A Panel Cointegration study of the long-run relationship between Savings and Investments in the OECD economies, 1970-2007

Di Iorio, Francesca and Fachin, Stefano

University of Naples Federico II, University of Rome "La Sapienza"

10 October 2010

A Panel Cointegration study of the long-run relationship between Savings and Investments in the OECD economies, 1970-2007

Francesca Di Iorio
University of Naples Federico II

Stefano Fachin*
University of Rome "La Sapienza"

Abstract

In this paper we test for the existence of a long-run savings-investments relationship in 18 OECD economies over the period 1970-2007. Although individual modelling provides only very weak support to the hypothesis of a link between savings and investments, this cannot be ruled out as individual time series tests may have low power. We thus construct a new bootstrap test for panel cointegration robust to short- and long-run dependence across units. Thid test provides evidence of a long-run savings-investments relationship in about half of the OECD economies examined. The elasticities are however often smaller than 1, the value expected under no capital movements.

Keywords:Savings, Investments, Feldstein-Horioka puzzle, OECD, Panel Cointegration, Stationary Bootstrap.
JEL codes: C23, C15, E2
(*) corresponding author. e-mail: s.fachin@caspur.it; tel: +39-06-49910834.

Revised October 2010

1 Introduction[1]

From the fundamental macroeconomic relationship $Y = C + I + B$, where as usual Y denotes Gross Domestic Product (GDP), C consumption, I capital formation and B the current account, we know that in an open economy capital formation is not constrained by domestic savings ($S = Y - C$), as $I = S - B$. However, Feldstein and Horioka (1980) documented a close correlation between savings and investments as ratios of GDP (hereafter, simply savings and investment ratios) in 16 OECD economies[2] over the period 1960-1974. This evidence, christened "Feldstein-Horioka puzzle" (in fact, "the mother of all puzzles", according to Obstfeld and Rogoff, 2000) for its stark contrast with the prevailing perception of capitals mobile enough to make the external constraint not binding, has stirred an enormous literature. The recent survey by Apergis and Tsoumas (2009), lists nearly 200 references, many of which empirical analyses greatly differing for methods applied and datasets (countries, time period, or both) studied. While Feldstein and Horioka (hereafter FH) used a cross-country regression of time-averaged variables, time series or panel studies now largely prevail in the literature; the variability of the samples studied, with respect to both the countries (LDC's, OECD, EU) and the time period, creates a comparable variability of results. Apergis and Tsoumas (2009) conclude that overall the evidence of a saving-investment relationship is weaker for developing countries than for richer ones, which is in fact not too surprising in view of the typical importance for the former of foreign aid and direct investments. In the OECD economies early evidence generally provided strong support to the existence of a one-to-one savings-investment relationship, but later studies are much more cautios. For instance, Jansen (1996) on the basis of a variety of tests strongly rejected the hypothesis of no long-run savings-investment relationship for a sample of 20 OECD economies from the early 1950's to the early 1990's. Using datasets starting in 1960 and ending in the early 1990's, Kim (2001) and Coiteux and Olivier (2000) respectively never rejected no cointegration for a panel of 19 OECD economies and only in four cases in a slightly larger group of 22 countries. Other references along this line are reported by Apergis and Tsoumas (2009). The obvious explanation of this decreasing correlation is the increasing financial integration, which considerably accelerated with the introduction of the Euro in 1999 in about half of the countries typically included in these samples.

Summing up, the question if, quoting Blanchard and Giavazzi (2002), we are close to "the end of the Feldstein-Horioka puzzle" is both open and challenging. To answer it we need to improve on the existing literature from two points of view: data and methods.

From the data point of view, a first remark is that the periods examined in the most updated studies on OECD economies (to the best of our knowledge, Coakley, Fuertes and Spagnolo, 2004, Pelgrin and Schich, 2008, AmirKhalkhali, Dar and AmirKhalkhali 2003) start around 1960 and do not extend beyond 2000. Hence, on one hand they include the 1950's and 1960's, decades of strict capital controls, on the other they fail to cover the most recent years of accelerating financial globalisation. To evaluate if the FH puzzle is currently valid we instead need to use a sample excluding the earlier decades and as updated as possible[3]. We will thus use a sample of 18 OECD economies (Austria, Australia, Belgium, Canada, Denmark, Finland, France, Ger-

[1]Research supported by the Department of Statistics of the University of Naples Federico II, University of Rome "La Sapienza" and MIUR. Comments and suggestions from participants to the conferences on "Bootstrap and Time Series" (Kaiserlautern, June 2008) and "Factor Structures for Panel and Multivariate Time Series Data" (Maastricht, September 2008) are gratefully acknowledged. Correspondence to: s.fachin@caspur.it, fdiiorio@unina.it.

[2]Austria, Australia, Belgium, Canada, Denmark, Finland, Germany, Greece, Ireland, Italy, Japan, Netherlands, Portugal, Sweden, UK, USA.

[3]Our aim is thus radically different from, *e.g.*, Hoffmann's (2004), who investigated the validity of the FH equation over periods stretching for about 150 years, from the mid-XIX century to the early 1990's.

many, Greece, Ireland, Italy, Japan, Netherlands, Portugal, Spain, Sweden, UK, USA) over the period 1970-2007, 38 annual observations including eight years of the Euro era. The latter is an important point, since 11 out of the 14 european countries of our sample adopt the european common currency.

From the econometric point of view our task is particularly challenging. A failure to reject the null hypothesis, which in our case is no saving-investment relationship, can be taken as a convincing piece of evidence only if a powerful statistical procedure is used. Since the savings and investment ratios are very likely to be non-stationary[4] the null hypothesis of interest is no-cointegration, and with a sample of 38 observations standard no-cointegration tests should be expected to be not very powerful. Moreover the question is on the general validity of the FH puzzle, rather than in some given country. It is thus quite natural to turn to panel procedures[5]. Unfortunately, the solutions adopted so far seems to be inadequate. For instance one of the most recent studies, Pelgrin and Schich (2008), concludes in favour of cointegration between savings and investment in the usual panel of 20 OECD economies (and analogously, Kim, Oh and Jeong, 2005, for 11 asian countries) on the basis of various panel cointegration tests. However, all these tests are severely oversized if the assumption of independence across units is not met (Banerjee, Marcellino and Osbat, 2004). Since this condition is very unlikely to hold for the closely integrated OECD economies, Pelgrin and Schich's conclusion in favour of a long-run saving-investment relationship in the OECD is very likely to be flawed. Summing up, to reach reliable conclusions we need a powerful panel cointegration test robust to short- and long-run dependence across units.

We shall now first examine the data and estimate Auto Regressive Distriubuted Lags (ARDL) models for the invidual economies (section 2). These overall find only some weak support for a relationship between savings and investment for our data set. To exclude that this may be simply due to the low power of the tests and to reach a conclusion for the panel based on a formal testing procedure we construct (section 3) and evaluate by simulation (section 4) a suitable bootstrap panel cointegration test, delivering considerable power gains. Applying this test to our dataset (section 5) we conclude that a long-run savings-investments relationship holds in about half of the OECD economies examined. Some overall conclusions are drawn in section 6.

2 Savings and Investments in the OECD, 1970-2007

As a first glance at the data, following FH we look at the averages over time of the investment and savings ratios for their set of countries expanded to include Spain and France, two major european economies which should not be ignored. As mentioned above, since our question is if the puzzle is currently valid we consider a much later period, 1970-2007 (instead of 1960-1974).

The visual impression given by Fig. 1 is indeed of some dependence between the two variables, but running OLS we obtain a regression coefficient (called saving retention ratio in the literature) equal to 0.16, only slightly higher than its standard error (0.12), hence not significant according to standard inference. The correlation, thus, seems to have fallen, as FH report a cofficient equal to 0.89 and strongly significant. Clearly, Greece, with the lowest saving ratio of the panel

[4] Although ratios obviously lie in the [0,1] interval, modelling is based on the logs, which are bounded only above. Since both ratios are typically small (around 0.20 for the OECD economies between 1970 and 2007) empirically the upper bound is irrelevant, and the variables may behave as realizations of unit root processes.

[5] In fact, the evidence from studies on multiple units (here, countries) is inevitably evaluted in an informal panel perspective, with conclusions holding for a large fraction of the cases considered as supported by the data. Clearly, a formal testing procedure, such as a panel test, is preferable.

but a fairly high investment ratio, is likely to have an influential role in these results. Although it may be hasty to consider it an outlier (Blanchard and Giavazzi, 2002, argue that it is a perfect example of the consequences of financial integration, a process which in the long run will affect all economies), considering the debt crisis this economy is facing since the end of 2009 (European Commission, 2010) it is nevertheless interesting to repeat estimation in a restricted sample excluding it. The coefficient is now 0.52, with a standard error of 0.15: significant, but still much smaller than FH's. A first interesting remark is thus that replicating exactly FH's computations for a later period we are left with the impression of a much weaker link between domestic savings and capital formation.

Of course, we know that looking at time average data can be deceveing and that these naive regressions have little, if any, meaning. We then move to the detailed time series for the individual countries, reported in Figs. 2 and 3. The overall impression is that the variables generally followed very close paths. The most notable exceptions are Greece, Ireland and the Netherlands. In Greece the investment ratio has indeed consistently been much higher than the saving ratio, while in Ireland this happened until the mid-90's. Finally, in the Netherlands, on the opposite, savings always exceeded investments, with a widening differential.

It is interesting to see that in the USA, where as well-known savings always fell short of investments, the two variables nevertheless share some large swings (troughs in early 1990's and around 2004, peak in early 2000's); very much the same holds for Australia. Finally, in some countries the association seems to became weaker after breakpoints varying betwen the early (Belgium and Finland) and the late 1990's (Portugal, Spain, perhaps Germany). In the latter case the timing naturally suggests a possibile influence of the introduction of the Euro in 1999. For Portugal, where savings follow a negative trend not shared by investments, this view is supported by Blanchard and Giavazzi (2002).

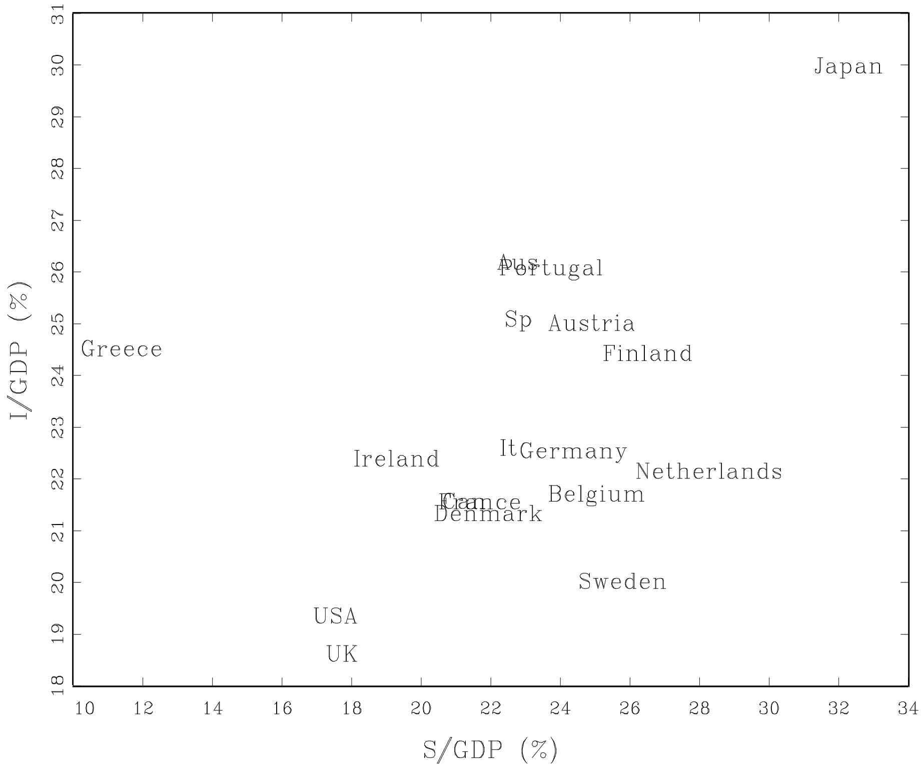

Fig. 1 Savings/GDP and Investment/GDP in OECD countries, 1970-2007 averages. Abbreviations: It: Italy; Sp: Spain. The countries with overlapping labels are: Australia and Portugal (at the top), France, Canada and Denmark (in the middle).

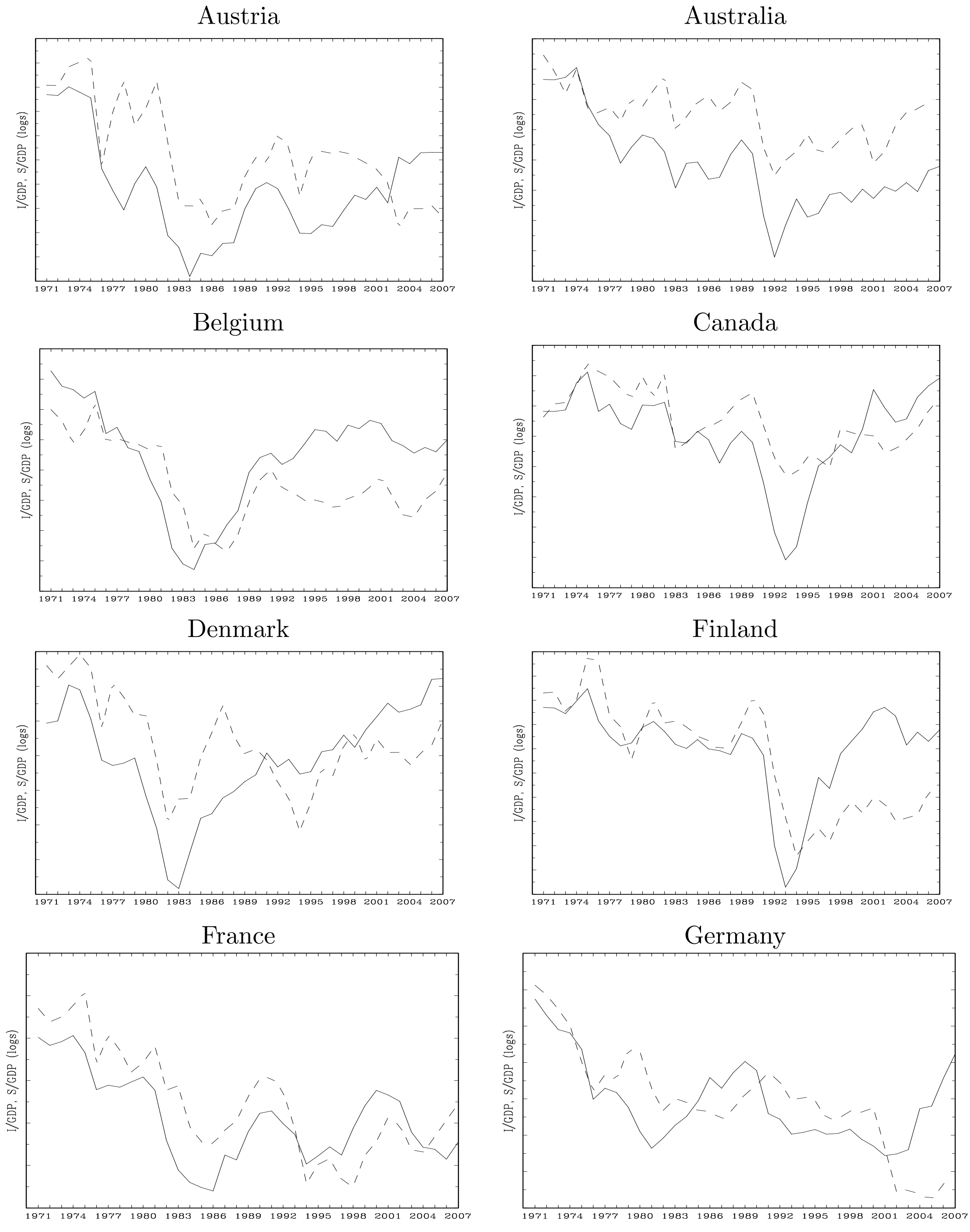

Fig. 2 Savings/GDP (solid line) and Investment/GDP (dashed line), 1970-2007 (logs).

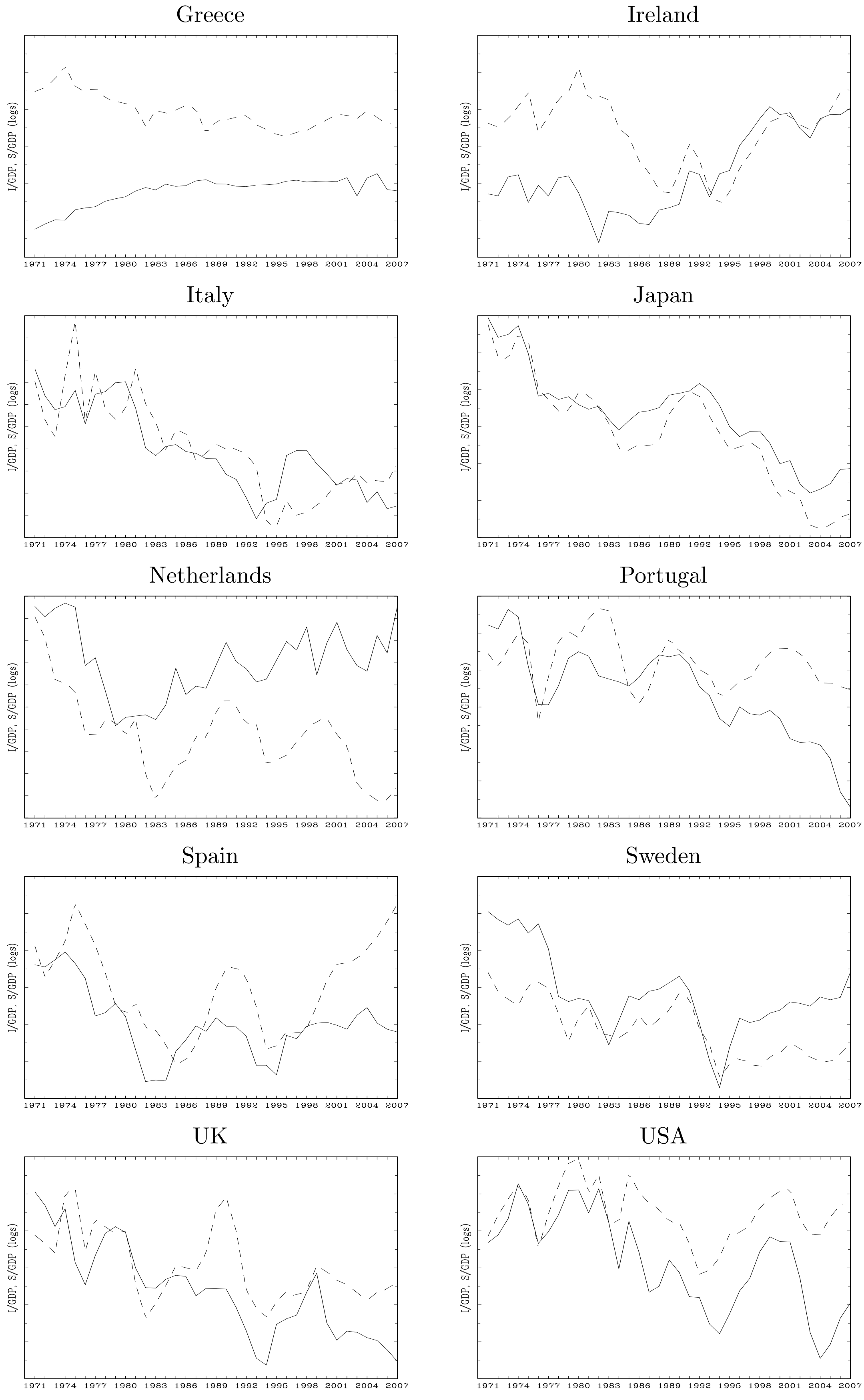

Fig. 3 Savings/GDP (solid line) and Investment/GDP (dashed line), 1970-2007 (logs).

Consistently with this graphical evidence and with the literature the results of ADF unit root tests, reported in detail in Table A1 in the Appendix, are largely in favour of unit roots. The only clear exception is investments in Portugal, for which the unit root hypothesis is rejected with a p-value of 0.6%. In the Netherlands and UK for the same variable the p-values are borderline with respect to the traditional 5% threshold (respectively, 5.0% and 5.1%).

Given that savings and investment ratios are generally, but not always, non-stationary, the natural next step is modelling their relationship using the ARDL approach (Pesaran, Shin and Smith, 2001). For a given country i the starting point is the conditional dynamic model

$$\Delta i_t = \psi + \beta_1 s_{t-1} + \beta_2 i_{t-1} + \sum_{i=0}^{p_1} \gamma_{1i} \Delta s_{t-i} + \sum_{i=1}^{p_2} {}_2\gamma_{2i} \Delta i_{t-i} + \varepsilon_t \qquad (1)$$

where $i_t = \ln(Investments/GDP)$ and $s_t = \ln(Savings/GDP)$. Equation (1) provides a basis for both estimation of the long-run saving-retention ratio as $\beta = -\beta_1/\beta_2$ and for tests of the existence of a relationship between the levels of the variables. To this end Pesaran, Shin and Smith (2001) recommend the $F-$test for $H_0 : \beta_1 = \beta_2 = 0$, while Banerjee, Dolado and Mestre (1998) proposed as a cointegration test the $t-$test for $H_0 : \beta_2 = 0$. We thus estimated equation (1) for each country of our panel, selecting the dynamic structure on the basis of the standard model selection criteria (AIC, Hannan-Quinn, Schwarz) and checking the presence of autocorrelation in the residuals through LM tests. The main results are reported in Table 1, with details of dynamic structure and diagnostics in Table A2 in the Appendix.

From Table 1 we immediately see that both the $F-$test for $H_0 : \beta_1 = \beta_2 = 0$ and the $t-$test for $H_0 : \beta_2 = 0$ mostly fail to reject even at 10% the hypothesis of no level relationship between savings and investment. More precisely, evidence of a relationship is found only in seven cases out of 18: Australia, Finland (only on the basis of the $F-$test under stationarity), Greece (where the elastiticity is however, as expected, negative, hence the relationship spurious), Italy, Japan, Portugal and UK. Excluding Greece, in half of these cases the saving-retention ratio is smaller than 1 (Australia, Portugal, UK), in two cases slightly larger than 1 (Italy, Japan), and finally in one case (Finland) so large that it is likely to be very imprecisely estimated.

Trying to shed some more light on the issue we turn to the Engle-Granger cointegration tests (Table 2). Excluding Greece, these are significant at 5% in three cases (Japan, Portugal, UK) plus one more at 0.10 (Australia), thus delivering a very similar picture.

Summing up, a long-run savings-investment relationship seem to exist in a minority of the countries of our panel, with coefficients often far from the theoretically expected value of 1. Should we conclude that there is no FH puzzle any more? Not necessarily. As discussed above, we may fail to reject the hypothesis of no long-run relationship simply because of the small sample size. We need a more powerful procedure. This will be constructed in the next section.

Table 1

Saving and Investments in the long-run: ARDL conditional modelling and tests of a level relationship

	Austria	Australia	Belgium	Canada	Denmark	Finland
β	0.48	0.69	−1.78	0.93	0.62	3.17
F	1.62	5.05*	3.83	2.51	1.33	4.30§
t	−1.78	−2.89*	−0.60	−1.57	−1.42	−1.32
	France	Germany	Greece	Ireland	Italy	Japan
β	1.38	1.42	−1.17	1.18	1.14	1.32
F	1.70	1.59	7.42*	2.38	6.52*	8.56*
t	−1.58	1.38	−3.84*	−1.22	−3.41*	−3.90*
	Netherlands	Portugal	Spain	Sweden	UK	USA
β	0.27	0.25	1.21	1.75	0.67	0.28
F	1.47	5.96*	3.67	2.71	5.17*	2.62
t	−1.71	−3.45*	−2.54	−0.76	−3.11*	−2.20

$\beta = -\beta_1/\beta_2$, long-run saving retention ratio;
F: $H_0 : \beta_1 = \beta_2 = 0$; 0.10 critical values : $I(0) : 4.04$, $I(1) : 4.78$;
t: $H_0 : \beta_2 = 0$; 0.10 critical values: $I(0) : -2.57$, $I(1) : -2.91$;
§: significant at 0.10 only in the stationary case;
*: significant at 0.10 both in the stationary and non-stationary cases.

Table 2

Saving and Investments in the long-run: Engle-Granger cointegration tests

Austria	Australia	Belgium	Canada	Denmark	Finland
−0.35	−3.23	−1.69	−2.23	−2.31	−1.34
[0.97]	[0.09]*	[0.68]	[0.42]	[0.38]	[0.82]
France	Germany	Greece	Ireland	Italy	Japan
−1.64	−0.88	−3.55	−1.45	−2.78	−4.24
[0.71]	[0.92]	[0.04]**	[0.78]	[0.19]	[0.003]**
Netherlands	Portugal	Spain	Sweden	UK	USA
−2.23	−3.95	−1.92	−2.07	−3.47	−3.05
[0.42]	[0.01]**	[0.57]	[0.50]	[0.04]**	[0.12]

EG: Engle-Granger cointegration test, asymptotic p-values in brackets
*,**: significant at 0.10, 0.05.

3 Panel cointegration testing via residual-based bootstrap

3.1 Panel cointegration tests: overview

The rate of expansion of the literature on the analysis of non-stationary panels, as revealed *e.g.,* by a comparison of the list of references in the surveys by Banerjee (1999) and Breitung and Pesaran (2006) is impressive. This growing interest is due to good reasons: first, as in our case, many important economic questions are naturally framed in a panel perspective; second, adding the cross-section dimension grants considerable improvements of the small samples properties of testing procedures, provided the possible linkages across units are properly accounted for. Excluding the panel full information maximum likelihood approach (Groen and Kleibergen, 2003) which, requiring the time dimension to be much larger than the cross-section dimension, cannot be applied to our dataset (nor to most macroeconomic panels in general), two main ways to tackle this problem have been suggested: (i) modelling the linkages as due to unobserved common factors; these can be estimated by principal components methods (Bai and Ng, 2004) and then removed from the data so to apply simple procedures for independent panels (Banerjee and Carrion-i-Silvestre 2006, Gengenbach, Urbain and Palm, 2006, Westerlund 2008); (ii) apply bootstrap algorithms designed to deliver estimates of the distribution of the statistics of interest conditional on the cross-section linkages as present in the dataset at hand. To the best of our knowledge two bootstrap (no-) cointegration procedures have been put forth so far. Fachin (2007) applied the Continuous-Path Block bootstrap (Paparoditis and Politis, 2001, 2003) separately to the right- and the left-hand side variables, hence generating unrelated pseudoseries obeying the null hypothesis of no cointegration, while Westerlund and Edgerton (2007) developed a Sieve Bootstrap procedure.

Unfortunately, neither the common factor nor the existing bootstrap approaches are fully satisfactory. Let us discuss them in turn. A first problem with the common factor approach is that, as Gengenbach *et al.* (2006) explicitly admit, it requires large samples. Thus, although investigating the possible common factor structure of the data could be very important for its own sake, in many empirical applications the available information set may simply be not rich enough. A second problem is that it hinges upon a series of assumptions which may be very restrictive. For instance, Banerjee and Carrion-i-Silvestre (2006) and Westerlund (2008) allow for common factors in the cointegrating residuals but not in the variables themselves. This more general set-up is allowed by Bai and Carrion-i-Silvestre (2005) and Gengenbach *et al.* (2006), but at the cost of other restrictions: the former assume homogeneous cointegrating vectors, and the latter that the matrix of factor loadings is full rank and block-diagonal, hence ruling out the empirically relevant case of a single source of non-stationarity common across units and variables. For instance, in our case the national stochastic trends in savings (and, if cointegration holds, investments) may be linked to a global stochastic trend.

Block bootstrap, model-free methods were showed by Fachin (2007) to be empirically useful tools in tackling the problems at hand. However his algorithm destroys *any* relationship between the modelled variables, not only long-run ones. Hence, the bootstrap pseudodata obey not only the null hypothesis which we are interested in testing (no long-run relationship) but also the nypothesis of no short-run relationships. This is clearly unsatisfactory. On the other hand, the Sieve Bootstrap (shown to be valid for inference on cointegrating regressions by Chang, Park and Song, 2006) hinges upon the assumption of a linear structure of the cointegrating residuals.

We thus need to improve on the available methods. To this end, a natural route is to extend to the analysis of cointegration the Residual-based Stationary Bootstrap (RSB) test for unit roots developed by Parker, Paparoditis and Politis' (2006), henceforth PPP. This test is closely related to the block bootstrap panel unit root test shown by Palm, Smeekes and Urbain (2008) to be asymptoticall valid; the key advantage of the Stationary Bootstrap over the block bootstrap is that the resampled pseudo-series series are stationary, hence the name (Politis and Romano, 1994). In both cases the resampling involves chaining blocks of observations of the originary series starting at random locations, with the difference that in the Stationary Bootstrap the length is also random, while in the block bootstrap it is fixed. These resampling schemes may be applied to both single-equation and system cointegration tests, but the extension to the former, very popular in panel studies[6], is particularly natural.

3.2 Set-up

To introduce our procedure let us first ignore the panel dimension and consider for simplicity two $I(1)$ variables, X and Y, linked by a linear relationship

$$y_t = \mu + \beta x_t + \epsilon_t, t = 1, \dots, T \tag{2}$$

where ϵ is the equation noise. Consider then the AR(1) equation for the residuals

$$\epsilon_t = \rho \epsilon_{t-1} + \nu_t. \tag{3}$$

It is immediately seen that when H_0 : no cointegration holds $\rho = 1$,while when it does not $|\rho| < 1$. The hypothesis of no cointegration is then equivalent to H_0: $\rho = 1$.

Two important remarks are in order here. First, (3) is *not* a model of the cointegrating residuals; its purpose is only to define a parameter expressing the null hypothesis of interest. Second, the $\nu_t' s$ are always stationary, either H_0 holds or not: they can thus be resampled via the Stationary Bootstrap. An algorithm along the lines put forth in PPP, mean zero case, will then proceed as follows:

1. Estimate (2) by OLS, obtaining $\{\widehat{\epsilon}_t\}$;
1. Estimate (3) by OLS, obtaining $\widehat{\rho}$ and the residuals $\widehat{\nu}_t = \widehat{\epsilon}_t - \widehat{\rho}\widehat{\epsilon}_{t-1}$;
2. Construct the pseudo-residuals $\{\nu_t^*\}$ applying the Stationary Bootstrap to $\{\widehat{\nu}_t\}$:

   2.1 generate $L_1, \dots, L_T$ i.i.d. from a geometric distribution with parameter θ;
   2.2 for each $t \in [1, T-1]$ let $K_t = \inf \{k : L_1 + \dots + L_T \geq t\}$ and $M_t = L_1 + \dots + L_{K_t}$;
   2.3 generate $i_1, \dots, i_{K_t}$ i.i.d. from a uniform distribution on $\{2, \dots, T\}$;
   2.4 for all $t \in [1, K]$ set $\nu_t^* = \widehat{\nu}_{\left[(i_{K_t} + (t - M_t)) \bmod (T-1)\right]+2}$.

3. Cumulate $\{\nu_t^*\}$ obtaining pseudoresiduals $\{\epsilon_t^*\}$ obeying the null hypothesis of no cointegration;
4. Compute $y_t^* = \widehat{\mu} + \widehat{\beta} x_t + \epsilon_t^*$;
5. Estimate the cointegrating regression on the dataset $\{y_t^*, x_t\}$: $y_t^* = \widehat{\mu}^* + \widehat{\beta}^* x_t + \widehat{\epsilon}_t^*$;
6. Estimate ρ^* applying (3) to the residuals $\widehat{\epsilon}_t^*$;
7. Repeat 2-6 B times;

[6]Note that, although single-equation modelling should in principle be limited to the case of weakly exogenous right-hand side variables, from Haug (1996) we know that using asymptotic critical distributions system and single-equation cointegration tests may be approximately equivalent if the signal/noise ratio (variance of the noise of the right-hand side variable divided by the variance of the noise of the left-hand side variable) is not too small.

8. Test the hypothesis $\mathrm{H}_0 : \rho = 1$ on the basis of the distribution of the $\rho^{*\prime}s$, which obey it. Note that the consistency results reported in PPP are in fact general enough to allow the use of more general statistics function of ρ, such as the ADF.

A few remarks are in order. First, the OLS estimator of ρ satisfies

$$\widehat{\rho} = \begin{cases} \rho + o_p(1) & \text{if } \rho < 1 \\ \rho + O_p(T^{-1}) & \text{if } \rho = 1 \end{cases} ; \tag{4}$$

Second, step 4 of the algorithm assumes exogeneity of the right-hand side variable X. In principle this assumption is easy to relax, extending the resampling to X as well. In practice, this is computationally demanding and will be left for future research. Obviously, in the simulation experiment we will check the robustness of the results with respect to this assumption.

Third, an important, and still largely unsettled, aspect of block bootstrap methods is the choice of block length. In the case of the Stationary Bootstrap the length is random, and the choice to be made is on the parameter θ of the geometric distribution (step 2.1 of the algorithm), which determines the mean block length. To prove the asymptotic validity of the test PPP assume $\theta \to 0$ as $T \to \infty$, so that $\sqrt{T}\theta \to \infty$. In practice, in the finite samples used in simulation experiments even naive choices for the mean block length, such as $0.10T$ used by Paparoditis and Politis (2001), seem to deliver good results. In our simulations we will thus fix the mean block length at either $0.10T$ or $0.15T$ (with a minimum of 4), leaving implementation of data-based methods, such as the Warp-Speed calibration of Palm, Smeekes and Urbain (2008), to future research.

Finally, it is instructive to compare the procedure proposed here with the model-based sieve bootstrap applied by Westerlund and Edgerton (2007). In this approach, assuming an autoregressive linear structure, (3) is replaced with a linear autoregressive model:

$$\epsilon_t = \rho\epsilon_{t-1} + \sum_{j=1}^{p} \phi_j \Delta\epsilon_{t-j} + \widetilde{\nu}_t \tag{5}$$

so to obtain empirically white noise $\widetilde{\nu}_t^{\prime}s$ on which simple resampling may be applied; the bootstrap residuals $\{\epsilon_t^*\}$ are then constructed recursively on the basis of (5). In other terms, in this approach the memory of the residuals is assumed to be captured by a linear autoregressive model. No assumption of this sort is required by PPP's approach, in which the memory of the process is captured resampling block of data.

Let us now introduce the panel dimension, ignored so far. The basic idea of unit root and panel cointegration tests is that of achieving power gains by pooling or averaging (the latter approach being more general, as no homogeneity constrains are imposed) the information from N individual units, indexed by i in the following discussion. However, the null and alternative hypotheses of these tests are a delicate issue worth a careful discussion (see also Pedroni, 2004). Given the null hypothesis of no cointegration in all units, *i.e.* $H_0 : \rho_i = 1$ for $i = 1, \ldots, N$, we can define four different alternative hypotheses:

(i) $H_1 : \rho_i < 1$ in *all* units;

(ii) $H_1 : \rho_i < 1$ in *at least one* unit;

(iii) $H_1 : \rho_i < 1$ in *most* of the units;

(iv) $H_1 : \rho_i < 1$ in *most* of the units *or* $\rho_i << 1$ in a smaller number of units.

The alternative hypothesis clearly dictates the choice of the summary statistic on which to base the panel test:

(i) $G = Max(\rho_i)$, as *all* $\rho_i's$ should be significantly smaller than 1 in order to reject $H_0 : \rho_i = 1$ against $H_1 : \rho_i < 1$ in *all* units;

(ii) $G = Min(\rho_i)$, as a *single* ρ_i significantly smaller than 1 is sufficient to reject $H_0 : \rho_i = 1$ against $H_1 : \rho_i < 1$ in *at least one* unit;

(iii) $G = Median(\rho_i)$, as to reject $H_0 : \rho_i = 1$ against $H_1 : \rho_i < 1$ in most of the units the mass of the distribution should be significantly far from 1;

(iv) $G = Mean(\rho_i)$, for the properties of the mean.

The same holds for transformations of ρ such as the ADF. Now, all asymptotic procedures are based on the mean of the individual statistics, and thus test the null hypothesis of cointegration in no units against the alternative hypothesis (*iv*). This, however, does not seem to be particularly meaningful from the empirical point of view. When the interest is centred on the properties of a given set of units (*e.g.*, the OECD countries, or the US states) the most relevant alternative hypothesis are (*ii*), cointegration in all units, and (*iii*), cointegration in the majority of them. This would require using as panel statistics respectively the maximum and the median, both of notoriously difficult treatment by asymptotic methods. Hence, bootstrap methods appear once again to have a high potential in panel cointegration testing; in fact, Fachin (2007), report good performances of a median-based bootstrap panel cointegration test.

This leads us to the next question: how to extend the algorithms outlined above to panel data sets? In fact, the task turns out to be easily accomplished. As already discussed above, an essential feature to be taken into account is dependency across units. In order to reproduce it in the pseudoseries, in both cases we simply need to apply the resampling algorithm to the entire cross-sections. In this way the (short- and long-run) cross-correlation structure of the data is exactly reproduced in the bootstrap data. More precisely, letting $\widehat{\nu}_{it} = \widehat{\epsilon}_{it} - \widehat{\rho}_i \widehat{\epsilon}_{it-1}$, in step 2 of the RSB algorithm we apply the stationary boostrap to the entire $T \times N$ matrix of the residuals $\mathbf{V} = [\widehat{\boldsymbol{\nu}}_1 \ldots \widehat{\boldsymbol{\nu}}_N]$, where $\widehat{\boldsymbol{\nu}}_i = [\widehat{\nu}_{1i} \ldots \widehat{\nu}_{Ti}]'$, In the final step the statistic of interest becames either the median or the mean of the cointegration EG statistics computed for each of the N units, so that the bootstrap estimate of the significance level of the test is $p^* = prop(S^* < \widehat{S})$, where S is the summary statistic adopted ($Median(\mathbf{EG})$, $Mean(\mathbf{EG})$, or $Max(\mathbf{EG})$, where $\mathbf{EG} = [EG_1 \ldots EG_N]$.) As mentioned above, Palm *et al.* (2008) prove that a block bootstrap unit root test is asymptotically valid, *i.e.* the bootstrap estimates of its p-values, p^*, converge asymptotically to the true p-values, p_0. To this end, they consider a fairly general common factor structure, allowing for both long- and short-run dependence across units. In the case of a no-cointegration test following the same approach would unfortunately be considerably more complicated, as the number of possible dependence structures grows geometrically with the dimensions of the process. A direct proof is also an exceedingly demanding task for tests based on order statistics, such as the median and the maximum. We shall thus follow a different strategy, proving the asymptotic validity of a RSB no cointegration test for time series, which in the case of independence is readily extended to the mean panel test, and evaluating by simulation the small sample performance of (*i*) the mean panel test under dependence and (*ii*) that of panel tests based on the median and the maximum.

3.3 Asympotic validity of the RSB No Cointegration test

To prove the asymptotic validity of the proposed boostrap test we follow the structure of the sequential limit arguments typical of the non stationary panel literature (see *e.g.*, Moon and

Phillips, 2000, Pedroni, 2004), examining first the asymptotics over time of the test applied to a single multivariate time series. The key point is that we do not actually need to derive the asymptotic distribution of the bootstrap test, but simply to show that it is the same of the test applied to the observed data (see, *e.g.*, Chang and Park, 2003). If this is the case, a hypothesis test based on critical values extracted from the RSB distribution will have asymptotic validity, as those critical values will be close to the ones extracted from the true null distribution. It is thus natural to concentrate our attention on the studentised statistic, the ADF, rather than a coefficient statistic, since for the former the fundamental results by Phillips and Ouliaris (1990) provide a convenient starting point.

Let $z_{it} = (y_{it}, x_{it})'$, $i = 1 \dots N$, be a bivariate integrated process with correlated innovations:

$$\begin{aligned} y_{it} &= y_{it-1} + w^y_{it} \\ w^y_{it} &= \beta_i w^x_{it} + \nu_{it} \\ x_{it} &= x_{it-1} + w^x_{it} \end{aligned}$$

where ν_{it} is a white noise process. The existence of short-run linear dependence between the two variables, y_i and x_i,can be emphasized rearranging the DGP as

$$\begin{aligned} y_{it} &= \beta_i x_{it} + \varepsilon^y_{it} \\ \varepsilon^y_{it} &= \varepsilon^y_{it-1} + \nu_{it} \\ x_{it} &= x_{it-1} + w^x_{it}. \end{aligned}$$

Since in this first step we are examining the asymptotic behaviour of the no-cointegration test for a given unit we can now drop, with some notational simplification, the index i. Assume that

A1 $\xi_t = (w^y_t, w^x_t)'$ is a linear process generated as in equation (3) of Phillips and Ouliaris (1990), hencefort PO, so that PO's conditions (C2) hold.

A2 the technical conditions $(i) - (v)$ in PPP hold for $\{\varepsilon^y_t\}$ and $\{\nu_t\}$.

Remark 1 *All stationary ARMA processes satisfy Assumption A1.*

Among the many results obtained by PO, for our purposes we are particularly interested in Lemma 2.1 and theorem 4.2. The latter describes the the asymptotic distribution of the cointegration ADF test for the model $y_t = \widehat{\beta} x_t + \widehat{\varepsilon}^y_t$, while the former states that the partial sum processes $S(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t$, $r \in [0, 1]$, (see PO, eqs. C1 and 2) converge to the vector Brownian motion

$$B(r) = (B_1(r), B_2(r))' \tag{6}$$

with covariance matrix $\Omega = \lim_{T \to \infty} T^{-1} E[(\sum_{t=1}^T \xi_t)(\sum_{t=1}^T \xi'_t)]$.

Consider now the bootstrap DGP under the null of no-cointegration (constant omitted for simplicity):

$$\begin{aligned} y^*_t &= \widehat{\beta} x_t + \varepsilon^*_t \\ \varepsilon^*_t &= \varepsilon^*_{t-1} + \nu^*_t \\ x_t &= x_{t-1} + w^x_t. \end{aligned}$$

where $\{\nu^*_t\}$ is obtained applying the SB to $\{\Delta \widehat{\varepsilon}^y_t\}$. This DGP satisfies all conditions required by PO's theorem 4.2. In particular, $\xi^*_t = (\nu^*_t, w^x_t)'$ is a linear process generated such that PO's equation (3) and condition (C2) hold. Further, since ν^*_t satisfies the Assumption A2, the Functional Central Limit Theorem 1 in PPP holds for its partial sum processes. Formally:

Lemma 1 *Define the partial sum process* $Z^*(r)$, $0 \le r \le 1$, *as:*

$$Z^*(r) = \frac{1}{\sqrt{T}\widehat{s}^*_r}(\sum_{t=1}^{\lfloor Tr \rfloor} \nu^*{}_t)$$

with $\widehat{s}^*_r = var^*(T^{-1/2}\sum_{j=1}^{\lfloor Tr \rfloor} \nu^*_j)$. *Let* $\xrightarrow{d^*}$ *denote convergence in distribution in the bootstrap world. Then under Assumption A2 the invariance principle:*

$$Z^*(.) \xrightarrow{d^*} W$$

where W *is the standard Wiener process, holds.*

Proof. See Theorem 1 in PPP. ■

Moreover, since the properties of the SB assure that $\widehat{\nu}^*_t$ is a stationary ARMA variable with variance converging to that of $\widehat{\nu}_t$ (PPP, lemma 4), the partial sum processes $S^*(r) = T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} \xi^*_t$, $r \in [0,1]$, converge to the vector Brownian motion (6). The propositions stating that the bootstrap cointegration and panel cointegration tests are asymptotically valid then follow.

Proposition 1 *Given assumptions A1 and A2, the asymptotic distribution of the cointegration test* ADF^* *for the bootstrap model*

$$y^* = \beta^* x_t + \widehat{\varepsilon}^*_t$$

converges to the asymptotic distribution described in PO, theorem 4.2, so that

$$\sup_{c \in \mathcal{R}} |P^*(ADF^* \le c | z_1 \dots z_T) - P_0(ADF \le c)| \xrightarrow{p} 0$$

where P_0 *is the probability measure obtained under the null hypothesis of no cointegration.*

Proof. Directly derive from above considerations. ■

Proposition 2 *Assume* $cov(w^x_{it}, w^x_{jt}) = cov(v_{it}, v_{jt}) = 0 \forall i \neq j$. *Then the panel cointegration bootstrap test based on the arithmetic mean of the statistics of the individual units is asymptotically valid:*

$$\sup_{c \in \mathcal{R}} \left| P^*(N^{-1}\textstyle\sum_{i=1}^N ADF^*_i \le c | z_{i1} \dots z_{iT}) - P_0(N^{-1}\sum_{i=1}^N ADF_i \le c) \right| \xrightarrow{p} 0$$

Proof. Directly from Proposition 1. ■

4 Monte Carlo evaluation

4.1 Design

Once established the asymptotic validity under independence of a test based on the mean of the individual tests we can now evaluate the small sample performance of this test as well as those of tests based on order statistics, such as the median and the maximum. Our Monte Carlo experiment is based on a DGP which is essentially a generalisation of the classic Engle and Granger (1987) DGP to the case of dependent panels, with the design of the panel structure related to those used by Kao (1999), Fachin (2007), and Gengenbach *et al.* (2006). Since panel DGPs are inevitably very complex, simulation experiments are computationally very demanding.

Hence, rather than aiming at the unfeasible task of a complete design our aim will be that of defining an empirically relevant set-up.

In our base case in the spirit of conditional modelling we assume a variable of interest, Y, known to be linked by a linear, possibly cointegrating, relationship to a right-hand side variable[7] X:

$$\begin{cases} y_{it} = \mu_{0i} + \beta_i x_{it} + \epsilon^y_{it} \\ \epsilon^y_{it} = \rho_i \epsilon^y_{it-1} + e^y_{it}, \qquad e^y_{it} \sim N(0, \sigma^2_{iy}) \end{cases} \tag{7}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$. When X_i and Y_i are not cointegrated $\rho_i = 1$, while $|\rho_i| < 1$ when instead they are. To mimick the empirically relevant case of rather slow adjustment to equilibrium, for the power simulations we will thus generate a set of $\rho_i' s$ as $Uniform(0.6, 0.8)$ across units and keep them fixed for all Monte Carlo simulations. To ensure some heterogeneity across units we analogously generate $\sigma^2_{iy} \sim Uniform(0.5, 1.5)$ and keep them fixed across experiments, while with no loss of generality we set $\mu_{0i} = \beta_i = 1 \; \forall i$.

The right-hand side variable X is constructed essentially as the sum of two terms. The first, u_x, is in turn the sum of a non-stationary factor common across units (F_1), a second stationary common factor (F_2) and an idiosyncratic stationary noise (ϵ^x_{it}). The second term, $a_i(\mu_{0i} + \epsilon^y_{it})$, captures the feedback from the left-hand side variable, absent when $a_i = 0$. Summing up:

$$\begin{cases} x_{it} = (1 - a_i\beta_i)^{-1}[u_{xt} + a_i(\mu_{0i} + \epsilon^y_{it})] \\ u_{xt} = \gamma_{1i} F_{1t} + \gamma_{2i} F_{2t} + \epsilon^x_{it} \end{cases} \tag{8}$$

To allow for some heterogeneity across units we generate the $a_i' s$ as $Uniform(0.2, 0.6)$, a rather wide range. The factor loadings are chosen so to ensure substantial cross-correlation in the $X's$; following Pesaran (2007, $\gamma_{ji} \sim Uniform(-1, 3) \; \forall i$, $j = 1, 2$. Again, both the $a_i' s$ and the $\gamma_{ji}' s$ are fixed across experiments. The common factors are generated as follows:

$$\begin{bmatrix} F_{1t} \\ F_{2t} \end{bmatrix} = \begin{bmatrix} F_{1t-1} \\ 0.4F_{2t-1} \end{bmatrix} + \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} \tag{9}$$

where, as in Gengenbach *et al.* (2006), both the common and idiosyncratic shocks are assumed to have a MA(1) structure:

$$\begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix} = \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} + \begin{bmatrix} \vartheta_{1i} & 0 \\ 0 & \vartheta_{2i} \end{bmatrix} \begin{bmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{bmatrix} \tag{10}$$

$$\epsilon^x_{it} = e^x_{it} + \varphi_i e^x_{it-1}, \tag{11}$$

where $\eta_{it} \sim N(0, 1), i = 1, 2$, and $e^x_{it} \sim N(0, \sigma^2_{ix})$, with $\sigma^2_{ix} \sim Uniform(1, 1.4)$. The $\varphi' s$ and the $\vartheta' s$ are generated as $Uniform(0.5, 0.7)$. Note that this DGP emphasises the linear dependence of Y on X: under the null hypothesis of no cointegration, when the linear dependence is confined the short-run, it can be easily rewritten so to define Y and X as cumulated sums of correlated stationary noises.

From the empirical point of view, our DGP is representative of many applications. Our FH equation is one case, but another obvious example is the case of regional consumption and income, with the common factors given by the trend and cycle in national GDP. The sample sizes considered in the experiment are also chosen trying to reproduce empirically relevant conditions. Considering that the OECD currently has 30 full members, we shall examine up to $N = 40$. We further assume that, as it often happens in practice, a full time sample of 160 observations (quarterly observations for 40 years) is available for the aggregate (average over all units), while the fully disaggregated sample only is available for $T = 20$ and 40.

[7] Exploratory simulations showed the performances of the test to be independent on the number of independent variables.

In order to compare the performances of our bootstrap test with those of the asymptotic test we shall first compute both tests on the aggregate data for $T = 20, 40, 160$, and then evaluate the gains possibly delivered by adding the panel dimension. We also consider an intermediate case, the first five units on which $T = 80$ time observations are assumed to be available. This allows checking the behaviour of the bootstrap panel cointegration test for growing T.

Finally, to strike a balance between experimental precision and computing costs the number of both Monte Carlo simulations and bootstrap redrawings has been set to 1000.

4.2 Results

The results are reported in tables 3-6 below. First of all, we check that in our set-up methods based on extracting the common factors from the residuals actually deliver poor results. To this end we report in Table 3 the Type I errors of the Durbin-Hausman group mean test DH_g by Westerlund (2008). The application of the test is obviously wrong here; a careful common factor analysis of the data would conclude that the residuals have no common factor, while the right-hand side variable does. However, unappropriate applications of available methods are pretty frequent in applied work, so that these results do have some interest. Since the common factor procedure fails to remove the dependence across units, the test heavily overrejects. In fact, when the X is generated according to the full specification (8)-(11) with two common factors the true null of no cointegration is *always* rejected. Letting $\gamma_{2i} = 0$ for all i in eq. (8) so that there is only one, non stationary common factor, the size bias falls but it is still very large, and, though shrinking with the time dimension, it worsens with the cross-section one for a fixed time sample. The problem is that, since the bias is exactly in the direction most welcome by practitioners (against H_0: no cointegration, hence in favour of the existence of a cointegrating relationship), they will probably be too happy of the results delivered by a routine application of the test to check carefully the validity of its assumptions.

Let us now move to the bootstrap tests. The results of the tests on the aggregate data (Table 4), show first of all that the cointegration RSB tests with the two different block sizes deliver essentially the same results. This is consistent with the performance of the unit root RSB tests in PPP and good news from the practitioner's point of view. Hence, to save space, for the panel tests we will report results only for mean block size $0.10T$, with those for mean block size $0.15T$ available on request. Second, the bootstrap tests deliver performances essentially comparable to those of traditional tests which compare the ADF cointegration statistics with MacKinnon (1991) critical values. Unfortunately, while Type I errors are always close to nominal values, power is rather disappointing in empirically relevant sample sizes: for instance, with $T = 40$ only slightly higher than 0.50 for a test with nominal size 0.05.

Can adding the panel dimension help? The results in Tables 5 suggest it can indeed. First of all, Type I errors are generally fairly close to nominal, although the test is generally oversized (recall that with 1000 Monte Carlo simulations the approximate 95% confidence intervals around 0.05 and 0.10 are respectively 0.04-0.06 and 0.08-0.12). Second, the power performance is very good: more than the high values of the rejection rates (which are conditional on the specific DGP and signal/noise ratio at hand), the important evidence here is their rapid growth with the cross-section dimension.

The good behaviour of the panel tests is confirmed by the results with $T = 80$ and the first five units (Table 6): Type I errors are only slightly higher than nominal sizes and power reaches 1. Hence, the overall conclusion is that the proposed test can help to answer to the question left open by the modelling of the individual economies: do we still have a FH puzzle in the OECD?

Table 3
Durbin-Hausman Common Factors
Group Mean DH_g
Panel Cointegration Test
Size

		Units			
T	α	5	10	20	40
20	0.05	0.06	0.42	0.70	0.88
	0.10	0.06	0.47	0.75	0.91
40	0.05	0.03	0.25	0.35	0.49
	0.10	0.04	0.30	0.43	0.58
80	0.05	0.05	0.08	0.09	0.11
	0.10	0.06	0.13	0.13	0.17

DGP:
X_i : cf. (8)-(11), $a_i = \gamma_2 = 0$
Y_i : cf. (7), $\rho_i = 1 \forall i$
H_0 : No cointegration.

Table 4
Asymptotic and Bootstrap
Aggregate Cointegration Tests

		Size			*Power*		
			RSB			*RSB*	
			Block size			Block size	
T	α	MK	$0.10T$	$0.15T^1$	MK	$0.10T$	$0.15T^1$
20	0.01	0.00	–	0.01	0.02	-	0.26
	0.05	0.02	–	0.06	0.16	-	0.45
	0.10	0.03	–	0.14	0.28	-	0.66
40	0.01	0.00	0.01	0.01	0.08	0.27	0.19
	0.05	0.05	0.04	0.04	0.42	0.61	0.49
	0.10	0.06	0.12	0.12	0.62	0.77	0.66
160	0.01	0.03	0.02	0.02	0.93	1.00	0.99
	0.05	0.05	0.06	0.07	1.00	1.00	1.00
	0.10	0.09	0.12	0.12	1.00	1.00	1.00

DGP: $\overline{X} = N^{-1} \sum_{i=1}^{N} X_i, \overline{Y} = N^{-1} \sum_{i=1}^{N} Y_i$;
X_i : cf. (8)-(11), $a_i = 0$;
Y_i : cf. (7), size: $\rho_i = 1 \forall i$; power: $\rho_i \sim Uniform(0.6, 0.8)$.
$test$: EG (ADF on cointegrating residuals), H_0 : No cointegration;
MK: Asymptotic test based on MacKinnon (1991) critical values;
RSB: Residual-Based Stationary Bootstrap
[1] :block size set to 4 when $T = 20$.

Table 5
Bootstrap Panel Cointegration Tests

		Units				*Units*				*Units*			
		5	10	20	40	5	10	20	40	5	10	20	40
		$Median(EG)$				$Mean(EG)$				$Max(EG)$			
T	α					*Size*							
20	0.05	0.06	0.07	0.04	0.03	0.07	0.07	0.05	0.03	0.08	0.06	0.05	0.03
	0.10	0.12	0.13	0.09	0.07	0.13	0.11	0.10	0.05	0.14	0.11	0.10	0.07
40	0.05	0.10	0.09	0.11	0.12	0.09	0.11	0.10	0.10	0.09	0.08	0.06	0.05
	0.10	0.15	0.18	0.21	0.23	0.18	0.18	0.19	0.22	0.15	0.15	0.13	0.10
						Power							
20	0.05	0.37	0.59	0.83	0.97	0.45	0.70	0.93	0.99	0.36	0.42	0.46	0.46
	0.10	0.54	0.76	0.92	0.99	0.62	0.82	0.97	1.00	0.51	0.54	0.58	0.57
40	0.05	0.95	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.88	0.89	0.91	0.92
	0.10	0.98	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.91	0.94	0.95	0.96

DGP:
X_i : cf. (8)-(11), $a_i = Uniform(0.2, 0.6)$; Y_i : cf. (7);
Size: $\rho_i = 1 \forall i$; Power: $\rho_i \sim Uniform(0.6, 0.8)$;
H_0 : No cointegration for all tests; H_1 : for $Median(EG)$, cointegration in most units;
for $Mean(EG)$, cointegration in a large number of units *or* strong cointegration in a smaller number of units; for $Max(EG)$, cointegration in all units.

Table 6
Bootstrap Panel Cointegration Tests
$T = 80$, $N = 5$

		α		
		0.01	0.05	0.10
$Median(EG)$	Size	0.01	0.06	0.14
	Power	1.00	1.00	1.00
$Mean(EG)$	Size	0.01	0.07	0.16
	Power	1.00	1.00	1.00
$Max(EG)$	Size	0.01	0.07	0.14
	Power	0.99	1.00	1.00

DGP: see Table 5;
Size: $\rho_i = 1\ \forall i$; Power: $\rho_i \sim Uniform(0.6, 0.8)$.
H_0, H_1 : see Table 5.

5 A Panel Cointegration test of the FH puzzle

In principle a critical point of the bootstap panel cointegration test developed in the previous section is the choice of the block size. In absence of well-established data-based methods a simple practical solution is to compute the test for a reasonably wide interval; in our case this can be 4 (slightly more than $0.10T$, the block size used in their simulations by Paparoditis and Politis, 2003) to 8 (slightly more than $0.20T$, a definitely large block size). We expect power to fall, hence p-values to grow, with the block size, as the bootstrap series tend to resemble more and more closely the data as the block size approaches T, when they will coincide. This indeed turns out to be the case, but as it can be appreciated from Table 7 the differences are very small. Computing the tests on the entire panel the p-value for the Mean of the individual EG tests is 0.12 with mean block size 4 and 0.15 with mean block size 8, while those of the Median are always larger than 0.20 (more precisely, 0.21 and 0.24). Dropping Greece from the panel all p-values increase slightly, with the smallest now becaming 0.16, still safely not significant. For the Eurozone the evidence in favour of the null hypothesis of no-cointegration is overwhelming even including Greece, with p-values close to 0.50 for the Mean and to 0.70 for the Median. Obviously, $Max(EG)$, which has a more restrictive alternative hypothesis than both $Mean(EG)$ and $Median(EG)$, never rejects.

At this point we can quite confidently conclude that investments does not seem to have been linked in the long-run to domestic savings in our sample of OECD economies as a whole, nor in the Eurozone. However, from the individual EG tests we know that this did happen in a few countries: excluding Greece, at the 10% level the hypothesis of no cointegration is rejected in Australia, Japan, Portugal, UK, while in a few other cases (USA, Italy) the estimated p-values are only slightly larger than 10%. This prompts a natural question: can we identify some subsample of our OECD group, obviusly wider than that limited to these four countries, such that the FH puzzle holds for every unit of the panel? From our discussion of the various alternative hypothesis we know that this hypothesis can be tested looking at the maximum of the cointegration statistics over the units.

We thus ordered the countries according to the p-values of the individual EG tests[8] and, starting with a panel including only the two units with the smallest p-values, Portugal and Japan, proceeded sequentially adding one country to the panel at each step and computing the p-value of the $Max(EG)$ statistic. In other terms, we are computing tests with the same null and alternative hypotheses (respectively,"cointegration in no unit" and "cointegration in all units") on a increasing sequence of nested samples of size $2, 3, \ldots, N$. Standard sequential tests, such as those proposed by Smeekes (2010) for this same purpose, keep the sample size and null hypothesis fixed, and change systematically the alternative hypothesis (here it would be in the first step "cointegration at least in unit 1", in the second "cointegration at least in units 1 and 2", etc.). The p-values (see Table 8) follow a very clear pattern: always smaller or equal to 1% for the first ten panels (hence up to the panel including Portugal, Japan, UK, Australia, USA, Italy, Denmark, Canada and Sweden), with a sudden increase to over 40% when Spain is added. While this last test is obviously not significant, evaluating the first ten requires some care. The point is that these tests, computed on a sequence of nested samples, are not independent. Hence, using some fixed significance level (such as the customary 0.05) we would completely lose control of the overall Type I error. A simple[9] solution is provided by the Bonferroni principle, which states that a multiple comparison test with individual significance level $\alpha_i^B = \alpha/N$, $i = 1, \ldots, N$ will have overall significance level α (see, *e.g.*, Savin, 1984). To evaluate the performances which can be expected in our set-up from a Bonferroni-type test we

[8] The ranking according to the value of the EG statistics is exactly the same from the third position included.

[9] Other approaches are described by Hanck (2009) and Smeekes (2010).

run a small simulation experiment with a DGP including 10 units and 40 time observations (as in the previous cases, 1000 Monte Carlo replications and Bootstrap redrawings). We computed the $Max(EG)$ test on all subpanels of cross-section size $1, 2, \ldots, 10$, and obtained the overall rejection rate as the proportion of simulations in which at least for one subpanel the $Max(EG)$ test rejected. The results, reported in Table 9, show that using Bonferroni individual significance levels indeed leads to overall Type I errors very close to overall nominal levels and high power.

Since all p-values of the first ten tests smaller or equal than 0.01, applying the Bonferroni principle we can then conclude that at an overall significance level of $0.01 \times 10 = 0.10$ there seem to have been a long-relationship between savings and investments in a subsample of our OECD panel, including Portugal, Japan, United Kingdom, Australia, United States, Italy, Denmark, Netherlands, Canada and Sweden.

Going back (Table 1) to the ARDL estimates, we can see that in four of these countries (Canada, Italy, Sweden and Japan) the long-run elasticity of investments to savings is close or even larger than 1, the value expected under no capital mobility. In three more cases (Australia, Denmark and UK) it is less than 0.70, while in USA, Netherlands and Portugal it is less than 0.30. Since in these countries cointegration holds we can compute the FM-OLS estimates, with the associated asymptotic standard errors. These estimates (Table 9) overall confirm the impression that, although in some of the OECD countries examined a link is likely to exist, it is also likely to be much weaker than the one-to-one relationship expected in absence of capital movements: Italy and Japan are the only countries apparently close to this condition. Unfortunately, the rather large standard errors advice against drawing more precise conclusions.

The countries where instead investments do not seem to depend on home savings are Spain, Belgium, France, Ireland, Finland, Germany and Austria. Size is obviously not a determinant of this clustering, nor is financial openness as measured by the averages of the Chinn-Ito index (Chinn and Ito, 2008). However, static classification on the basis of averages is not very satisfactory, as over the last decades the index is generally non stationary, with some evident jumps. Further work investigating the possible presence and relevance of breaks caused by the financial liberalisation process is needed.

Table 7

The long-run Saving-Investments relationship
in the OECD, 1970-2007
Bootstrap panel cointegration tests

	$Mean(EG)$	$Median(EG)$
Entire panel	-2.30 [0.12 − 0.15]	-2.23 [0.21 − 0.24]
Panel without Greece	-2.22 [0.16 − 0.23]	-2.23 [0.22 − 0.26]
Eurozone	-1.88 [0.45 − 0.48]	-1.64 [0.68 − 0.70]

EG: Engle-Granger cointegration test.
in brackets: $p-$values for block sizes 4 and 8, 1000 redrawings.
Entire panel: Austria, Australia, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, Portugal, Spain, Sweden, UK, USA;
Eurozone: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain.

Table 8

The long-run Saving-Investments relationship
in some OECD countries, 1970-2007
Bootstrap panel cointegration tests

Panel	N	*Countries*	$Max(EG)$
1	2	Portugal, Japan	-3.95 [0.0078]
2	3	Panel 1 + UK	-3.47 [0.0060]
3	4	Panel 2 + Australia	-3.23 [0.0020]
4	5	Panel 3 + USA	-3.05 [0.0018]
5	6	Panel 4 + Italy	-2.78 [0.0022]
6	7	Panel 5 + Denmark	-2.39 [0.0110]
7	8	Panel 6 + Netherlands	-2.23 [0.0056]
8	9	Panel 7 + Canada	-2.23 [0.0020]
9	10	Panel 8 + Sweden	-2.07 [0.0028]
10	11	Panel 9 + Spain	-0.84 [0.4214]

$Max(EG)$: Maximum of the Engle-Granger cointegration tests over the units included in the panel, testing:
H_0 : cointegration in *no* country of the panel , against
H_1 : cointegration in *all* countries of the panel ;
in brackets: $p-$values with mean block size 6, 5000 redrawings.

Table 9
Rejction rate of Bonferroni
Panel Cointegration Tests
$T = 40$, $N = 10$

α	*Size*	*Power*
0.05	0.08	0.95
0.10	0.10	0.91

DGP: see Table 5;
Size: $\rho_i = 1\ \forall i$; Power: $\rho_i \sim Uniform(0.6, 0.8)$.
For each subpanel of size $n = 1, 2, \dots, 10$:
H_0 : no cointegration;
H_1 : cointegration in all units;
Rejction rate: proportion of simulations in which
for at least one subpanel H_0 is rejected at $\alpha_i^B = \alpha/N$.

Table 9
Saving and Investments in the long-run:
FM-OLS estimates of the long-run saving retention ratio

Australia	Canada	Denmark	Italy	Japan
0.45 [0.11]	0.47 [0.17]	0.18 [0.28]	0.84 [0.19]	1.19 [0.11]
Netherlands	Portugal	Sweden	UK	USA
0.18 [0.36]	0.16 [0.16]	0.38 [0.16]	0.36 [0.24]	0.36 [0.15]

standard errors in brackets.

6 Conclusions

We started with a very simple question: is there a long-run savings-investments puzzle in the OECD? Previous answers were not satisfactory either because obsolete (the most recent datasets studied stopped at the end of the 1990's, exactly when global financial integration accelerated considerably) or metholodogically flawed. We then examined a panel of 18 OECD economies for a period as updated as possible (1970-2007) both individually and as a panel.

Individual modelling suggested a long-run relationship to be present in a small minority of the countries, often with rather small saving retention coefficients. To be sure that the failure to detect in more cases a relationship is not due to low power of the tests we developed and applied a novel bootstrap panel cointegration test. The conclusion is that there is evidence of a long-run savings-investments relationship in about half of the OECD economies examined (Portugal, Japan, United Kingdom, Australia, United States, Italy, Denmark, Netherlands, Canada and Sweden), but not in the other half (Spain, Belgium, France, Ireland, Finland, Germany and Austria)[10]. Except in two cases, Italy and Japan, the long-run elasticity is however much smaller than 1, the value expected in absence of capital movements. Essentially, in these countries imperfections in financial markets seem to create a partial home bias. Neither size, nor financial openness, as measured by the averages of the Chinn and Ito (2008) index, seem able to explain why the puzzle holds in the first cluster but not in the second. A possibly important issue to be explored is the role of the financial liberalisation process, and associated possible breaks

[10] In Greece the two variables have been linked by a spurious inverse relationship (not surprising in view of the 2009 financial crisis).

in the long-run saving-investment relationship. To this end we are currently working on the generalisation of the procedure put forth in this paper to the case of breaks at unknown dates.

7 Appendix

7.1 Data source and definitions

All data, in national currency at current prices, have been downloaded from the OECD.stat database on 26 June 2009. Definitions are as follows:

Investment: Gross capital formation (transaction code: P5S1).

Savings: Net savings (transaction code B8NS1) plus Consumption of fixed capital (transaction code K1S1).

Gross Domestic Product: transaction code B1_GS1.

7.2 Unit root tests

Table A1

Saving and Investments: ADF Unit root tests

	Austria	Australia	Belgium	Canada	Denmark	Finland
$ln(I/Y)$	−2.12 [0.24]	−2.82 [0.06]	−2.15 [0.23]	−1.92 [0.32]	−2.26 [0.19]	−2.08 [0.28]
$ln(S/Y)$	−1.77 [0.39]	−2.49 [0.12]	−2.44 [0.13]	−2.18 [0.21]	−1.87 [0.35]	−2.72 [0.07]
	France	Germany	Greece	Ireland	Italy	Japan
$ln(I/Y)$	−1.98 [0.29]	−2.12 [0.24]	−1.88 [0.34]	−1.80 [0.38]	−2.27 [0.18]	−1.18 [0.69]
$ln(S/Y)$	−2.50 [0.12]	−2.31 [0.17]	−2.67 [0.09]	−1.07 [0.72]	−1.85 [0.35]	−1.97 [0.30]
	Netherlands	Portugal	Spain	Sweden	UK	USA
$ln(I/Y)$	−2.94 [0.05]	−3.61 [0.06]	−1.76 [0.40]	−2.37 [0.16]	−2.85 [0.05]	−2.68 [0.09]
$ln(S/Y)$	−2.08 [0.25]	−1.11 [0.71]	−2.00 [0.29]	−2.54 [0.11]	−1.83 [0.36]	−1.33 [0.61]

Tests with constant, p-values in brackets; AR order selection: Ng and Perron (1995).

Table A2

Saving and Investments in the long-run:
ARDL conditional modelling

	Austria	Australia	Belgium	Canada	Denmark	Finland
$p(LM)$	0.38	0.82	0.47	0.27	0.35	0.28
p_1	0	0	0	0	0	0
p_2	-	-	1	1	1	2
	France	Germany	Greece	Ireland	Italy	Japan
$p(LM)$	0.62	0.20	0.94	0.52	0.99	0.72
p_1	0	0	0	0	1	1
p_2	-	1	1	1	2	2
	Netherlands	Portugal	Spain	Sweden	UK	USA
$p(LM)$	0.99	0.77	0.77	0.20	0.40	0.17
p_1	0	0	0	1	0	0
p_2	1	1	1	1	1	1

$p(LM)$: p-value of Breusch-Godfrey LM autocorrelation test of order 1
p_1 : maximum lag of Δs
p_2 : maximum lag of Δi (-: Δi excluded from the model)

8 References

AmirKhalkhali S, Dar A, AmirKhalkhali S. 2003. Saving–investment correlations, capital mobility and crowding out: some further results. *Economic Modelling* **20**: 1137–1149.

Apergis N, Tsoumas C. 2009. A survey of the Feldstein Horioka puzzle: What has been done and where we stand. *Research in Economics* **63**: 64-76.

Bai J, Carrion-i-Silvestre J. L. 2005 Testing Panel Cointegration with Unobservable Dynamic Common Factors mimeo, University of Barcelona.

Bai J, Ng S. 2004. A PANIC Attack on Unit Roots and Cointegration. *Econometrica* **72**: 1127-1177.

Banerjee, A. 1999. Unit Roots and Cointegration: an Overview. *Oxford Bulletin of Economics and Statistics*, **61**: 697-629.

Banerjee A, Dolado J, Mestre R. 1998. Error-correction mechanism tests for cointegration in single-equation framework. *Journal of Time Series Analysis,* **19**: 267–283.

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

Banerjee A, Marcellino M, Osbat C. 2004. Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data. *The Econometrics Journal,* **7**: 322-340.

Blanchard O, Giavazzi F. 2002. Current Account Deficits in the Euro Area: The End of the Feldstein-Horioka Puzzle? *Brookings Papers on Economic Activity,* **2**: 147-209.

Breitung J, Pesaran MH. 2006. Unit Roots and Cointegration in Panels In .*The Econometrics of Panel Data* Matyas L, Sevestre P. (eds), Kluwer Academic Publishers: Dordrecht (NL).

Chinn, MD, Ito, H. 2008. A New Measure of Financial Openness *Journal of Comparative Policy Analysis: Research and Practice*, **10**: 309-322.

Coakley J, Fuertes AM, Spagnolo F. 2004. Is the Feldstein Horioka puzzle history? *The Manchester School* **72**: 569-590.

Chang Y, Park JY. 2003. A Sieve Bootstrap for the Test of a Unit Root. *Journal of Time Series Analysis,* **24**: 379-400.

Chang Y, Park JY, Song K. 2006. Bootstrapping cointegrating regressions *Journal of Econometrics* **133**: 703-739.

Coiteux M, Olivier S. 2000. The saving retention coefficient in the long run and in the short run: Evidence from panel data. *Journal of International Money and Finance,* **19**: 535-548.

Engle RF, Granger, C. 1987. Co-integration and Error Correction: Representation, Estimation, and Testing *Econometrica* **55**: 251-276.

European Commission. 2010. *Report on Greek government deficit and debt statistics* Brussells (BE).

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.

Feldstein M, Horioka C. 1980. Domestic Saving and International Capital Flows *Economic Journal* **90**: 314-329

Gengenbach C, Palm F, Urbain, JP. 2006 Panel Cointegration Testing in the Presence of Common Factors. *Oxford Bulletin of Economics and Statistics* **68-S1**: 683-719.

Gregory AW, Haug A, Lomuto N. 2004. Mixed signals among tests for cointegration. *Journal of Applied Econometrics* **19**: 89–98.

Groen JJJ, Kleibergen F. 2003. Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models *Journal of Business and Economic Statistics* **21**: 295-318.

Hanck C. 2009. For which countries did PPP hold? A multiple testing approach *Empirical Economics* **37**: 93–103.

Haug AH. 1996. Tests for cointegration: a Monte Carlo comparison. *Journal of Econometrics* **71**: 89-115.

Hoffmann M. 2004. International capital mobility in the long run and the short run: Can westill learn from saving investment data? *Journal of International Money and Finance* **23**: 113-131.

Jansen WJ. 1996. Estimating saving-investment correlations: evidence for OECD countries based on error correction model. *Journal of International Money and Finance,* **15**: 749-781.

Kao C. 1999. Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics*, **90**: 1-44.

Kim SH. 2001. The saving–investment correlation puzzle is still a puzzle. *Journal of International Money and Finance* **20**: 1017–1034.

Kim H, Oh K, Jeong C. 2005. Panel cointegration results on international capital mobility in asian economies. *Journal of International Money and Finance* **24**: 71-82.

MacKinnon J. 1991. Critical Values for Cointegration Tests in *Long-Run Economic Relationships*, Engle RF, Granger C. (eds), Oxford University Press: Oxford (UK).

Ng S, Perron P. 1995. Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Association*, **90**: 268-281.

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

Palm FC, Smeekes S, Urbain JP. 2008 Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests METEOR Research Memoranda N. 48, Maastricht Research School of Economics of Technology and Organization.

Parker C, Paparoditis E. and Politis DN 2006 Unit root testing via the Stationary Bootstrap *Journal of Econometrics* **133**: 601-638.

Paparoditis E, Politis DN 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).

Paparoditis E, Politis, DN. 2003. Residual-based block bootstrap for unit root testing *Econometrica* **71**: 813-855.

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.

Pelgrin F, Schich S. 2008. International capital mobility: What do national saving-investment dynamics tell us? *Journal of International Money and Finance* **27**: 331-344.

Pesaran HM. 2007. A Simple Panel Unit Root Test In The Presence Of Cross Section Dependence. *Journal of Applied Econometrics* **22**: 265-312.

Pesaran HM, Shin Y, Smith RJ. 2001. Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics* **16:** 289-326.

Phillips PCB, Ouliaris S. 1990. Asympotic Properties of Residual Based Tests for Cointegration. *Econometrica* **58**: 165-93.

Moon HR, Phillips, PCB. 2000. Estimation Of Autoregressive Roots Near Unity Using Panel Data. *Econometric Theory* **16**: 927-997.

Politis DN, Romano J.P. 1994. The Stationary Bootstrap. *Journal of American Statististical Association* **89**: 1303-1313.

Savin NE. 1984. Multiple Hypothesis Testing. In *Handbook of econometrics*, vol. 2 ch. 14, Griliches Z Intriligator M (Eds), North-Holland: Amsterdam.

Smeekes S. 2010 Bootstrap Sequential Tests to Determine the Stationary Units in a Panel. *Working Paper*, Maastricht University.

Westerlund J. 2008 Panel Cointegration Tests of the Fisher Effect. *Journal of Applied Econometrics* **23:** 193-233.

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