic of obvious interest, which has not received a fully satisfactory treatment in the still very young literature on non-stationary panels. Under the cross-section independence assumption we find Gutierrez (2005) and Westerlund (2006a,b); the former proposes to combine the p-values of GH tests computed for the individual units, while the latter derives panel cointegration tests allowing for breaks in the deterministic kernel (level and trend) of the cointegration regression. Acknowledging the importance of the dependence issue, Westerlund (2006b) proposes a bootstrap procedure complementing the asymptotic test mentioned above. However, some caution is suggested by the fact that this procedure is based upon simple resampling of the FMOLS or DOLS cointegrating residuals: these are weakly dependent if cointegration holds, and non-stationary if it

[1] Financial support from the Department of Statistics of the University of Naples Federico II, University of Rome "La Sapienza" and MIUR is gratefully acknowledged. We are grateful to Anindya Banerjee and Josep Carrion-i-Silvestre for kindly providing the investment and savings data set. Correspondence to: s.fachin@caspur.it, fdiiorio@unina.it.

does not[2]. Taking a completely different route, both Banerjee and Carrion-i-Silvestre (2006) and Westerlund and Edgerton (2006b) model dependence with a common factor approach. Unfortunately, in both cases some restrictive assumptions on either the breaks or the form of cross-section dependence are needed. More precisely, Banerjee and Carrion-i-Silvestre's asymptotics requires the break to fall in the same period in all units, while Westerlund and Edgerton's excludes any sort of short- and long-run dependence of the explanatory variables across units. Further, in both cases large sample sizes are required. Westerlund and Edgerton's LM test has acceptable power for $T = 200$ and $N = 20$, but it is disappointing (power generally lower than 50%) for $T = 100$. In Banerjee and Carrion-i-Silvestre's framework the usual single-equation definition of cointegration (stationary residuals in the cointegrating equation) is accepted if the null hypothesis of non-stationarity is rejected both for the estimated common factor and the idiosyncratic residuals, an event which in their simulations has a frequency generally much lower than 50% even with rather large sample sizes such as $T = 100$ and $N = 40$.

Hence, we must conclude that a panel cointegration test fully robust to possibly heterogenous breaks and cross-section dependence and with an acceptable small sample performance is not available yet. Building on Fachin (2007), our proposal is to exploit Paparoditis and Politis' (2001) Continuous-Path Block Bootstrap (CBB). As we will see, the proposed procedure can account for any form of dependence and, to some extent, heterogenous breaks, delivering satisfactory small sample size and power properties which may prove of considerable help in applied work.

We shall now in section 2 introduce the set-up and outline the testing procedure and in section 3 present the design and results of a Monte Carlo experiment. An empirical illustration on the relationship between the investment/GDP and savings/GDP ratios, the so-called Feldstein-Horioka puzzle, already examined by Banerjee and Carrion-i-Silvestre (2004) and Gutierrez (2005), is presented in section 4. Some conclusions and suggestions for future research are finally discussed in section 5.

2 Bootstrap Panel Cointegration Testing with Breaks: Set-up

Let us consider for simplicity a standard bivariate panel cointegration set-up, with the right- and left-handside variables, denoted as usual by X and

[2] Incidentally, very much the same remark applies to the panel cointegration bootstrap test proposed in Westerlund and Edgerton (2006a), where an AR model is fitted to the cointegration residuals in order to carry out the sieve bootstrap. If cointegration does not hold the procedure hinges crucially upon precise estimation of the unit root, a notoriously difficult task (on this issue, see Fachin, 2004).

Y, observed over N units and T time periods, indexed respectively by i and t. In each unit X and Y are linked by a linear, not necessarily cointegrating, relationship with a slope break in period t_i^b. A panel cointegration test allowing for breaks may be defined very simply, following Pedroni's (1999) group mean test approach, as the mean of cointegration statistics with breaks computed for the individual units, *i.e.* the minimum of a sequence of tests computed for all possible breakpoints[3]. Similarly to the case of panel cointegration tests, the bootstrap is a natural candidate for solving the problem of inference under the general set-up of dependent units, with the additional advantage of allowing the use of a robust statistics such as a median, untractable for convential asymptotic methods. To this end, we need to design a resampling scheme delivering pseudodata:

(i) reproducing both the autocorrelation and cross-correlation properties of the data;

(ii) accounting for the break;

(iii) obeying the null hypothesis of no cointegration.

As mentioned above, as in Fachin (2007), the algorithm we propose is based upon Paparoditis and Politis (2001) "Continuous-Path Block Bootstrap" (CBB), a resampling scheme designed to construct non-stationary pseudo-series from data of the same type. Requirement (i) is readily satisfied by resampling together all the units, while (iii) by resampling separately the X's and the Y's, thus constructing pseudodata matching a given time observation for the former with a different one for the latter. We thus need to discuss only (ii). The key point here is that our implementation of the CBB assumes that the series are non-stationary with a constant drift, which can then be consistently estimated. Now, if X is $I(1)$ with drift, this assumption is easily seen to be violated by a slope change in the cointegrating equation, equivalent to a change in the drift of the $\Delta Y's$. Hence, we have to resample the data separately before and after the break. Since to satisfy (i) the resampling scheme is applied to all columns for a given block of time observations, this requirement introduces a complication in our procedure: unless the breaks happen to fall all in the same period we need to impose the constrain $\widehat{t}_i^b = \widehat{t}^b \; \forall i$. A natural choice is $\widehat{t}^b = median(\widehat{t}_1^b, \ldots, \widehat{t}_N^b)$. To account for the estimation error in the $\widehat{t}_i^{b\prime} s$ caused by this constraint we will take the pseudodata of the block centred on $\widehat{t}^b$ to be equal to the actual data.

[3] Two remarks are in order here. First, for simplicity we will refer to a summary statistic of the individual cointegration tests as a "panel test", although in Pedroni's terminology this term is reserved for tests obtained imposing an homogeneity assumption. Second, although GH examined all most common residual-based no cointegration statistics, Phillips' Z_α and Z_t and the ADF, in this paper we will concentrate, without much loss of generality but some computational advantage, only on the latter.

Summing up, the bootstrap testing procedure we propose to implement is the following:

1. compute for each unit $i = 1, \dots, N$, of the data set under study, $\{X_1 \dots X_N, Y_1 \dots Y_N\}_{t=1}^{T}$ the no-cointegration Gregory-Hansen statistic $Min(\widehat{ADF_i}) = \inf ADF_i(t)$ for $t \in [\tau_1 T, \tau_2 T]$; the trimming coefficients τ_1, τ_2 must chosen to ensure stability of the statistic at the endpoints;

2. compute the panel cointegration statistic as, *e.g.*, the mean or median of the N individual statistics (respectively, $\widehat{G}_m = \sum_{i=1}^{N} Min(\widehat{ADF_i})/N$) and $\widehat{G}_{me} = median(Min(\widehat{ADF_1}), \dots, Min(\widehat{ADF_N}))$;

3. estimate the breakpoints for each unit; a natural choice is the breakpoint $\widehat{t^b_i}$ associated to $Min(\widehat{ADF_i})$;

4. estimate a common breakpoint $\widehat{t^b}$, e.g., $\widehat{t^b} = median(\widehat{t^b_1}, \dots, \widehat{t^b_N})$;

5. apply the CBB with block length s (assumed to be even; the choice of s is discussed in more detail below) separately to the X's and the Y's over the intervals before ($[1, \dots, \widehat{t^b} - \frac{s}{2} - 1]$) and after the break ($[\widehat{t^b} + \frac{s}{2} + 1, \dots, T]$), obtaining four matrices of pseudodata: for the X's $\{X_1^* \dots X_N^*\}_{t=1}^{\widehat{t^b} - \frac{s}{2} - 1}$ and $\{X_1^* \dots X_N^*\}_{t=\widehat{t^b} + \frac{s}{2} + 1}^{T^*}$, for the Y's $\{Y_1^* \dots Y_N^*\}_{t=1}^{\widehat{t^b} - \frac{s}{2} - 1}$ and $\{Y_1^* \dots Y_N^*\}_{t=\widehat{t^b} + \frac{s}{2} + 1}^{T^*}$; note that $T^* < T$, as some observations are lost in the chaining;

6. construct the pseudo-dataset for the entire time sample joining the pseudodata constructed before and after the break with a central block of actual data: $\mathbf{X}_i^* = \left[x_{i1}^* \dots x_{i\widehat{t^b} - \frac{s}{2} - 1}^* : x_{i\widehat{t^b} - \frac{s}{2}} \dots x_{i\widehat{t^b} + \frac{s}{2}} : x_{i\widehat{t^b} + \frac{s}{2} + 1}^* \dots x_{iT^*}^* \right]'$, $i = 1, \dots, N$, and analogously for the Y's;

7. compute the Group statistics G_h^* ($h = m, me$; see steps 1-2) for the pseudo-data set,
   $\{X_1^* \dots X_N^*, Y_1^* \dots Y_N^*\}_{t=1}^{T^*}$;

8. repeat steps (5) to (7) a large number (say, B) of times;

9. compute the boostrap significance level; assuming that the rejection region is the left tail of the distribution, $p^* = prop(G_h^* < \widehat{G}_h), h = m, me$.

Although exploratory simulations showed the results to be quite robust to the choice of block length, in principle this is a critical point of the algorithm. Here for computational convenience we applied a simple rule-of-thumb, fixing it at $T/10$. In future work we plan to implement Politis and White's (2003) algorithm.

3 Monte Carlo Experiment

3.1 Design

We will base our simulations on a Data Generation Process (DGP) which is essentially a generalisation of the classical bivariate Engle and Granger (1987) DGP to the case of dependent panels; it is very similar to that considered by Kao (1999), and it has been recently adopted by Fachin (2007). The two variables of interest, X and Y, are linked by a possibly cointegrating, breaking relationship:

$$x_{it} = u^x_{it} \tag{1a}$$

$$y_{it} = \begin{cases} \mu_{0i} + \beta_0 x_{it} + u^y_{it}, & t \leq t^b_i \\ \mu_{1i} + \beta_1 x_{it} + u^y_{it}, & t > t^b_i \end{cases} \tag{1b}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$. Both errors u^j, $j = x, y$, are assumed to be the linear combination of a common component, $f^j, j = x, y$, and an idiosyncratic one, ϵ^j, $j = x, y$,:

$$\begin{cases} u^x_{it} = \gamma^x_i f^x_t + \epsilon^x_{it} \\ u^y_{it} = \gamma^y_i f^y_t + \epsilon^y_{it} \end{cases} \tag{2}$$

$$\begin{cases} \epsilon^x_{it} = e^x_{it} + \theta \\ \epsilon^y_{it} = \phi_i \epsilon^y_{it-1} + e^y_{it} \end{cases} \tag{3}$$

where the $e^j_{it} \sim N(0, \sigma^2_{ij})$, $j = x, y$, are white noise.

To have a fully general dependence set-up we assume f^x to be non-stationary, thereby inducing cointegration across units in the X's as well as in the Y's in case of cointegration between X_i and Y_i ($|\phi_i| < 1$)[4]. In this case the common factor f^y is stationary and induces short-run correlation between the Y's across units, while when cointegration does not hold ($\phi_i = 1$) it is non-stationary so that cross-units cointegration in the Y's still holds. Summing up:

$$f^x_t = \sum_{s=1}^{t} \psi^x_s \tag{4}$$

[4] This set-up is likely to be representive of many empirical applications: for instance, in the case of regional consumption and income f^x will be the national stochastic GDP trend causing income, and hence consumption, to be cointegrated across regions.

and

$$f_t^y = \begin{cases} \psi_t^y & if\ \phi_i = 1\ \forall i \\ \sum_{s=1}^t \psi_s^y & else \end{cases} \tag{5}$$

with $\psi_t^j \sim N(0,1)$, $j = x, y$. Note that for simplicity we are ruling out the possibility of cointegration holding in some units only, but the design could be easily generalised further to include this case also.

The simulation framework outlined above is very complex, and the tests to be evaluated computationally demanding (this issue is discussed in more detail below). Hence, rather than aiming at the unfeasible task of a complete design we will define as a base case an empirically relevant set-up and then explore a few interesting variations. Let us first discuss the design parameters common to all experiments. Similarly to GH we take the model as correctly specified; with no loss of generality we set both constant and slope to 2 before the break, with the slope halved after it. The factor loadings are chosen so to ensure substantial cross-correlation in the Y's and cross-cointegration of the the X's: $\gamma_i^j \sim Uniform(-1,6)\ \forall i, j$. In the power simulations we consider on the average rather slow adjustment to equilibrium, with some heterogeneity: $\phi_i \sim Uniform(0.6, 0.8)$. Finally, further heterogeneity across units is given by the noise variances: $\sigma_{ij}^2 \sim Uniform(0.5, 1.5),\ \ j = x, y$. Given the rather short time series analysed in most experiments, in order to ensure computational stability the trimming coefficient is never smaller than 25%. Let us now examine the various experiments (six altogether) in some detail.

1. *Base case*: $T = 40$, $N = 5, 10, 20, 40$; break date Uniform over units in $[0.5T \pm 3] = [17, 23]$. The time span is medium in terms of annual data, but pretty small with a quarterly frequency, so to make it relevant for actual empirical applications. It is definitelly smaller than those considered in the simulation studies on the other cointegration tests with breaks available in the literature. The breaks are distributed over six periods centred in the middle of the time sample, with the testing procedure searching over the interval $[14, 26]$, corresponding to a 35% trimming at both ends of the sample. The drift θ in ϵ_{it}^x is set to zero both in the DGP and, following GH's assumption of correct model specification, in the estimation.

2. *Base case - Wide search:* as Base case, with the testing procedure searching over the interval $[10, 30]$, corresponding to 25% trimming. The aim of this exercise is to assess the importance of the choice of the trimming parameter.

3. *Large time sample*: $T = 160$, $N = 5, 10, 20, 40$; break date Uniform over units in $[0.5T \pm 3] = [77, 83]$. The time span is long in terms of

annual data, but medium with a quarterly frequency, so that it is still relevant for actual empirical applications. Since we need the results from this experiment to be closely comparable to those from the Base case we mimick a situation in which more observations (more precisely, 60) become available at both ends of the sample; hence, the breaks are distributed, and the search takes place, over the same sets of periods centred in the middle of the time sample.

4. *Drift in X*: as Base case, with drift in ϵ_{it}^{x} set to $\theta = 0.01$. This case is designed to evaluate if, similarly to the unit root tests with trend breaks studied by Ioannidis (2005), the estimation of the drift term is critical for our bootstrap algorithm.

5. *Late break*: as Base case - Wide Search, with break date Uniform over units in $[0.75T \pm 3] = [27, 33]$. Note that this a demanding set-up, as half of the interval in which the breaks may fall is outside the searching interval.

A critical point of the bootstrap algorithm described above is the partition of the samples before and after the estimated breakpoint, constrained to be homogenous across units at the median of the individual estimates. This is intuitively acceptable if all units are affected by breaks stemming from a common cause. However, even assuming each unit to be affected by at most one break over the period of interest, these may be widely dispersed over units, for instance because they stem from different causes. The following case is designed to investigate this scenario:

6. *Twin breaks:* as Base case - Wide Search with break date Uniform in $[0.3T \pm 3] = [9, 15]$ in the odd-numbered units, and in $[0.6T \pm 3] = [25, 31]$ in the even-numbered ones. Note that in both cases the break may take place marginally outside the search interval, $[10, 30]$.

The last issue to be discussed is the number the number of Monte Carlo replications. In all simulation exercises this is chosen trying to strike a balance between the contrasting requirements of precision in the results and control of the cost and time scale of the experiment. Here this balance is particularly difficult to achieve because of the combined effects of the the panel structure of the data and the recursive nature of the statistics evaluated: the number of loops executed is the product of bootstrap redrawings, units, periods included in the searching interval, and number of Monte Carlo replications. With 1000 redrawings, 40 units and search over 20 periods the product of first three terms is equal to 800.000; fixing the Monte Carlo replications to 1000 will thus require the execution of 800 million loops for each experiment, with an approximate confidence interval $p \pm 2\sqrt{p(1-p)/1000}$ for, *e.g.*, 5% equal to $[3.6\%, 6.4\%]$. Although this interval may appear not very precise, even improving it marginally to $[4.0\%, 6.0\%]$ requires doubling the number of replications from 1000 to 2000 and thus the total number of loops to the considerable figure of 1.6 billion. We thus decided that 1000 replications is a reasonable choice.

3.2 Results

The results are reported in Tables 1-6 below. In the first column (label "Asy") we reported the mean rejection rates from individual GH tests run on all units. Although, as we will see, the comparison with the panel tests leads to clear cut conclusions in all cases, it should be remarked that it is strictly speaking legitimate only with the panel tests with al 40 units, as in the other cases the units involved are not the same. Hence, in those cases the comparison should be taken as merely suggestive of the relative performances that can be expected.

The results for the Base case ($T = 40,\ N = 5, 10, 20, 40$) are fairly similar for both searching intervals (Tables 1 and 2). Consistently with GH, we find that the asymptotic test on individual series is considerably biased against the null, with power unsurprisingly high. On the contrary, both the mean and median panel cointegration bootstrap tests are somehow slightly undersized but deliver high power both with the mean and median versions, provided the significance level and cross-section sample size are not too small (in practice, in our simulations $\alpha \geq 5\%$ or $N \geq 10$). Hence, in this Base case exploiting the panel dimension using the proposed bootstrap procedure seems to provide a considerable improvement with respect to individual time series analsysis with small time samples. With a larger time sample (Table 3) the Type I errors of the bootstrap panel tests essentially converge to nominal levels taking into account Monte Carlo estimation error, while the GH tests on individual time series still overreject considerably. In both cases power is always close to 100% . When a drift term is allowed in the DGP of the x's (Table 4) the rejection rates of the bootstrap panel tests increase under both H_0 and H_1, with units 6 to 10 causing large overrejection: for a 5% mean test, the Type I error jumps from 3.9% for the first five units to 16.6% for the first 10. The size bias of the asymptotic test also increases, with power somehow adversely affected. The important lessons here is that when using very small cross-section sample sizes great care is needed.

When the breaks take place at the end of the sample (possibily outside the searching interval) we find that the rejection rates of both the GH and the bootstrap panel tests fall under both H_0 and H_1 (Table 5). As far as the the asymptotic test is concerned this has the effect of reducing the size bias, but at the cost of disappointing power (only marginally higher than 50% for a test with an approximate actual 5% size). Obviously, the negative size bias found for panel tests in the Base case (break around the middle of the sample) is even larger here, with the actual size of a 10% test close to 1%. Taking somehow into account the size bias, which advises against use of test with nominal size smaller than 10%, power is nevertheless high for all cross-section sample sizes. Very much the same holds for the panel tests[5] when the breaks are clustered in two different intervals (Table 5), with an additional note of caution on the use of very small cross-section sample sizes (here $N = 5$).

Table 1
Base Case: $T = 40$, N from 5 to 40
Rejection Rates$\times 100$

	N								
	1	5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
				A. Size: $\phi_i = 1\ \forall i$					
1.0	15.8	0.5	0.4	0.6	0.5	0.2	0.4	0.4	0.3
5.0	22.5	2.2	1.8	2.0	1.1	2.6	2.9	4.7	3.4
10.0	28.0	4.9	5.7	6.3	4.3	7.9	9.5	10.3	7.4
				B. Power: $\phi_i \sim Uniform(0.6, 0.8)$					
1.0	93.0	42.4	54.7	63.6	84.6	36.9	47.1	60.5	68.9
5.0	96.7	73.0	80.6	84.2	84.1	70.1	78.5	84.4	85.1
10.0	97.9	83.5	87.3	89.2	88.1	81.3	86.1	88.6	88.0

DGP: eqs. (1a)-(3), $t_i^b \sim Uniform(0.5T \pm 3)$;
search interval: $[0.5T \pm 6]$;
Asy: average rejection rates of invidual no cointegration tests over all 40 units, Gregory and Hansen (1996) asymptotic critical values;
Boot-mean/median: bootstrap test on the mean/median across units of the no cointegration statistics;
Bootstrap: 1000 redrawings, block size $T/10$;
Montecarlo: 1000 replications.

[5] Since this case is of special interest only from the panel point of view the results for the asymptotic tests are not reported, as they will essentially duplicate those from the "Late break" case.

Table 2

Base Case - Wide Search: $T = 40$, N from 5 to 40

Rejection Rates$\times 100$

	N								
	1	5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
	A. Size: $\phi_i = 1\ \forall i$								
1.0	16.4	0.1	0.3	0.3	0.2	0.2	0.1	0.2	0.2
5.0	24.5	1.7	2.4	4.3	2.7	1.1	1.0	0.9	0.4
10.0	31.5	6.3	7.4	8.9	5.6	3.4	3.6	5.3	1.9
	B. Power: $\phi_i \sim Uniform(0.6, 0.8)$								
1.0	94.1	42.3	53.5	68.2	75.9	49.9	58.6	71.9	73.2
5.0	97.4	77.6	86.2	92.5	93.6	81.1	88.1	93.1	93.4
10.0	98.4	90.1	94.7	98.2	97.4	92.1	96.4	98.7	97.4

$t_i^b \sim Uniform(0.5T \pm 3)$;search interval: $[0.5T \pm 10]$.

All other abbreviations and definitions: see table 1.

Table 3

Large time sample: $T = 160$, N from 5 to 40

Rejection Rates$\times 100$

	N								
	1	5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
	A. Size: $\phi_i = 1\ \forall i$								
1.0	10.0	0.8	0.9	0.5	0.5	0.6	0.6	0.6	0.5
5.0	16.4	4.3	4.0	5.5	3.4	4.6	4.3	3.6	2.9
10.0	20.6	8.1	8.8	10.9	7.2	8.8	9.1	8.0	7.1
	B. Power: $\phi_i \sim Uniform(0.6, 0.8)$								
1.0	96.5	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
5.0	98.5	99.9	99.9	100.0	100.0	99.9	99.9	100.0	100.0
10.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

$t_i^b \sim Uniform(0.5T \pm 3)$; search interval: $[0.5T \pm 6]$.

All other abbreviations and definitions: see table 1.

Table 4

Drift: $T = 40$, N from 5 to 40

Rejection Rates$\times 100$

	N								
	1	5	10	20	40	5	10	20	40
α	*Asy*		*Boot-Mean*				*Boot-Median*		
				A. Size: $\phi_i = 1\ \forall i$					
1.0	29.3	0.5	0.6	0.6	0.5	1.2	5.5	1.6	1.0
5.0	39.7	3.9	16.6	6.3	5.6	5.7	27.9	8.1	5.2
10.0	45.8	8.7	36.2	16.7	15.6	11.6	47.1	15.7	13.6
				B. Power: $\phi_i \sim Uniform(0.6, 0.8)$					
1.0	76.1	58.0	82.2	94.0	96.4	65.7	92.0	97.0	97.7
5.0	86.4	92.6	99.9	99.8	99.8	92.8	100.0	99.6	99.9
10.0	90.4	98.2	100.0	100.0	100.0	97.1	100.0	100.0	99.9

$t_i^b \sim Uniform(0.5T\ \pm 3)$; search interval: $[0.5T \pm 6]$.

All other abbreviations and definitions: see table 1.

Table 5

Late break: $T = 40$, N from 5 to 40

Rejection Rates$\times 100$

	N								
	1	5	10	20	40	5	10	20	40
α	*Asy*		*Boot-Mean*				*Boot-Median*		
				A. Size: $\phi_i = 1\ \forall i$					
1.0	2.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5.0	8.0	0.1	0.2	0.0	0.2	0.3	0.5	0.2	0.5
10.0	13.7	0.5	1.5	0.3	0.6	0.9	1.9	0.5	1.0
				B. Power: $\phi_i \sim Uniform(0.6, 0.8)$					
1.0	50.1	30.5	45.2	73.6	89.4	24.4	36.0	71.0	89.6
5.0	64.8	64.8	83.4	94.4	98.9	53.1	77.6	92.7	97.5
10.0	72.7	80.1	94.9	98.1	99.2	71.0	90.6	97.6	99.0

$t_i^b \sim Uniform(0.75T\ \pm 3)$; search interval: $[0.5T\ \pm 6]$

NB:$0.75T + 3 = 33$, $0.5T + 6 = 30$.

All other abbreviations and definitions: see table 1.

Table 6

Twin breaks: $T = 40$, N from 5 to 40

Rejection Rates$\times 100$

	N							
	5	10	20	40	5	10	20	40
α	*Boot-Mean*				*Boot-Median*			
	A. Size: $\phi_i = 1\ \forall i$							
1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5.0	0.3	0.1	0.5	0.4	0.3	0.0	1.2	0.7
10.0	1.3	0.1	0.5	0.4	1.0	0.0	1.9	1.5
	B. Power: $\phi_i \sim Uniform(0.6, 0.8)$							
1.0	24.7	36.1	66.8	71.0	9.9	27.1	68.0	75.6
5.0	52.1	62.7	85.9	89.2	28.6	52.9	86.3	89.7
10.0	70.1	75.8	92.6	94.8	41.5	67.9	91.5	94.3

$$t_i^b \sim \begin{cases} Uniform(0.3T \pm 3) & i = 1, 3, \ldots, N-1 \\ Uniform(0.6T \pm 3) & i = 2, 4, \ldots, N \end{cases}$$

search interval: $[10, 30]$;

NB: $0.3T - 3 = 9, 0.6T + 3 = 31$.

All other abbreviations and definitions: see table 1.

4 Empirical illustration: the Feldstein-Horioka Puzzle

One the major empirical puzzles of contemporary macroeconomics (six altogether according to Obstfeld and Rogoff, 2000) is the so-called Feldstein-Horioka Puzzle, *i.e.* the evidence supporting the existence of a long-run link between the investment (I) and savings (S) to GDP (Y) ratios in advanced economies, where high capital mobility may allow the current account to be unbalanced for long periods. Banerjee and Carrion-i-Silvestre (2004) investigated the issue on a data set including 14 European economies (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, UK) over the period 1960-2002 using panel cointegration tests allowing for a single break in the cointegrating coefficients. Although Banerjee and Carrion-i-Silvestre were not able to reach a clear conclusion, their findings appear on the whole rather favourable to the cointegration-with-break hypothesis, indeed plausible from the plots reported in Figs. 1-2. However, these results may not be entirely reliable, as the bootstrap procedure used (abandoned in the revised version of the paper, Banerjee and Carrion-i-Silvestre, 2006) implied fitting an AR model to a MA process with a unit root under no cointegration. We thus need to explore some alternatives. Confirming the doubts on their empirical limits, unfortunately neither of the cointegration tests with breaks currently available in the literature may be applied. Since savings (the right-hand side variable) are generally correlated in the short-run, and in some cases cointegrated, across economies (results not reported here available on request) the assumptions underlying Westerlund and Edgerton's (2006b) test are not satisfied, and the small sample size as well as the likely presence of heterogenous breaks advises against use of Banerjee and Carrion-i-Silvestre's (2006) procedure. It is thus of some interest to find out the results of applying the procedure proposed in this paper.

As a first step of our analysis we computed ADF tests to check the properties of the series, choosing the order of the autoregression on the basis of the significance of the last lag (maximum four). The results, reported in Table 7, suggest that the Savings/GDP ratio may be stationary in Finland and Portugal. We thus excluded these two countries and proceed to compute the individual and panel cointegration tests. Examining the individual statistics (Table 8; since essentially similar results have been obtained with 25% and 12.5% trimming we report only the latter) we find that, consistently with theoretical expectations and somehow contrary to those formed on the basis of visual inspection of the plots, only in five countries (Netherlands, Denmark, France, Spain and UK) out of 12 the $Min(ADF)$ tests reject the null hypothesis of no cointegration according to the asymptotic critical values. This evidence (or, better, lack of) should however be evaluated keeping in mind the properties of the test with the dataset at hand, which has two features suggesting caution: small size and breaks often falling far from the middle of the sample. Since on the basis of both GH's and our own results very low power has to be expected in these circumstances, the failure to reject cannot be taken as a conclusive piece of evidence. We thus turn to the bootstrap panel cointegration tests, reported in Table 9. The mean p-values are now very small with both 25% and 12.5% trimming (respectively, 3% and 1%), while the median ones are slightly higher but still rather small (respectively, 14% and 10%). While the mean p-values are definitely strongly significant, to correctly evaluate the median ones we should keep in mind that, as already stressed above, the breaks are widely dispersed across units. From our simulations ("Twin breaks" case, Table 6) we know that in these circumstances our panel cointegration tests are likely to be severly undersized: hence, both median p-values should be regarded as significant, and the no cointegration hypothesis for the panel as a whole rejected according to both mean and median criteria and with both trimming values considered.

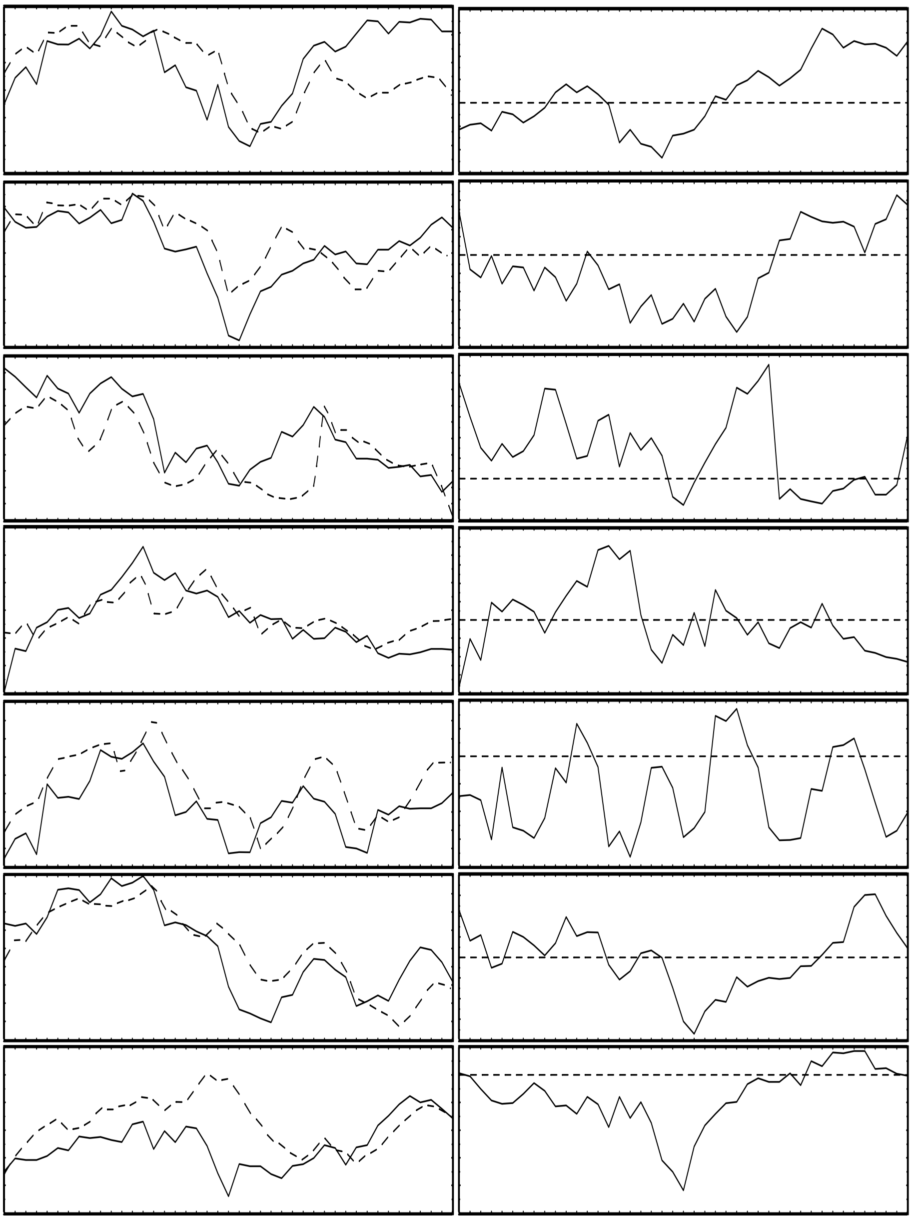

Fig. 1. Savings (S) and Investments (I) to GDP (Y) ratios dynamics, 1960-2002. Left Column: S/Y and I/Y (respectively solid and dotted line, logs). Right Column: Current Account/GDP = $(S-I)/Y$ (solid line) and zero (dotted line). Top to bottom: Austria, Belgium, Denmark, Finland, France, Germany, Greece.

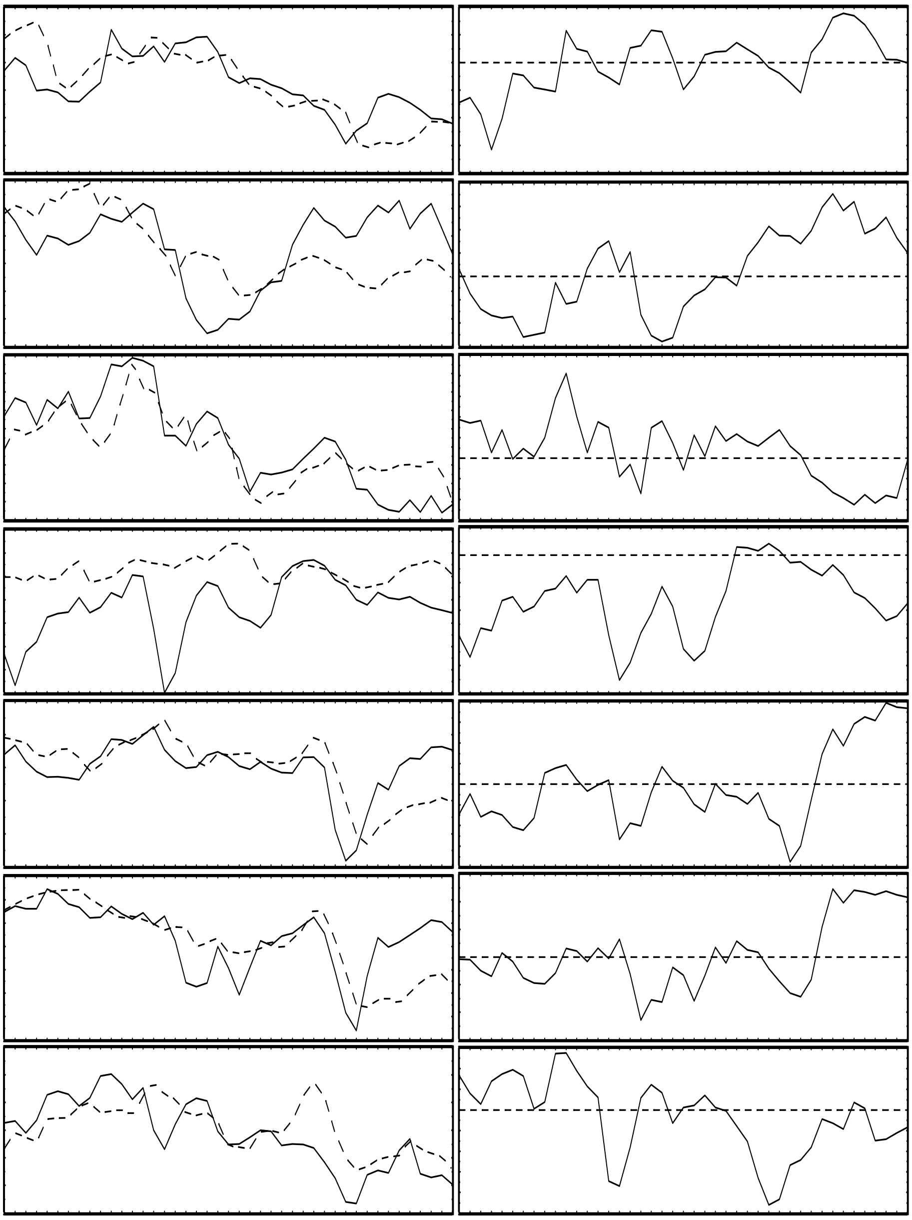

Fig. 2. Savings (S) and Investments (I) to GDP (Y), 1960-2002. Left Column: S/Y and I/Y (respectively solid and dotted line, logs). Right Column: Current Account/GDP $= (S - I)/Y$ (solid line) and zero (dotted line). Top to bottom: Ireland, Italy, Netherlands, Portugal, Spain, Sweden, UK.

Table 7

Investment and Savings to GDP ratios: ADF Unit Root Tests

	Austria	Belgium	Denmark	Finland	France	Germany	Greece
I	-1.31	-1.57	-1.42	-2.16	-1.26	-2.00	-2.13
S	-0.62	-1.46	-2.03	-3.34^{*}	-1.49	-1.66	-1.11
	Ireland	Italy	Netherlands	Portugal	Spain	Sweden	UK
I	-2.34	-1.32	-1.11	-3.42	-2.93	-1.54	-2.16
S	-1.85	-1.29	-1.94	-4.87^{**}	-2.42	-2.06	-0.79

*: significant at 5%; **: 1%.

Table 8

Investment and Savings, 1960-2002
Min(ADF) Cointegration Tests with Break
and estimated breakpoints

	Austria	Belgium	Denmark	France	Germany	Greece
Min(ADF)	-4.28	-4.44	-5.96^{***}	-4.95^{***}	-4.22	-4.45
Break	1991	1982	1990	1994	1994	1999
	Ireland	Italy	Netherlands	Spain	Sweden	UK
Min(ADF)	-3.36	-4.25	-5.50^{***}	-6.01^{***}	-3.97	-4.92^{*}
Break	1987	1981	1976	1979	1970	1996

trimming: 12.5% (searching interval: 1964-1999); *break*: argmin(ADF);
critical values: $1\% : -5.47; 5\% : -4.95; 10\% : -4.68$.
*: significant at 10%; **: 5%;***: 1%.

Table 9

Investment and Savings, 1960-2002
Bootstrap Panel Cointegration Tests

Trimming	*Mean* (p^*)	*Median* (p^*)
12.5%	-4.69 (0.6)	-4.44 (10.2)
25%	-4.62 (2.8)	-4.41 (13.7)

panel: Austria, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Netherlands, Spain, Sweden, UK;
Mean/Median: mean/median of the individual $Min(ADF)$ statistics;
p^*: bootstrap p-values $\times 100$, 1000 redrawings.

5 Conclusions

Testing panel cointegration in dependent panels allowing for breaks at unknown periods is a challenging task, as two forms of dependence (between the tests computed with the break fixed at different periods and for different units) must be accounted for. Building upon Fachin (2007), in this paper we propose to solve this problem using the bootstrap. Simulation results suggest that the proposed panel testing procedures improve considerably on the performances of pure time series Gregory and Hansen (1996) tests. Further, it appears to be more flexible than other available panel cointegration tests with breaks, which assume either homogenous breaks (Banerjee and Carrion-i-Silvestre, 2006) or independence of the explanatory variables (Westerlund and Edgerton, 2006b). These expectations are confirmed by an empirical application to the Feldstein-Horioka Investment-Savings Puzzle for a panel of 12 european countries. Assuming free capital movements there are no reasons to expect any long-run relationship between these two variables to hold, but the visual inspection of the plots suggests that this may have been possibly the case, provided breaks are allowed. The majority of individual Gregory and Hansen tests for individual countries fail to reject the null of no cointegration, thus supporting theretical expectations. However, the bootstrap panel tests overturn this conclusion, supporting the view of a long-run relationship with breaks holding in the panel as a whole and suggesting that the individual tests fail to reject merely because of low power.

Clearly, much work is still needed on several aspects of the test procedure proposed: just to mention a few, data-driven choice of the CBB block size and exploring tests performance with multivariate models or partially cointegrated panels.

6 References

Banerjee, A., Carrion-i-Silvestre J.L. (2004) "Breaking Panel Cointegration", *Mimeo*, European University Institute.

Banerjee, A., Carrion-i-Silvestre J.L. (2006) "Cointegration in Panel Data With Breaks and Cross-Section Dependence", *Working Paper Series* n. 591, European Central Bank.

Breitung, J., Pesaran M. H. (2006) "Unit Roots and Cointegration in Panels" Forthcoming in L. Matyas, and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Kluwer Academic Publishers, Dordrecht (NL).

Engle, R.F., Granger, C.W.J. (1987) "Co-integration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-176.

Fachin, S. (2004) "Bootstrap inference on Fully Modified Estimates of Cointegrating Coefficients: A Comment", *Economics Bulletin*, Vol. 3, No. 13 pp. 1-8.

Fachin, S. (2007) "Long-Run Trends in Internal Migrations in Italy: a Study in Panel Cointegration with Dependent Units", *Journal of Applied Econometrics*, forthcoming.

Gutierrez, L. (2005) "Tests for Cointegration in Panels with Regime Shifts", *Mimeo,* Department of Agricultural Economics, University of Sassari.

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

Ioannidis, E.E. (2005) "Residual-based block bootstrap unit root testing in the presence of trend breaks" *Econometrics Journal*, 8, 323-351.

Johansen, S. (2000) "A Bartlett Correction Factor for Tests on the Cointegrating Relations" *Econometric Theory,* 16, 740-778.

Johansen, S. (2002) "A Small Sample Correction for Tests of Hypotheses on the Cointegration Vectors" *Journal of Econometrics,* 11, 195-221.

Li, H., Maddala, G.S. (1997) "Bootstrapping Cointegrating Regressions" *Journal of Econometrics,* 80, 297-318.

Kao, C. (1999) "Spurious regression and residual-based tests for cointegration in panel data" *Journal of Econometrics,* 90,1-44.

Omtzigt, P., Fachin, S. (2006) "The Size and Power of Bootstrap and Bartlett-corrected Tests of Hypotheses on the Cointegrating Vectors" *Econometric Reviews*, 25, 41-60.

Obstfeld, M., Rogoff, K. (2000) "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" *NBER Working Paper Series* n. 7777.

Paparoditis, E., Politis, D.N. (2001) "The Continuous-Path Block Bootstrap" In *Asymptotics in Statistics and Probability. Papers in honor of George Roussas.* Madan Puri (ed.). VSP Publications: Zeist (NL).

Pedroni, P. (1999) "Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors", *Oxford Bulletin of Economics and Statistics*, 61, 653-670.

Pedroni, P. (2004) "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis" *Econometric Theory,* 20, 597-625.

Politis, D. N., White, H. (2003) "Automatic Block-Length Selection for the Dependent Bootstrap", *Econometric Reviews*, 23, 53-70.

Shiller, R.J., Perron, P. (1985) "Testing the Random Walk Hypothesis: Power versus Frequency of observations" *Economics Letters*, vol. 18, 381-386.

Westerlund, J. (2006a) "Testing for panel cointegration with a level break", *Economics Letters*, 91, 27-33.

Westerlund, J. (2006b) "Testing for Panel Cointegration with Multiple Structural Breaks", *Oxford Bulletin of Economics and Statistics*, 68, 101-132.

Westerlund, J., Edgerton, D. (2006a) "A Panel Bootstrap Cointegration Test" *Mimeo*, Lund University.

Westerlund, J., Edgerton, D. (2006b) "Simple Tests for Cointegration in Dependent Panels with Structural Breaks" *Mimeo*, Lund University.