Testing for breaks in cointegrated panels - with an application to the Feldstein-Horioka puzzle*

Francesca Di Iorio
University of Naples Federico II

Stefano Fachin
University of Rome "La Sapienza"

Abstract

Stability tests for cointegrating coefficients are known to have very low power with small to medium sample sizes. In this paper we propose to solve this problem by extending the tests to dependent cointegrated panels through the stationary bootstrap. Simulation evidence shows that the proposed panel tests improve considerably on asymptotic tests applied to individual series. As an empirical illustration we examined investment and saving for a panel of European countries over the 1960-2002 period. While the individual stability tests, contrary to expectations and graphical evidence, in almost all cases do not reject the null of stability, the bootstrap panel tests lead to the more plausible conclusion that the long-run relationship between these two variables is likely to have undergone a break.

Keywords: Panel cointegration, stationary bootstrap, parameter stability tests, FM-OLS.

JEL codes: C23, C15

* 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 and for helpful discussions, and to participants to the SIS Turin and Cambridge Panel Data 2006 conferences for suggestion and comments. Joakim Westerlund's comment and two anonymous reports greatly helped us in revising the paper. The usual disclaimers apply. Correspondence to: s.fachin@caspur.it, fdiiorio@unina.it.

Revised November 2007

1 Introduction

The analysis of cointegration in non-stationary panels has been recently rapidly expanding in two main directions. The first, urged by the nature of the data actually used in empirical applications, is the effort to generalise panel cointegration tests to the case of dependent units, either by modelling the dependence (*inter alia*, Gengenbach, Palm, Urbain, 2006) or reproducing it through the bootstrap (Fachin, 2007, Westerlund and Edgerton, 2006). The second direction follows steps already taken by the cointegration literature in the early '90's, tackling the issues of testing (*i*) cointegration allowing for breaks and (*ii*) the stability of a cointegrating relationship. In this stream of the literature, the first problem seems to have received more attention (*e.g.*, Banerjee and Carrion-i-Silvestre, 2004 and 2006, Gutierrez, 2005, Westerlund, 2006) than the second (to the best of our knowledge, only Emerson and Kao, 2001, 2005, for trend regressions, Kao and Chiang, 2000, for homogenous panel regressions). This is somehow surprising, as stability tests with unknown break points may have very low power with even medium sample sizes. For instance, the rejection rates under H_1 simulated by Gregory *et al.* (1996) for $T = 100$ in a Data Generating Process (DGP) with medium speed of adjustment to equilibrium are only marginally higher than Type I errors, and actually *lower* than the significance level: these tests are essentially useless from the empirical point of view. Exploring panel extensions, which may provide power gains, thus seems a promising area of research. A second surprising aspect of the current debate is that so far the developments in the treatment of dependence across units seems to have been largely ignored in the "panel with breaks" literature[1]. The tests proposed so far should thus be regarded essentially as a first step in the construction of empirically relevant procedures, very much like the first generation panel cointegration tests. On the contrary, in this paper we tackle the dependence issue from the outset, proposing a panel generalisation of Hansen (1992) stability tests based on the stationary bootstrap which is completely robust to cross-section dependence, and may thus be helpful for actual empirical work. A fitting empirical illustration is the so-called Feldstein-Horioka (1980) puzzle, *i.e.* the widespread evidence supporting the existence of a long-run link between the investment (I) and savings (S) to GDP (Y) ratios in advanced economies which should be characterised by

[1]Noticeable exceptions include the panel cointegration tests with breaks by Banerjee and Carrion-i-Silvestre (2004, 2006) and Westerlund (2006), which however leave many questions open. Westerlund applies simple resampling to data which, provided cointegration holds, are weakly dependent, while Banerjee and Carrion-i-Silvestre's (2004) procedure implies fitting an AR model to a MA process with a unit root under no cointegration (the same remark applies to Westerlund and Edgerton, 2007). Finally, Banerjee and Carrion-i-Silvestre (2006) test appears to have very good properties, but its small sample performance is not clear (the smallest sample sizes considered in Banerjee and Carrion-i-Silvestre's simulations with dependent units are T=50, N=40).

high international capital mobility. Since capital mobility is known to have increased in the late 1980's as a consequence of a worldwide shift towards financial liberalisation (see *e.g.,* Frankel, 1992) any investigation of the existence of this relationship should allow for breaks. This point has been taken into account by both Banerjee and Carrion-i-Silvestre (2004) and Di Iorio and Fachin (2007), who applied different panel cointegration tests allowing for breaks to a sample of fourteen European economies for the period 1960-2002. In both cases the results on the whole support the hypothesis of a possibly breaking relationship between the two variables. The next question to answer is if the relationship did actually undergo a break: to this end, in the empirical section we apply to the same dataset the procedure put forth in this paper.

We shall now (section 2) introduce the set-up and outline the testing procedure, then present the design and results of a Monte Carlo experiment (section 3) and the empirical illustration (section 4). Some conclusions and suggestions for future research are finally discussed (section 5).

2 Testing parameter stability in cointegrated panels

2.1 Set-up

Consider a $(k+1)-$dimensional $I(1)$ random variable Z observed over N units and T time periods (respectively indexed by i and t), naturally partitioned as $\mathbf{Z}_{it} = [Y_{it}, \mathbf{X}'_{it}]'$, where $\mathbf{X}'_{it} = [X_{1it} \dots X_{kit}]'$. Assume previous cointegration testing led to conclude that cointegration holds between Y_i and $\mathbf{X}'_i$. Then, as long as no long-run relationships among the $X's$ exist, we can estimate the N cointegrating vectors $\boldsymbol{\beta}_i, i = 1, \dots, N$, by applying some single-equation method. Although in the general case of non-diagonal disturbances some system approach such as FM-SOLS or FM-SUR (Moon, 1999) should be adopted, in this paper we shall limit the analysis to the simpler case of separate estimation of FM-OLS for each of the N time series.

Hansen (1992) proposed three tests for the hypothesis that the $\boldsymbol{\beta}$'s are stable over time when no a priori information on the location of the possible breaks t_i^b is available: (i) the maximum of the Chow tests computed at all possible break points ($SupF$); (ii) their mean ($MeanF$); (iii) a Lagrange-Multiplier test of the hypothesis that the coefficients follow a martingale process of zero variance (L_c). A panel extension along the lines of Pedroni's (1999) group mean test is readily obtained taking some summary statistic of the individual tests. The choice of the statistic is dictated by the view of the alternative hypothesis[2]: while the panel null hypothesis is $H_0^P =$

[2]The following discussion develops points raised by Joakim Westerlund in his comment uploaded on *Economics* webpage.

$\bigcap_i^N H_0^i$: "coefficient stability in all units", we can have different alternative hypothesis. A first possibility is H_1^P :"break in the coefficients in all units". Since the rejection region of the individual test statistics is the right tail of the distribution, in this case the suitable summary statistic is the minimum over units, as one unit not rejecting H_0 : "coefficient stability" is enough for not rejecting it in the panel as a whole. The opposite stance is that instability in one unity is enough to reject the null hypothesis in the entire panel; formally, H_1^P :"break in at least one unit". In this case the maximum should be used. However, the panel alternative hypothesis most likely to suit a typical empirical analysis is H_1^P : "break in a reasonable majority of the units", analogously to Pedroni's (2004) view of the alternative of panel cointegration tests as "cointegration in a reasonable majority of the units". In this case the outcome of the panel test depends on where the mass of the distribution of the break statistics lies. Hence, natural summary statistics are the mean (possibly in a robust, $\alpha-$trimmed version) or the median of the statistics computed for the individual units. Since our inference will be based on the bootstrap the traditional analytical reasons to favour the mean are obviously irrelevant, so that the best choice is likely to be the median.

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. To this end, we need to design a resampling scheme delivering pseudodata (*i*) obeying the null hypothesis of coefficient stability both when it holds and when it does not, and reproducing (*ii*) the autocorrelation and (*iii*) cross-correlation properties of the data. Let us discuss the three point in turn. The first requirement is easily satisfied by taking as systematic part of the bootstrap DGP $\widehat{\boldsymbol{\beta}}_i^0 \mathbf{X}_{it}$, with $\widehat{\boldsymbol{\beta}}_i^0$ the FM-OLS estimate of the cointegrating vector obtained under the hypothesis of no coefficient change. The second and third requirements (which hinge upon the algorithm to be applied to the cointegrating residuals in order to produce the bootstrap noises) are also easily satisfied, provided a key point is taken into account. Since the stochastic trend in the bootstrap DGP is given by $\widehat{\boldsymbol{\beta}}_i^0 \mathbf{X}_{it}$, the bootstrap noises to be added must be stationary either if H_0 holds or not: it is easily seen that only the residuals (say, $\widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1 \dots \widehat{\mathbf{u}}_N]$) computed using the model estimated under the most general hypothesis, *i.e.* H_1, satisfy this condition. To estimate this model we clearly need to choose a breakpoint; we will discuss this point in detail below. Since the residuals of FM-OLS are weakly dependent, their autocorrelation properties will be reproduced by any block boostrap scheme[3], such as *e.g.* the stationary bootstrap by Politis and Romano (1994). We thus only need to handle

[3]In block boostrap resampling schemes the pseudoseries are constructed stacking blocks of observations with random starting point and length proportional to the memory of the series, hence ensuring that the pseudodata reproduce the autocorrelation properties of the data.

cross-correlation. This is easily reproduced: resampling the entire $T \times N$ matrix of residuals $\widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1 \dots \widehat{\mathbf{u}}_N]$, rather than single series, will yield a matrix of pseudoresiduals $\mathbf{U}^* = [\mathbf{u}_1^* \dots \mathbf{u}_N^*]$ in which the rows, but not the columns, of $\mathbf{U}$ are reshuffled. The cross-correlation properties of the data are thus preserved.

Let us now go back to the issue of the specification of the break points for the models under H_1. An apparently appealing option is the period associated with $SupF$ in each unit, *i.e.* $\widehat{t}_i^b = \arg\max(\widehat{F}_{it})$, so that break location is allowed to vary across units. In fact, some exploratory simulations show that this turns out to be a good choice when there actually is a break in the data, but not when H_0 : "coefficient stability" holds. While existing breaks are accurately located (the median error is one period only; see histogram in Fig. 1), when no break takes place this criterion often leads to picking periods near the sample ends (see Fig. 2), causing overfitting and spuriously small estimated residuals. As a consequence of the latter, the bootstrap pseudodata tend to exhibit spuriously high signal/noise ratios, and the bootstrap stability tests to be severely oversized. It turns out that superior results are obtained when the restriction of a common break located at the median of the individual estimates of break periods is imposed: *i.e.*, $\widetilde{t}^b = median(\widehat{\mathbf{t}}^b)$, $\widehat{\mathbf{t}}^b = \left[\widehat{t}_1^b \widehat{t}_2^b \dots \widehat{t}_N^b\right]$, with $\widehat{t}_i^b = \arg\max(\widehat{F}_i)\ \forall i$. The histogram reported in Fig. 2 suggests that when the null hypothesis of no break holds this will be roughly equivalent to placing the break in the middle of the sample. Of course this will not be true when there actually is a break.

Summing up, denoting by S_i the stability statistic of interest for unit i, we propose to estimate the $p-$value of the group stability statistic S by the following algorithm:

1. Obtain estimates $\widehat{\boldsymbol{\beta}}_i^0$ of the cointegrating vectors under H_0 : "coefficient stability";

2. Compute the individual stability statistics $\widehat{S}_i$ and estimate break locations $\widehat{t}_i^b$;

3. Compute the group stability statistic $\widehat{S}$, *e.g.*, $\widehat{S}_m = \sum_{i=1}^N \widehat{S}_i/N$, or $\widehat{S}_{me}$ $= median(\widehat{\mathbf{S}})$, where $\widehat{\mathbf{S}} = \left[\widehat{S}_1, \dots, \widehat{S}_N\right]$;

4. Estimate models allowing for breaks at the period $\widetilde{t}^b = median(\widehat{\mathbf{t}}^b)$, $\widehat{\mathbf{t}}^b = \left[\widehat{t}_1^b \widehat{t}_2^b \dots \widehat{t}_N^b\right]$, and store the matrix of residuals $\widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1 \dots \widehat{\mathbf{u}}_N]$;

5. Resample the $T \times N$ matrix $\widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1 \dots \widehat{\mathbf{u}}_N]$ applying the stationary bootstrap (Politis and Romano, 1994) and obtain a matrix of pseudoresiduals $\mathbf{U}^* = [\mathbf{u}_1^* \dots \mathbf{u}_N^*]$;

6. Construct the pseudodata under H_0 : "coefficient stability" as $y_{it}^* = \widehat{\boldsymbol{\beta}}_i^0 \mathbf{X}_{it}' + u_{it}^*$;

7. Compute the group stability statistic S^* for the pseudo-data set $[Y_{it}^*, \mathbf{X}_{it}']$, $i = 1, \ldots, N, t = 1, \ldots, T$;

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

9. Compute the boostrap estimate of the $p-$value as $p^* = prop(S^* > \widehat{S})$.

Two remarks are in order:

(i) The hypothesis of partial (involving only some of the coefficients) stability is easily handled by modifying accordingly the equations estimated in step (4) and the stability statistics adopted;

(ii) 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.

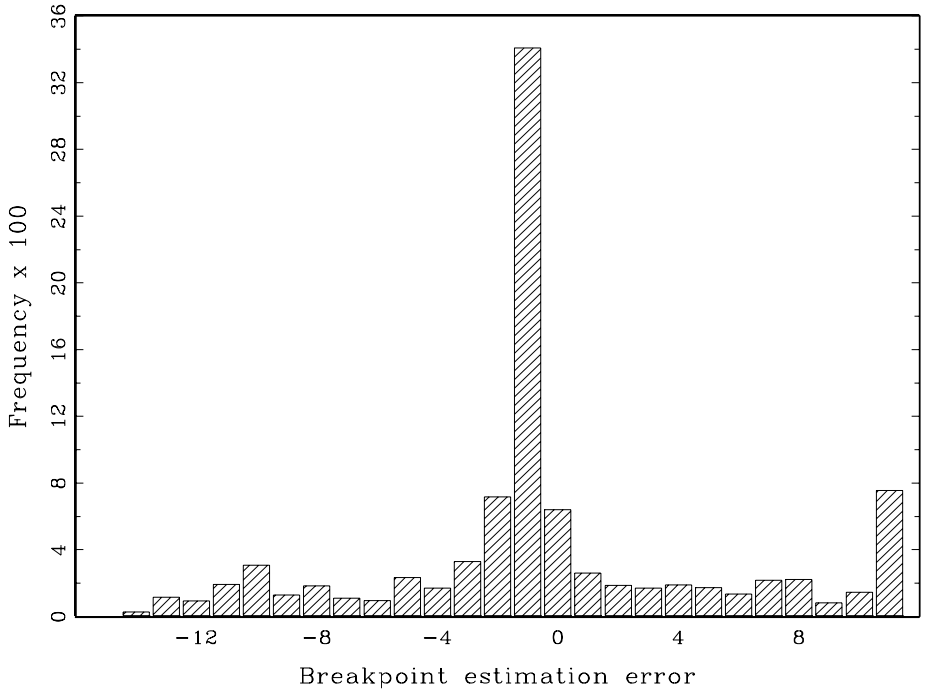

Fig. 1. Distribution of the error in the estimation of the breakpoint $\widehat{t}_i^b - t_i^b$,where $\widehat{t}_i^b = \arg\max(\widehat{F}_{it})$ and $t_i^b \sim Uniform[0.5T - 3, 0.5T + 3]$ with $T = 50$. 25% trimming at each sample end, pooled results from 500 Montecarlo replications for 40 units.

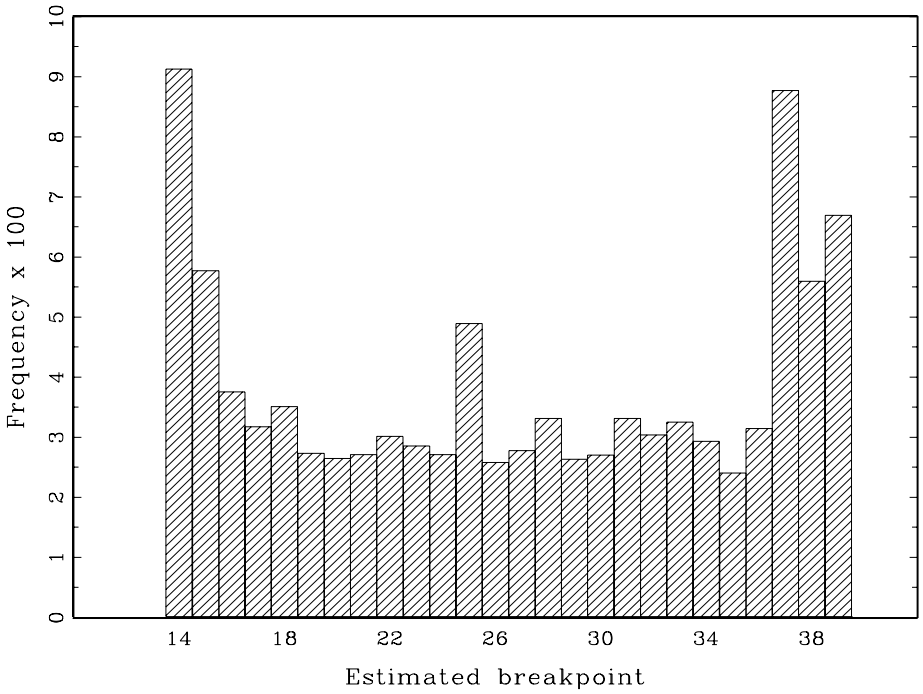

Fig. 2. Distribution of $\widehat{t}_i^b = \arg\max(\widehat{F}_{it}), T = 50$, 25% trimming at each sample end, when there is no break in the cointegrating coefficients. Pooled results from 500 Montecarlo replications for 40 units.

3 Monte Carlo Experiment

3.1 Design

The simulation experiment is based on the design adopted by Fachin (2007), essentially a generalisation of the Engle and Granger (1987) and Gonzalo (1994) classical DGP to the case of dependent panels (a similar design is also employed by *e.g.* Kao, 1999).

Considering for the sake of simplicity the bivariate case $\mathbf{Z}_{it} = [Y_{it} X_{it}]'$ the DGP can be summarised as follows. Following Pesaran (2007), short-run dependence is induced by defining the shocks driving Y and X ($u^j, j = x, y$) as the sum of a idiosyncratic component ($\epsilon^j, j = x, y$) and a single stationary common factor ($f_t^j, j = x, y$); long-run dependence is caused by an explanatory variable common across units. Since single-equation methods assume weak exogeneity of the left-hand side variable for the cointegrating coefficient we exclude any long-run feedback from Y to X. Letting t_i^b be the period in which the break takes place in unit i, we then have:

$$x_{it} = u_{it}^x \tag{1}$$

$$y_{it} = \begin{cases} \mu_{0i} + \beta_0 x_{it} + u_{it}^y, & t \leq t_i^b \\ \mu_{1i} + \beta_1 x_{it} + u_{it}^y, & t > t_i^b \end{cases} \tag{2}$$

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, $f^j \sim N(0,1), j = x, y$, and an idiosyncratic one, ϵ^j, $j = x, y$:

$$\begin{cases} u_{it}^x = \gamma_i^x f_t^x + \epsilon_{it}^x \\ u_{it}^y = \gamma_i^y f_t^y + \epsilon_{it}^y \end{cases} \tag{3}$$

The factor loadings $\gamma_i^j, j = x, y$, determine the strength of the short-run cross-correlation across units. Considering that in the dataset on a panel of European economies used in our empirical application the maximum cross-correlation between the log differences of savings (the right-hand side variable) and that in the residuals of the Feldstein-Horioka equations are respectively about 0.62 and 0.65 we set $\gamma_i^j \sim Uniform(-1, 6)\ \forall i, j$, which yields a cross-correlation close to 0.65.

The structure of the idiosyncratic component is:

$$\begin{cases} \epsilon_{it}^x = \sum_{j=1}^t (e_{it-j}^x + \theta) \\ \epsilon_{it}^y = \phi_i \epsilon_{it-1}^y + e_{it}^y \end{cases} \tag{4}$$

where ϕ_i are the parameters governing the speed of adjustment to the long-run equilibrium ($\phi_i = 0$ would imply instantaneous adjustment). This parameter affects in the same way the asymptotic and the panel tests: both deliver better results when the long-run relationship stands out more clearly from the data, *i.e.* when the speed of adjustment is higher. Since the primary aim of our experiment is assessing the scope for improving over pure time series testing using a panel approach fixing the $\phi's$ does then not cause any no loss of generality. A natural choice are the values adopted by Pesaran (2007) in a recent work on the closely related issue of panel unit root tests, namely $\phi_i \sim Uniform(0.2, 0.4)$.Finally,

$$\begin{cases} e_{it}^x \sim N(0, \sigma_{ix}^2) \\ e_{it}^y \sim N(0, \sigma_{iy}^2) \end{cases} \tag{5}$$

with $\sigma_{ij}^2 \sim Uniform(0.5, 1.5)$, $j = x, y$, so to allow for some heterogeneity across units. This range is likely to be reasonably representative of empirical datasets: after normalising on the cross-section average, in ten out of the 12 economies included in our panel the variances of the log differences of savings fall in the interval $[0.40, 1.72]$, while nine of those of the equation residuals are included in the interval $[0.38, 1.49]$.

This DGP is obviously quite complex. Rather than aiming at the unfeasible task of a complete design[4] we will define as a base case an empirically

[4]The number of loops to be executed in each experiments grows geometrically with the number of units, breakpoints, Monte Carlo replications and bootstrap redrawings. For instance, with $N = 40$, 20 possible breakpoints, 500 Monte Carlo replications and 1000 bootstrap redrawings we have a total of 400 million loops.

relevant set-up and then explore a few interesting variations. Considering that the simple bivariate DGP often used in simulation experiments is clearly unrealistic, but in single-equation cointegration modelling the number of explanatory variables is usually limited, we generally set $k = 2$ in both the DGP and estimated model. With no loss of generality we set both constant and slopes to 3 before the break (the same value chosen by Banerjee and Carrion- i-Silvestre, 2004, for the slope); after the break all coefficients are halved.

Since Gregory *et al.* (1996) report a tendency to overrejection of the asymptotic test in models with 3 or 4 explanatory variables we also run a separate experiment with $k = 4$. Finally, a key point is that given the rather short time series analysed in most experiments, in order to ensure computational stability we fixed the trimming coefficient at 25%. The cases considered are six altogether.

1. *Base case*: $T = 50$, N from 5 to 40; in the power simulations break date Uniform over units in $[0.5T \pm 3] = [22, 28]$. Since recursive stability tests assume rather large sample sizes we chose to fix the time sample in all experiments except the following one to 50. This is admittedly a rather large sample in terms of annual data, but pretty small if a quarterly frequency is assumed. It may thus be considered relevant for actual empirical applications (note that it is much smaller than those typically considered in simulation studies on stability tests, where generally $T \geq 100$).

2. *Large T*: $T = 100$, $N = 3, 5$; in the power simulations break date Uniform over units in $[0.5T \pm 3]$. The aim of this experiment is mainly checking the time-asymptotic behaviour of the tests. For computational convenience only two very small cross-section sample sizes are examined, with one ($N = 5$) present in the Base Case also allows comparisons for fixed N and different T, and the other ($N = 3$) allows evaluating behaviour for different N and given T.

3. *Larger model*: $T = 50$, N from 5 to 40, $k = 4$; break date Uniform over units in $[0.5T \pm 3]$. This case is designed exactly like the Base case, except the number of explanatory variables in both the DGP and estimated model.

4. *Late break*: $T = 50$, N from 5 to 40; break date Uniform over units in $[0.75T \pm 3]$, that is $[35, 41]$. Since 25% of the sample is trimmed at each end, the estimation sample is $[13, 38]$: the break can thus fall very close or even after the end of the estimation sample, a very demanding set-up.

The bootstrap algorithm described above is based on residuals of coin-

tegrating regressions estimated for all units with a break at the median of the individual estimated break points, which is intuitively acceptable if we assume all units to be 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, two rather different set-ups may arise: (*i*) the break periods may be widely disperse over units, for instance because they stem from different causes, each one relevant to only some units; (*ii*) some of the units may be not affected by a break at all. The last two are designed to investigate these two scenarios in turn:

5. *Twin breaks:* as Base case, but in half of the units the break date is Uniform in $[0.3T \pm 3]$, and in the other half in $[0.6T \pm 3]$.

6. *Partial break*: $T = 50$, N from 10 to 40, break date Uniform in $[0.5T \pm 3]$ over $0.7N$ units (the first seven in each block of ten), no break in the remaining units. This case is designed to examine tests performances when the cointegrating relationship is mostly (here 70% of the cases), but not always, unstable. Note that since this view of the test makes sense on with fairly large cross-section sample sizes we set $N \geq 10$.

After some experimentation with different options we decided to fix the number of Monte Carlo replications at 500 and that of bootstrap redrawings at 1000. Higher numbers of either would have delivered a small increase in the precision of the results not worth the large increase of the cost and time scale of the experiment (which, because of the recursive nature of the statistics evaluated, is computationally very demanding).

To evaluate the improvements (in terms of both power gains and reduction in size bias) which could be expected by moving from a standard time series to a panel set-up we computed the average rejection rates of the asymptotic tests based on Hansen (1992) asymptotic critical values computed for all individual units involved in each experiment[5], namely 40 in all experiments except "Large T" where it is limited to 5. These averages must be compared only with the rejection rates for the panel tests with all units (either 40 or 5).

In all power experiments no size-adjustment is applied to the rejection rates[6]. Although size-adjustments make power comparisons easier we believe

[5]Except the "Partial break" case, where they will simply be a weighted average of the size and power of the test, with weights respectively given by the fractions of non-breaking and breaking units.

[6]Since we use the bootstrap distribution directly to compute the p-values of the empirical statistic under the null hypothesis, rejecting in the power simulations if $p^* > \widehat{\alpha}$, where p^* is the bootstrap p-value and $\widehat{\alpha}$ the empirical size of the test (proportion of rejections of the test under the null hypothesis, assumed to have been previously simulated) would deliver size-adjusted power.

that in this case this benefit is not worth the loss of relevance for empirical work which is caused by the adjustment (Horowitz and Savin, 2000, describe size-adjusted power as "irrelevant for empirical research"). The comparisons made in our Monte Carlo study involve three tests (L_c, $MeanF$ and $SupF$) in the panel and time series versions, so that we can compare either(i) different tests in the same version (*e.g.*, L_c vs. $MeanF$ in the panel set-up);(ii) different versions of the same test (*e.g.*, time series L_c vs. panel L_c). As we will see below, the results of latter comparison are obvious even without size-adjustment: for instance, for a 5% $SupF$ test in the Base case with $N = 40$ we have: time series, $\widehat{\alpha} = 0.5\%$, power = 1.8% ; panel test, $\widehat{\alpha} = 6.9\%$, power = 92.8%. The power differential is so large that the size-adjustment (which will increase the power of the time series test and lower that of the panel test) will not change in any significant way our appreciation of the comparative properties of the two tests. The advantages of the size-adjustment would then in practice be confined to the first type of comparisons, between the three tests in their panel version. It turns out that often (though by no means always) the Type I errors of the panel tests are very similar (for instance, for a 5% test in the Base case with $N = 40$, we have $2.4\%, 3.2\%$ and 2.4%, respectively for the L_c, $MeanF$ and $SupF$ tests), so that the size-adjustment will often have only a minor impact on the relative power of the three panel tests.

Summing up, size-adjustment will help only in a minority of the comparisons to be carried out between the panel versions of the three tests. Given this somehow limited benefits, we preferred to plan our experiment so to shed some light on the properties of the tests exactly as they might be used in an empirical application. In applied work it is important to know that a certain test tends to reject very often because oversized, or viceversa, as this knowledge may guide its empirical use.

3.2 Results

The results are reported in Tables 1A-6B below. In the Base case ($T = 50$, N from 5 to 40) the Type I errors (Table 1A) of the bootstrap panel tests have some positive size bias for $N = 5$ but converge fairly closely to nominal significance levels as N increases. The asymptotic tests on individual series deliver variable performances: the L_c test is slightly oversized, while both the $MeanF$ and the $SupF$ appear to be conservative (more the latter than the former). The power gains offered by the panel tests are remarkable. Consistently with *a priori* expectations, the asymptotic tests have negligible power, while that of the panel tests is generally acceptable and definitely good for $\alpha = 10\%$ and $N \geq 10$ (e.g., 92% for $N = 40$, with Type I error 11%; Table 1B). Hence, using the panel tests grants considerable improvements with respect to aggregate tests in terms of both reduction of size bias and increase in power. In fact, with this time sample a panel approach seems

to be the only viable option. In comparative terms, we find the Type I errors to be very similar for all the three tests, while the $SupF$ test appears to be somehow marginally less powerful than the L_c and $MeanF$. The results of the mean and median panel tests also appear very similar. Since these findings hold approximately in all the cases examined the following comments are mostly expressed in general terms, with no reference to the specific tests.

Allowing for the different speed of adjustment of the DGP's employed (ϕ in our design), the "Large T" results (Tables 2A-2B) for the asymptotic tests are fully consistent with Gregory *et al.* (1996): as we can see, the size bias is still noticeable, and power very poor. On the other hand, the Type I errors of the bootstrap panel tests essentially converge to nominal significance levels, and their power approaches 100% even with extremely small N. Hence, even when a rather large time sample is available, considerably superior results may be obtained following a panel approach, even with N as small as 5.

When $T = 50$ and breaks around 3/4 of the time sample (Table 3) power falls dramatically, rarely reaching 50% for the mean test; the performance of the median test, although not brilliant, appear somehow more robust. Since the upper extreme of the break interval ($t = 41$) falls after the end of the actual estimation sample ($t = 38$) these findings are not surprising, and make clear the great care necessary in using recursive stability tests.

The two experiments designed to check the robustness of the bootstrap procedure with respect to the nature of the breaks deliver comforting results (Table 4). Here the issue is that the cointegrating equation used to estimate the residuals to be bootstrapped is severely misspecified. The equation assumes one common breakpoint (located at the median of the individual estimated breaks), while the breakpoints are heterogenous and grouped in two different clusters. It is interesting to see the test turns out to be quite robust, as there is only a minimal power loss.

It is also interesting to see that when 70% of the units are affected by the break (Table 5) the rejection rates seem to fall approximately in the same proportion (e.g., for $N = 40$ and $\alpha = 10\%$ from 92.2% to 66.8%). Hence, the panel test is likely to reject the stability hypothesis when it does not hold in the majority of the units.

In a larger model with four explanatory variables (Tables 6A-B) we notice that the performance of the asymptotic tests is even worst than in the Base case. The Type I errors of the panel tests appear similar to the base case with only two variables, but unfortunately their power somehow smaller, possibly because of the larger number of coefficient estimated.

The overall conclusions to be drawn are now rather clear: consistently with Gregory *et al.* (1996) our experiments suggest that with a small or moderately large sample size ($T \leq 100$) Hansen (1992) asymptotic test has power ranging from very low to close to zero. A fairly general solution to this serious empirical shortcoming seems to be provided by a panel approach

based on the bootstrap: in out experiments the Type I errors turned out to be generally close to nominal sizes and converging rather rapidly over both over T and N to nominal levels, and power from acceptable to good with $\alpha =$ 10% when the break is located around the middle of the sample. Although tests power does not appear to be much affected by a wide dispersion of the breaks across units and to be (correctly) roughly proportional to the fraction of breaking units, it is important to keep in mind that it can be disappointing if the breaks fall towards the end of the sample (which is not surprising, since with a small time sample the marginal information becomes very small).

Table 1A: Base Case: $T = 50$, N from 5 to 40 − Size
(*Rejection Rates*×100)

		N							
		5	10	20	40	5	10	20	40
α	Asy^a	*Boot-Mean*^{b}				*Boot-Median*^{c}			
					A. L_c				
1.0	3.9	1.6	0.0	0.0	0.0	0.8	0.0	0.0	0.0
5.0	12.1	10.4	0.8	0.8	2.4	8.8	1.6	3.2	4.0
10.0	19.3	20.8	4.0	4.0	11.2	23.2	6.4	9.6	11.2
					B. $MeanF$				
1.0	0.5	0.8	0.0	0.0	0.0	2.4	0.0	0.0	0.0
5.0	3.1	10.4	1.6	0.8	3.2	13.6	2.4	0.8	1.6
10.0	6.2	16.0	4.8	6.4	9.6	24.8	7.2	8.0	15.2
					C. $SupF$				
1.0	0.0	2.4	0.0	0.8	0.0	2.4	0.0	0.0	0.0
5.0	0.2	11.2	1.6	2.4	2.4	13.6	1.6	0.0	1.6
10.0	0.5	20.8	7.2	5.6	8.8	24.0	5.6	6.4	12.8

DGP: No Break;
H_0*:* No break*;*
(a) *Asy:* average rejection rates of invidual tests over all 40 units, Hansen (1992) asymptotic critical values;
(b,c) *Boot-mean/median*: bootstrap test on the mean/median across units of the stability statistics;
Bootstrap: 1000 redrawings, block size $T/10$;
Montecarlo: 500 replications.

Table 1B: Base Case: $T = 50$, N from 5 to 40 − Power
(*Rejection Rates*×100)

		N							
		5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
					A. L_c				
1.0	3.5	6.6	5.6	6.4	5.2	7.4	10.6	11.6	10.0
5.0	11.5	36.4	39.8	55.0	59.4	41.0	44.4	55.0	58.6
10.0	19.3	57.0	73.2	87.6	92.2	62.0	70.2	77.6	84.8
					B. $MeanF$				
1.0	0.8	6.8	7.0	9.6	6.8	5.4	9.8	14.8	11.4
5.0	3.6	35.8	48.0	61.8	62.8	37.6	51.6	61.2	62.8
10.0	6.9	61.4	80.2	87.4	92.8	61.2	78.2	86.6	90.0
					C. $SupF$				
1.0	0.1	2.4	2.0	4.4	1.8	2.0	3.0	7.6	3.8
5.0	0.7	21.8	28.2	35.6	29.2	24.4	29.8	37.2	39.6
10.0	1.8	48.6	63.0	62.0	67.2	42.4	59.8	66.2	70.2

DGP: Break Uniform in $[0.5T \pm 3] = [22, 28]$;
H_0*:* No break*;*
All abbreviations and definitions: see table 1A.

Table 2A: Large T: $T = 100$, $N = 3$, 5− Size
(*Rejection Rates*×100)

		N			
		3	5	3	5
α	Asy^a	*Boot-Mean*		*Boot-Median*	
		A. L_c			
1.0	2.6	1.0	2.0	2.2	2.6
5.0	18.2	4.8	5.4	5.0	6.2
10.0	33.2	9.0	8.8	9.8	11.0
		B. $MeanF$			
1.0	3.3	1.4	1.6	1.4	2.4
5.0	13.9	4.6	4.4	5.4	5.6
10.0	22.8	8.0	8.6	8.8	10.4
		C. $SupF$			
1.0	0.6	1.4	1.2	1.4	1.6
5.0	6.4	5.0	4.8	6.0	5.2
10.0	12.0	10.8	10.4	10.2	10.8

DGP: No break;
H_0*:* No break.
(a) *Asy:* average rejection rates of invidual tests over all 5 units, Hansen (1992) asymptotic critical values;
All other abbreviations and definitions: see table 1A.

Table 2B: Large T: $T = 100$, $N = 3, 5-$ Power
(*Rejection Rates*$\times 100$)

	N				
		3	5	3	5
α	Asy^a	*Boot-Mean*		*Boot-Median*	
		A. L_c			
1.0	13.5	88.6	95.4	71.8	77.0
5.0	33.0	99.0	99.8	88.8	91.0
10.0	44.0	99.8	100.0	94.2	96.0
		B. $MeanF$			
1.0	7.9	96.2	99.6	86.8	90.6
5.0	24.2	99.8	100.0	96.6	98.8
10.0	33.9	100.0	100.0	98.8	99.8
		C. $SupF$			
1.0	3.1	95.2	98.8	90.2	93.4
5.0	10.6	99.6	100.0	98.6	99.6
10.0	17.9	99.8	100.0	99.2	100.0

DGP: Break Uniform in $[0.5T \pm 3]$;
H_0*:* No break.
(a) *Asy:* average rejection rates of invidual tests over all 5 units, Hansen (1992) asymptotic critical values;
All other abbreviations and definitions: see table 1A.

Table 3A: Larger model: $T = 50$, N from 5 to 40− Size
(*Rejection Rates*×100)

		N							
		5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
		A. L_c							
1.0	1.3	1.0	0.2	0.2	0.4	0.6	0.2	0.0	0.0
5.0	8.9	5.0	1.4	1.8	2.0	6.2	0.8	3.2	1.8
10.0	17.2	11.2	4.6	7.0	5.6	11.8	4.4	8.0	5.2
		B. $MeanF$							
1.0	0.1	0.6	0.0	0.0	0.2	0.6	0.0	0.2	0.0
5.0	1.2	5.4	1.4	2.4	2.2	5.6	0.8	2.0	0.8
10.0	3.6	10.0	4.6	6.4	6.4	10.4	4.0	7.8	6.2
		C. $SupF$							
1.0	0.0	0.8	0.2	0.2	0.0	0.6	0.2	0.0	0.0
5.0	0.0	4.4	1.6	2.4	1.8	4.0	1.0	2.0	1.6
10.0	0.1	10.4	5.0	5.8	6.4	10.8	3.6	7.2	5.4

DGP: No break, four explanatory variables;
H_0*:* No break*;*
All abbreviations and definitions: see table 1A.

Table 3B: Larger model: $T = 50$, N from 5 to 40− Power
(*Rejection Rates*×100)

		N							
		5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
		A. L_c							
1.0	1.2	2.0	2.4	2.4	5.8	2.0	2.2	5.6	5.2
5.0	6.2	11.6	23.8	37.0	57.8	8.8	27.4	42.2	36.0
10.0	12.2	25.0	53.0	71.8	87.2	17.2	59.0	64.0	60.4
		B. $MeanF$							
1.0	0.1	4.0	2.8	6.2	9.0	4.0	2.6	7.4	7.2
5.0	1.4	22.6	31.6	48.0	66.8	15.6	28.6	40.0	40.6
10.0	3.2	32.6	62.4	80.4	93.6	27.4	57.8	65.8	64.8
		C. $SupF$							
1.0	0.0	1.0	0.4	0.4	0.4	0.8	0.8	1.2	1.8
5.0	0.1	7.2	6.8	8.6	18.8	8.0	11.2	13.8	14.2
10.0	0.4	15.4	19.6	29.6	49.2	15.8	26.4	29.6	30.8

DGP: Break Uniform in $[0.5T \pm 3]$, $k = 4$;
H_0*:* No break*;*
All abbreviations and definitions: see table 1A.

Table 4: Late break: $T = 50$, N from 5 to 40
(*Rejection Rates*×100)

		N							
		5	10	20	40	5	10	20	40
α	*Asy*	*Boot-Mean*				*Boot-Median*			
		A. L_c							
1.0	3.5	2.4	0.8	0.6	0.2	6.4	1.6	4.2	0.8
5.0	11.4	25.0	16.0	24.2	15.0	31.8	19.6	37.0	33.2
10.0	20.8	47.6	38.6	61.6	50.2	49.2	42.4	66.4	63.0
		B. $MeanF$							
1.0	0.8	2.4	0.4	0.6	0.2	6.4	1.0	3.4	1.8
5.0	3.6	23.8	13.2	20.8	18.6	31.6	17.4	37.8	40.4
10.0	6.9	45.8	38.4	56.4	58.2	49.6	46.4	67.8	75.6
		C. $SupF$							
1.0	0.1	2.2	0.8	1.0	0.6	3.0	0.6	2.0	1.0
5.0	0.7	18.6	12.2	19.0	20.0	20.8	15.4	27.8	30.2
10.0	1.8	37.4	35.4	45.8	55.0	39.4	42.2	54.4	62.2

DGP: Break Uniform in $[0.75T \pm 3] = [35, 41]$;
H_0*:* No break*;*
All abbreviations and definitions: see table 1A.

Table 5: Twin breaks: $T = 50$, N from 5 to 40
Rejection Rates×100

	N							
	5	10	20	40	5	10	20	40
α	*Boot-Mean*				*Boot-Median*			
	A. L_c							
1.0	6.4	10.0	7.4	11.6	7.2	15.2	23.6	41.8
5.0	20.2	38.0	36.8	56.2	30.0	45.4	58.2	77.0
10.0	40.0	59.2	65.6	82.8	45.0	61.8	76.8	85.4
	B. $MeanF$							
1.0	5.0	10.4	6.0	7.6	7.8	17.4	19.6	35.2
5.0	21.2	37.4	32.2	46.2	30.4	43.8	56.4	73.8
10.0	40.0	55.4	63.0	77.0	44.2	58.4	73.2	86.2
	C. $SupF$							
1.0	5.6	10.4	8.6	7.0	5.2	13.6	14.4	23.0
5.0	22.8	34.2	31.4	34.8	23.8	41.0	44.4	60.6
10.0	38.6	53.6	54.2	65.0	39.4	55.0	63.2	77.6

DGP: Units $1, 3, \ldots, N-1$ break Uniform in $[0.3T \pm 3]$,
Units $2, 4, \ldots, N$ break Uniform in $[0.6T \pm 3]$;
H_0*:* No break*;*
All abbreviations and definitions: see table 1A.

Table 6: Partial break: $T = 50$, N from 10 to 40
(*Rejection Rates*×100)

	N					
	10	20	40	10	20	40
α	*Boot-Mean*			*Boot-Median*		
	A. L_c					
1.0	0.8	2.4	2.2	4.4	6.0	2.0
5.0	20.2	28.6	28.4	30.4	35.0	26.4
10.0	45.8	59.8	66.8	50.0	55.0	55.8
	B. $MeanF$					
1.0	2.4	3.4	2.6	3.2	4.8	2.8
5.0	22.8	35.4	33.4	29.2	36.6	32.6
10.0	50.8	67.8	74.6	54.8	64.0	62.4
	C. $SupF$					
1.0	0.8	1.4	0.6	0.8	3.0	1.4
5.0	16.4	21.2	18.6	18.6	25.4	23.0
10.0	39.0	48.6	48.8	41.2	49.8	47.2

DGP: Break Uniform in $[0.5T \pm 3]$ in
$0.7N$ units (the first seven in each block of ten);
H_0*:* No break*;*
All other abbreviations and definitions: see table 1A.

4 Empirical illustration: the Feldstein-Horioka Puzzle

As discussed in the Introduction, the apparent existence of a long-run link between the investment and savings in advanced economies, where high capital mobility may allow the current account to be unbalanced for long periods, is one the major empirical puzzles of contemporary macroeconomics (six altogether according to Obstfeld and Rogoff, 2000). Since changes in financial regulations may clearly have considerable effects on the relationship between savings and investment in an individual economy this issue is clearly best examined using a cointegration test allowing for breaks. Fitting starting points for our analysis are thus Banerjee and Carrion-i-Silvestre (2004) and Di Iorio and Fachin (2007), who studied 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. The conclusions of both studies are globally in favour of the existence of a relationship; hence, in the panel as a whole investment and saving do appear to cointegrate if breaks are allowed. Taking

these conclusions for granted we shall now proceed to next question: did a break actually took place, or, else, are the cointegrating coefficients stable? The tests developed in the previous section will help us in trying to reach an answer.

Plots of both variables reported in Fig. 3A-B; note that since in our previous work (Di Iorio and Fachin, 2007), we found the Savings/GDP ratio to be stationary in Finland and Portugal these two countries are not included in the panel. Consistently with a priori expectations (Frankel, 1992), in several cases (*e.g.*, Austria, Belgium, Netherlands, Sweden) the graphical analysis indeed suggests a weaker association between the two variables towards the end of the period, so that coefficients shifts appears plausible enough to be investigated.

Recalling that the choice of the trimming coefficient may affect considerably the results we computed all tests with both 25% and 12.5% trimming, obtaining always very similar results. Examining the individual statistics (Table 7; to save space we report only the results for 12.5% trimming) we find extremely strong evidence of instability in Belgium, while most of the remaining statistics are not significant. The failure of the asymptotic tests to reject the hypothesis of stability for the individual countries is puzzling in view of the graphical evidence, and the natural suspicion is that it may be merely due to the extremely low power to be expected from the tests with such a small sample size. In fact, moving to the panel tests we can see (Table 8) that the means of all statistics suggest strong rejection of the null hypothesis of stability, with p-values smaller than 5% (actually zero for the $MeanF$ and $SupF$ statistics). Since this outcome may be due to the strong evidence for instability in Belgium it is important to look also at the median of the individual statistics, which will not be influenced by a single case. Here the evidence for rejection is weaker, with p-values between 10% and 15% for the L_c and $MeanF$. However, recalling (cf. Table 1B) that with a panel of 12 units power must be expected to be rather low, such p-values should nevertheless be regarded as small enough to grant rejection. We can thus appreciate how applying the panel procedure does grant a power gain with respect to the individual tests, allowing to reach the more plausible conclusion that in this group of countries investment and savings do seem to be linked by a long-run relationship, but this is likely to have changed over time at least once.

The next natural step is to estimate for each economy an investment-savings relationship allowing for coefficient breaks at the estimated breakpoints, $\widehat{t}_i^b = \arg\max(\widehat{F}_{it})$:

$$\ln(I/Y)_{it} = (\theta_{0i} + \theta_{1i} D_{it}) + (\beta_{0i} + \beta_{1i} D_{it}) ln(S/Y)_{it} + \epsilon_{it} \qquad (6)$$

where $D_{it} = 1$ if $t > \widehat{t}_i^b, 0$ else. Although the small time sample suggests great care, especially when the break falls near the extremes of the sample, the results (reported in table 9) are of some interest.

In seven countries (Austria, Belgium, Germany, France, Ireland and Sweden, thus including two of the largest continental European economies), the retention ratio falls significantly after the break, consistently with the expectations of a progressive weakening of the long-run link between investments and savings in the advanced economies. In the case of the United Kingdom the results are peculiar, as the retention ratio is negative before 1977 and turns positive afterwards. However, neither estimates are significant, suggesting that in this case there may not be an actual causal link of any relevance running from domestic savings to investment. This hypothesis is consistent with Kejriwal (2007), who using quarterly data over the period 1957:1-2006:1 found no evidence for cointegration for this country.

Finally, in the four remaining cases (Italy, Spain, Greece, Denmark), contrary to expectations, the retention ratio seems to increase. However, two remarks are in order: first, the change in the retention rate, measured by the coefficient β_1, is never significant (nor the individual stability statistics, with the exception of Greece); second, in two cases (Italy and Spain) the estimated break points falls at the extremes of the interval in which they are constrained to lie (respectively, 1970 and 1991). From Fig. 2 we know that this is typical of cases when no break actually took place. Unfortunately, with the available sample size no reliable conclusions for individual cases can be reached, so it is impossible to shed more light on the issue. Clearly, the great care invoked above is fully necessary.

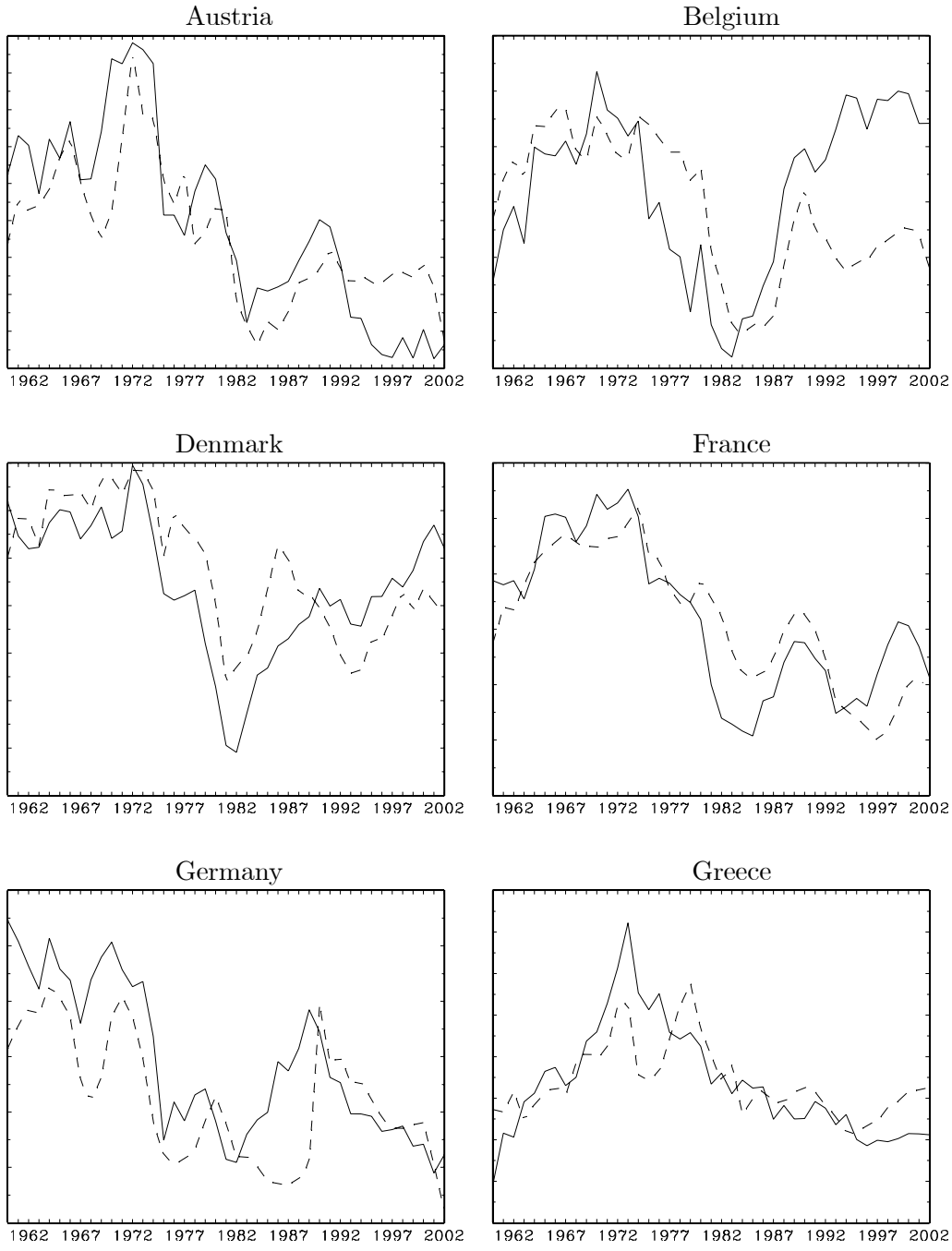

Fig. 3A. Savings/GDP (solid line) and Investments/GDP (dotted line), 1960-2002.

Fig. 3B. Savings/GDP (solid line) and Investments/GDP (dotted line), 1960-2002.

Table 7

Individual stability tests of the investment-savings long-run relationship, 1960-2002

	Austria	Belgium	Denmark	France	Germany	Greece
L_c	0.27	1.28***	0.12	0.08	0.26	0.35
$MeanF$	2.19	45.15***	0.75	0.53	2.48	5.12**
$SupF$	4.07	163.34***	1.65	1.18	10.94	27.52***
	Ireland	Italy	Netherlands	Spain	Sweden	UK
L_c	0.25	0.19	0.22	0.17	0.17	0.05
$MeanF$	3.17	1.23	1.86	1.51	4.90	0.75
$SupF$	14.57**	6.80	3.19	5.63	12.50	12.36

trimming: 12.5%;

*: significant at 10%; **: 5%;***: 1%.

Table 8

Panel tests of stability of the investment-savings long-run relationship, 1960-2002

p-values $\times 100$

	mean			median		
Trimming	L_c	$MeanF$	$SupF$	L_c	$MeanF$	$SupF$
25%	3.1	0.0	0.0	14.4	12.1	44.7
12.5%	3.4	0.0	0.0	16.7	14.9	0.2

panel: Austria, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Netherlands, Spain, Sweden, UK;

bootstrap: 1000 redrawings.

mean/median: of the test statistics across units.

Table 9
The investment-savings long-run relationship, 1960-2002
FM-OLS estimates

	β_0	β_1	θ_0	θ_1	*breakpoint*
Austria	0.93 [0.10]	−1.07 [0.49]	0.45 [0.55]	3.18 [1.58]	1991
Belgium	0.71 [0.15]	−0.75 [1.15]	0.94 [0.47]	2.21 [3.70]	1989
Denmark	0.67 [0.05]	0.19 [0.20]	1.09 [0.14]	−0.76 [0.60]	1974
France	0.59 [0.05]	−0.23 [0.18]	1.32 [0.15]	0.58 [0.55]	1975
Germany	0.92 [0.36]	−0.72 [0.39]	0.17 [1.19]	2.26 [1.28]	1972
Greece	0.72 [0.15]	0.11 [0.20]	0.79 [0.48]	−0.21 [0.64]	1989
Ireland	1.03 [1.51]	−0.83 [1.54]	0.04 [4.33]	2.44 [4.42]	1970
Italy	0.80 [0.52]	0.45 [0.56]	0.74 [1.63]	−1.53 [1.75]	1970
Netherlands	0.89 [0.18]	−1.07 [0.49]	0.45 [0.55]	3.18 [1.58]	1985
Spain	0.67 [0.24]	0.27 [0.31]	1.08 [0.74]	−0.84 [0.96]	1991
Sweden	0.75 [1.85]	−1.66 [1.89]	0.83 [5.84]	4.82 [5.97]	1974
UK	−0.25 [0.48]	−0.47 [0.57]	3.72 [1.44]	−1.49 [1.68]	1977

model: $\ln(I/Y)_{it} = (\theta_{0i} + \theta_{1i}D_{it}) + (\beta_{0i} + \beta_{1i}D_{it})ln(S/Y)_{it} + \epsilon_{it}$,
$D_{it} = 1$ if $t > \widehat{t}_i^b, 0$ else;
breakpoints estimated using 25% trimming at both ends (hence, constrained in the interval 1970-1991);
standard errors in brackets.

5 Conclusions

Our overall conclusion is that the proposed panel stability tests may grant considerable advantages. With time sample sizes rather common in macroeconomic datasets (*e.g.*, 50 observations) the asymptotic tests appear to be essentially of no use, while the proposed panel bootstrap tests have Type I errors close to nominal sizes and acceptable power. An empirical illustration on the Feldstein-Horioka puzzle for a panel of 12 economies over the period 1960-2002 shows how the bootstrap panel stability tests lead to a more plausible conclusion (at least one break in the cointegrating relationship in the panel as a whole) than the asymptotic tests applied to each individual country (which, with a few exceptions, do not reject stability). Among the points on our research agenda we can mention generalising our procedures to tests of the hypothesis of breaks limited to only some of the variables, implementing some block-length selection algorithm, and exploring the use

of the Bewley (1979) transform.

6 References

Bai J., Ng S. 2004 "A PANIC Attack on Unit Roots and Cointegration" *Econometrica* 72:1127-1177. Further information in IDEAS/RePEc

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. Further information in IDEAS/RePEc

Bewley, R.A. (1979) "The Direct Estimation of the Equilibrium Response in Linear Models" *Economics Letters* 3:375-381. Further information in IDEAS/RePEc

Di Iorio, F., Fachin S. (2007) "Feldstein-Horioka Revisited: Testing for Cointegration with Breaks in Dependent Panels" *Working Paper* http://mpra.ub.uni-muenchen.de/3280/

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

Emerson, J., Kao, C. (2001) "Testing for Structural Change of a Time Trend Regression in Panel Data: part I" *Journal of Propagations in Probability and Statistics* 2:57-75.

Emerson, J., Kao, C. (2005) "Bootstrapping and Hypothesis Testing in Non-Stationary Panel Data" *Applied Economics Letters* 12:313-318. Further information in IDEAS/RePEc

Fachin, S. (2007) "Long-Run Trends in Internal Migrations in Italy: a Study in Panel Cointegration with Dependent Units" *Journal of Applied Econometrics* 22:401-428. DOI:10.1002/jae.907 Further information in IDEAS/RePEc

Feldstein, M. and C. Horioka (1980) "Domestic Saving and International Capital Flows" *Economic Journal*, 90:314-329. Further information in IDEAS/RePEc

Frankel, J.A. (1992) "Measuring International Capital Mobility: A Review" *American Economic Review* 82:197-202. Further information in IDEAS/RePEc

Gengenbach, C., Palm, F., Urbain, J.P. (2006) "Panel Cointegration Testing in the Presence of Common Factors" *Oxford Bulletin of Economics and Statistics* 68:S1 683-719. DOI:10.1111/j.1468-0084.2006.00452.x Further information in IDEAS/RePEc

Gonzalo, J. (1994) "Five alternative methods of estimating long-run equilibrium relationships" *Journal of Econometrics* 60:203 -233. Further information in IDEAS/RePEc

Gregory, A.W., Hansen, B.E. (1996) "Residual-based tests for cointegration in models with regime shifts" *Journal of Econometrics* 70:99-126. Further information in IDEAS/RePEc

Gregory, A.W., Nason, J.M., Watt, D. (1996) "Testing for Structural Breaks in Cointegrated Relationships" *Journal of Econometrics* 71:1-2 321-341. Further information in IDEAS/RePEc

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

http://129.3.20.41/eps/em/papers/0505/0505007.pdf

Hansen, B.E. (1992) "Tests for Parameter Instability in Regressions with I(1) Processes" *Journal of Business and Economic Statistics* 10:321-335. Further information in IDEAS/RePEc

Horowitz, J.L. and N.E. Savin (2000) "Empirically relevant critical values for hypothesis tests: A bootstrap approach" *Journal of Econometrics* 95:375-389. Further information in IDEAS/RePEc

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

Kao, C., Chiang, M-H. (2000) "Testing for Structural Change of a Cointegrated Regression in Panel Data" *Mimeo*, Center for Policy Research, Syracuse University. Further information

Kejriwal, M. (2007) "Cointegration with Structural Breaks: An Application to the Feldstein-Horioka Puzzle" *Studies in Nonlinear Dynamics & Econometrics* Forthcoming.

Moon, H.R. (1999) "A note on fully-modified estimation of seemingly unrelated regressions models with integrated regressors" *Economics Letters* 65: 25-31.

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

Paparoditis, E., Politis, D.N. (2001) "The Continuous-Path Block Bootstrap" In Madan Puri (ed.), *Asymptotics in Statistics and Probability. Papers in honor of George Roussas.* 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:(0) 653-670. Further information in IDEAS/RePEc

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. Further information in IDEAS/RePEc

Pesaran, M.H. (2007) "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence" *Journal of Applied Econometrics*, 22:265-312. Further information in IDEAS/RePEc

Politis, D.N., Romano, J.P. (1994) "The Stationary Bootstrap" *Journal of the American Statistical Association* 89:1303-1313.

Politis, D. N., White, H. (2003) "Automatic Block-Length Selection for the Dependent Bootstrap" *Econometric Reviews*: 23:53-70. Further information in IDEAS/RePEc

Westerlund, J. (2006) "Testing for Panel Cointegration with Multiple Structural Breaks" *Oxford Bulletin of Economics and Statistics* 68:101-132. Further information in IDEAS/RePEc

Westerlund, J., Edgerton, D. (2007) "A Panel Bootstrap Cointegration Test" *Economics Letters* 97:185-190.